Dating *Spodoptera* clades

Using fossil, secondary and ecological evidence

Mark Lammers
MSc Thesis
2012-2013

Mark Lammers
Registration number: 890206498090
e-mail: mark.lammers@wur.nl

MSc Thesis Biosystematics
BIS-30436

1st Supervisor: Dr. Freek T. Bakker, Biosystematics Group (BIS), Wageningen University (WUR)
e-mail: freek.bakker@wur.nl
2nd Supervisor: Drs. Marja J. van der Straten, Plant Protection Service (PPS), Netherlands Food and Consumer Product Safety Authority (NVWA)
e-mail: m.j.van.der.straten@minlnv.nl

Date: Wednesday, 03 July 2013

Preface

The beauty of insects fascinated me since I was 13 years old. At first I was happy with catching some and releasing all of them. But later I learned that many species need to be collected for proper documentation of the observation. At that time my insect collection was started. Out of this hobby came the first condition for my MSc thesis subject: It must be insect related. Secondly, I wanted a subject which has a more or less direct societal relevance. Doing fundamental research is satisfying, but I also want to be able to explain why my research benefits everyone. By working on a crop pest, everyone understands why someone (me) should spend more than half a year on it. Thirdly, I wanted a subject in which I can combine many novel analyses and learn new skills. And lastly, it should have a realistic possibility of publishing my research, so that I improve my chances of obtaining a PhD position. In the 'Systematics of *Spodoptera*' (as this project was initially called) I found it all.

Although I could not perform every part of the originally described proposal I still am very satisfied with the results. I gained experience in laboratory work, learned to work with DNA sequence analyses and even my little ten year old sister understands why this research is beneficial for humans. I was able to formulate new ideas and test some of them. I went to the International Lepidopterersists' Conference in Denver, Colorado in the summer of 2012 to present one of these ideas in front of a room with about 150 respected scientists. This was an even better experience than expected. If possible, I would like to test all other ideas I describe as well, but that would not fit into an MSc thesis project.

Unfortunately (for me) during my thesis Kergoat *et al*. (2012) was published, which forced me to shift the focus of my thesis. The contents of their study and the consequences for me are all described in this thesis. I start with zooming out from the subject, giving a short overview of systematics. Then, step-by-step I zoom in on *Spodoptera*. I hope that you (the reader) enjoys my work at least as much as I did doing mine!

TABLE OF CONTENTS

List of tables and figures

Summary

Spodoptera is a cosmopolitan genus of moths (Lepidoptera: Noctuidae) containing 31 species of which many are polyphagous. About half of these have been considered crop pests. These species also have a high fecundity and good dispersal ability and therefore pose a threat to agriculture.
Using DNA sequence data of 6 protein coding markers (total 4539 bp with gaps) of 21 *Spodoptera* species and 76 other Noctuoid species and outgroups I reconstructed the phylogeny of the Noctuid family in order to find the position of *Spodoptera* in the family and investigate its monophyly. I time-calibrated the resulting phylogenetic reconstruction using primary and secondary calibration points and a new concept I call 'ecological' dating. This is based on the idea that the emergence of an organ in one clade is related to the emergence of an organ in another clade; in this case the former are the tympanal organs in nocturnal Lepidoptera which are assumed to be in response to the emergence of bat echo-location. Readers interested in this topic might want to start reading at the end of the Material and Methods section (2.6.3 'Dating the clades')
Spodoptera was recovered as monophyletic and it has clearly discerned clades which I named according to the species in the clade. The sister genus is *Galgula*, but the position of this two-genus lineage remains unresolved. The estimated crown age of *Spodoptera* is between 5 and 11 million years under all relevant calibration combinations. Therefore I declared the calibration points as internally consistent.
Simultaneous with my thesis a study by Kergoat *et al*. was published, this turned out to have a major overlap with my objectives. The results of the molecular phylogenetic part of their study are similar to my reconstruction, but the results from the dated phylogenetic reconstructions are incongruent. All differences and similarities are described in a separate chapter.
In this thesis I set a first step in shedding light on the evolutionary history of this economically important genus. Hopefully this will lead to a better understanding of the functional genomics underlying the pest syndrome in *Spodoptera*. And we will gain insight in the host specificity and invasive potential of these crop pests in our changing world where food for man is likely to become scarce.

1. Introduction

The genus *Spodoptera* is commonly known as a pest genus with many species having a major impact on agriculture and international trade (see section 1.4.1 for more information and references on this subject). When I started this project there was no published proper molecular phylogenetic reconstruction of the genus, nor did we know where to position the genus in a broader sense, meaning where to position it in a Noctuoid phylogenetic reconstruction. This changed when we heard that Kergoat *et al*. (2012) were publishing a molecular phylogenetic reconstruction of *Spodoptera*. Details on that study were released in June, when my study was already partly finished. Therefore this thesis has to some extend overlap with their work. However, it appears that there is major incongruence between the results. In chapter 5 I review the work of Kergoat *et al*., describe the differences and suggest possible causes.
In this thesis I investigated many aspects of the systematics of the genus *Spodoptera*. Therefore I'll in this report first zoom out from the topic, evaluating the historical aspects of insect systematics and then step by step zooming in on the Lepidoptera and all the way in to *Spodoptera* clades.

1.1. Historical scope on Systematics

1.1.1. Morphological and molecular systematics

Traditionally, the only characters available for sytematici and taxonomists were morphological characters, leading to the field of cladistics (Hennig 1966). Over the past decades, molecular systematics has taken great flight, powered by new DNA sequencing techniques, availability of public data (GenBank, Benson *et al*. 2008) and increased computational power. Moreover, new evolutionary models of nucleotide substitution (e.g. the General Time Reversible model, Tavaré 1986; among-site heterogeneity, Yang *et al*. 1996) were developed and implemented in software packages (see section 1.1.4 'Analytical Concepts' for more information on this topic). These developments often led to new classifications at many different levels (see below for examples).

1.1.2. Some differences between botanical en entomological molecular phylogenetics

Unlike the almost completely resolved family-level plant phylogenetic reconstruction by the Angiosperm Phylogeny Group (APGIII 2009) and the 1kP project (a phylogenetics project sequencing the total RNA of 695 plant species so far, Johnson *et al*. 2012), there hasn't been a communal effort to investigate the Arthropod phylogeny. There are numerous phylogenetic papers on the arthropods (e.g. Meusemann *et al*. 2010, Regier *et al*. 2010, Letsch *et al*. 2012, Yeates *et al*. 2012) or a subset (e.g. Holometabola: Beutel *et al*. 2010), but most, if not all, have a limited data set. Moreover, there does not seem to be a community consensus in which markers should be used. With the progress of the 1KITE project ('1000 Insect Transcriptome Evolution' project, Misof 2012, http://www.1kite.org/) this might change.
There are other fundamental differences between plant and arthropod systematics. Most importantly, from the phylogenetic perspective, in plants is polyploidy a common phenomenon (e.g. Otto and Whitton 2000) and has many consequences (e.g. problems in assessing orthology, Mayfield-Jones *et al*. 2013), while (to my knowledge) there are no cases described in arthropods other than in some geographically parthenogenetic species (Otto and Whitton 2000, Mable 2004).

1.1.3. Challenges in insect molecular phylogenetics

Insects are the largest clade of all eukaryotic organisms on Earth, over a million are described and many more expected (May 1988, May 2010, Mora *et al*. 2011). This sheer number of species results in several problems for taxonomists and systematists: What is a species? How do you recognize them and what are the differences with its related species? How certain are we that we sampled all the diversity in our clade of interest? There are no short answers to these questions; maybe these questions cannot be answered at all. And I will certainly not be able to answer them here. Species are often defined based on genital morphology: The genitalia fit like a lock and key between the males and females of a species (e.g. Mikkola 2008). Therefore the morphology of the genitalia determines with which individuals a certain individual can mate, thus also from which it is reproductively isolated. This reproductive isolation is therefore the most used and most suitable species concept in insect taxonomy (Schoonhoven *et al*. 2005)
Another important feature in the evolutionary history of insects is the occurrence of so-called ancient rapid radiations. Insects diverged fast in the Permian and Jurassic, leading to short internal branches between crucial nodes in phylogenetic reconstructions, generally resulting in low support for the produced tree (Whitfield and Kjer 2008). Later rapid radiations occurred in the Lepidoptera, following the diversification of angiosperm

plants in the Cretaceous (Schoonhoven *et al*. 2005, Whitfield and Kjer 2008). This is the likely cause of the problems found in the reconstruction of the phylogeny of Noctuid moths (Mitchell *et al*. 2006, Zahiri *et al*. 2011, Zahiri *et al*. 2012).

Bakker (2007) described several problems encountered in a pilot study on DNA extraction from collection material older than about ten years. Some storage methods commonly used for the resulting extracted DNA quality were tested. PCR on isolated DNA from specimens older than 10 years often fails and is therefore discouraged.
The most important problem is DNA breakdown (Mandrioli *et al*. 2006). The extent hereof depends on the storage and preparation conditions of the specimens. Storage of specimens in acetone yields the best results for later DNA extraction and is recommended for future collections (Mandrioli *et al*. 2006). When the DNA is degraded, short fragments remain, which can still be sequenced and can function as 'mini' barcodes (Hajibabaei *et al*. 2006b, Shokralla *et al*. 2011). For Angiosperm plant herbarium material the extent of the damage was investigated and methods for overcoming these problems were published (Staats *et al*. 2011). Another project testing the quality of DNA of preserved insects was also carried out (Staats *et al*. PLoSOne in press.). They found that material older than 5 year has DNA breakdown into fragments of about 200 bp. The markers I isolated are at least 540 bp long, thus required less old material.
With Next-Generation Sequencing (NGS) it is possible to extract a full mitochondrial genome from just one leg of an old museum specimen (Staats *et al*. PLoSOne 2013). However, the current costs of NGS were too high for my project.

1.1.4. Analytical concepts

All the important concepts of phylogenetics are well-described in the textbook by Page and Holmes (1998), 'The Phylogenetic Handbook' (Lemey *et al.*2009) and the review by Yang and Rannala (2012). I will therefore only summarise the most important concepts (for my thesis) here from these three publications.
The basis of a molecular phylogenetic data set consists of sequenced genetic markers for a number of taxa. The selection of these markers can be based on different criteria: being single copy in the genome, having low (or high) sequence divergence between species, being of organelle origin (and therefore available in many copies). Each marker therefore has different properties, leading to different difficulties in the sequencing or assembling process and/or differences in the phylogenetic signal of the markers. Moreover, not all markers have an independent evolutionary history. All mitochondrial markers are by definition dependent on one another, because of the sole maternal inheritance and a transmission bottleneck in oogenesis. In this thesis I report examples of all these issues.
Molecular phylogeneticists nowadays follow either, or a combination of, the Maximum Parsimony (MP), Maximum Likelihood (ML) or Bayesian Inference (BI). Support for nodes in the first two is calculated by bootstrapping and in the latter it is expressed in posterior probability. These are conceptually different measures of support and should not be confused and only be compared with caution.
A phylogenetic reconstruction of a marker is called a gene tree. This is not the same as a species tree *per se*. Therefore most studies use more than one (the more the better) markers for phylogenetic reconstructions of species, thereby hoping that the recovered consensus tree can be used as a proxy for the species tree.
Many research questions in phylogenetics have an time-related component. For this purpose phylogenetic hypotheses can be time-calibrated by several methods after ultrimatrisation. A distinction is made between primary and secondary calibrations, where the former means that evidence from the fossil record or dated biogeographic events is directly included in the analysis. It is called a secondary calibration when an estimation of a node age is adopted from another time-calibrated study. A combination of these two types of calibrations can be applied.
Ultimately these methods are a tool in investigating evolutionary or ecological hypotheses and (for me) mostly not a goal in itself.

1.2. Phylogenetics of Lepidoptera

Lepidopteran molecular phylogenetics is a fairly new field with major taxonomic changes being proposed often (e.g. Wiegmann *et al*. 2002, Regier *et al*. 2009, Mutanen *et al*. 2010, Cho and Zwick *et al*. 2011, Regier *et al*. 2013). These changes led to a new official family-level classification (Nieukerken *et al*. 2011). Some superfamilies undergo revision, but the Noctuoidea remain stable at this level (when the Doidae are excluded) (Mutanen *et al*. 2010, Regier *et al*. 2013). One order-wide reconstruction by Mutanen *et al*. (2010) is shown in figure 1. In these new phylogenetic reconstructions the Noctuoidea are consistently placed in a highly derived position.

Some popular clades within the order are well studied, for example the Nymphalid phylogenetic relationships are studied by The Nymphalidae Systematics Group, led by Niklas Wahlberg (see http://nymphalidae.utu.fi/index.htm).

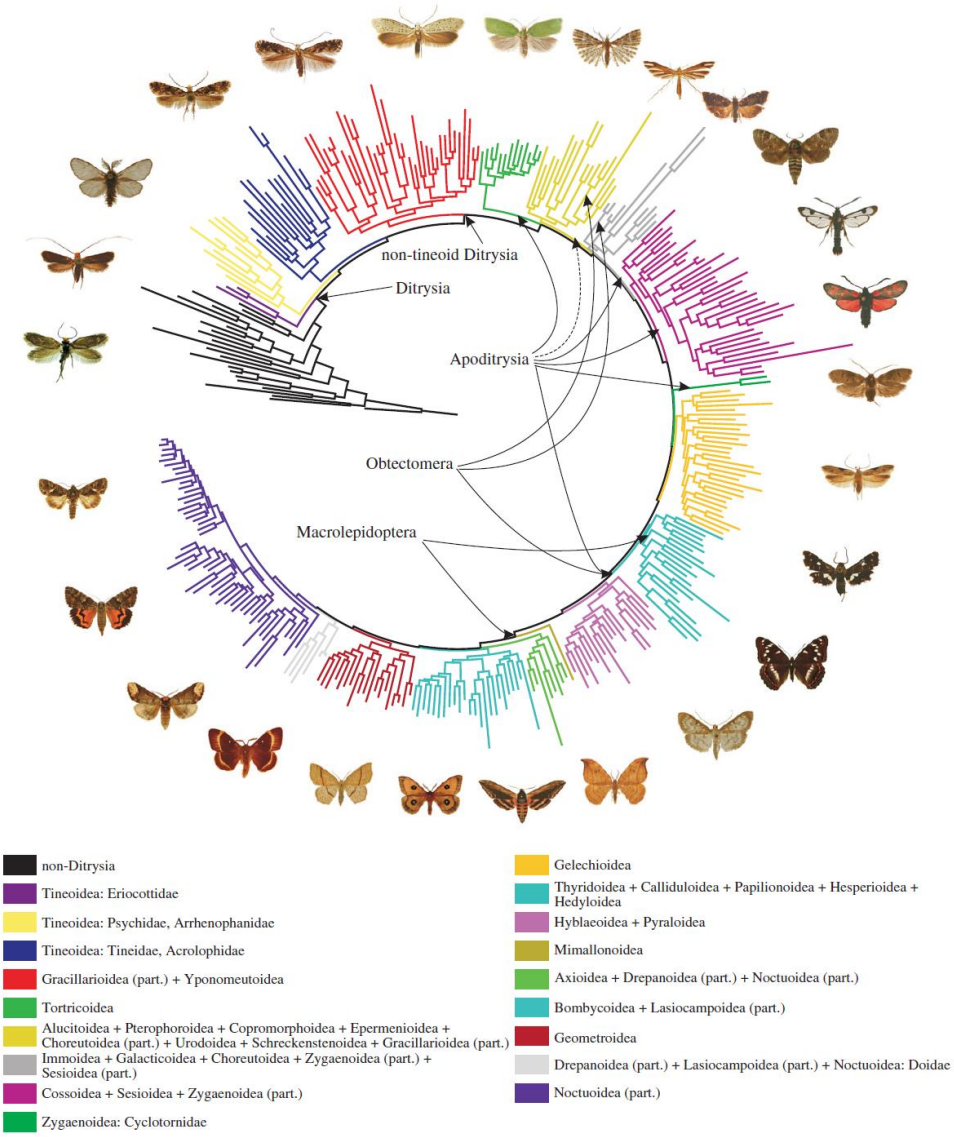


Figure 1. One of the latest order-wide Lepidoptera phylogenetic reconstructions. Analysis was based on RAxML maximum likelihood analysis of 6157 bp sequence data (with gaps) for 350 taxa, rooted on Micropterygidae. The Noctuoidea are tinted dark purple (from Mutanen *et al*. 2010).

1.2.1. Lepidoptera-specific challenges

Moths have their bodies and wings covered in scales, which easily detach. This has had a major influence for the chances they have in getting trapped in clay or resin (amber) and thus the chance of becoming a fossil. This led to an overall lack in fossil Lepidoptera, leading to much ambiguity in classical systematic morphological studies (Kitching *et al*. 1998, Kristensen *et al*. 2007). Sohn *et al*. (2012) published a catalogue of all known Lepidopteran fossils, containing a total of 667 records dealing with at least 4,568 specimens of 131 fossil genera and 72 extant genera. This lack of fossils also influences the number of available calibration points in the dating of phylogenetic reconstructions. That challenged me to find new ways of calibrating phylogenies, leading to a new concept I call 'ecological' dating (see below).

1.2.2. Available markers

Several markers are used for phylogenetic reconstructions in Lepidoptera: cytochrome oxidase subunit I gene (COI) and cytochrome B (CytB) from the mitochondrial genome (see figure 2 for the location of these genes on the mitochondrial genome) and several genes from the nuclear genome: Elongation Factor-1 alpha (EF-1α), wingless (WG), Ribosomal protein S2 (RpS2), Ribosomal protein S5 (RpS5), Carbamoyl-phosphate synthase domain protein (CAD), Cytosolic malate dehydrogenase (MDH), Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), Isocitrate dehydrogenase (IDH), malate dehydrogenase (MDH) and dopa decarboxylase (DDC) (Wahlberg and Wheat 2008). Another set of 19 suitable markers was published by Regier *et al*. (2012).
EF1α occurs in two copies in bees (Danforth and Ji 1998, Brady and Danforth 2004), beetles (Jordal 2002), flies (Hovemann *et al*. 1988) and some spiders (Hedin and Maddison 2001), but in Lepidoptera it is single copy (Wahlberg and Wheat 2008, http://nymphalidae.utu.fi/index.htm).
COI is the universal barcoding region used across all the different kingdoms of life (Hebert *et al*. 2003, Hebert *et al*. 2010, www.ibol.org). Especially barcoding in Lepidoptera has taken great flight, facilitated by the AllLeps-barcoding program: http://www.lepbarcoding.org/. COI is chosen because of its high phylogenetic signal on the species-level, thus for an optimal number of autapomorphies for the terminal taxa (Hebert *et al*. 2003).
In NCBI GenBank (Benson *et al*. 2008) and BOLD (Ratnasingham and Hebert 2007) several of these genes are available for *Spodoptera*, but no clear overview hereof is published. These genes all provide a different phylogenetic resolution at different taxonomic levels. Therefore not all are suitable for work on the species level, as I will do in the majority of this study (Wahlberg and Wheat 2008, Wilson 2010).

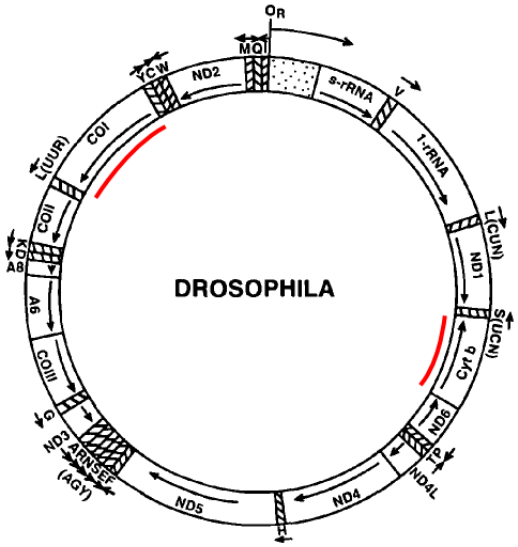


Figure 2. Overview of the insect mitochondrial genome.
The locations of COI and CytB are indicated with a red line (modified from Simon *et al*. 1994).

1.3. Phylogenetics of Noctuoidea

The monophyletic superfamily Noctuoidea, which currently holds 42,376 species in 3,771 genera (Nieukerken *et al*. 2011), recently experienced major revisions as well. In the published literature between five and thirteen families were recognised, based on morphologic (e.g. Poole 1995, Speidel *et al*. 1996, Kitching and Rawlings 1998) and molecular characters (Mitchell *et al*. 2006, Zahiri *et al*. 2011). The morphological studies were mainly based on the wing venation and the shape of the tympanal organs ('ears'). These characters are conserved, but nonetheless show marked differences between major clades (e.g. superfamilies).

Mitchell *et al*. (2006) already showed that wing venation does not correspond well to the molecular evidence of the phylogenetic relationships (using 2 nuclear protein-coding genes, total 2100 bp with gaps, maximum likelihood analysis, GTR+Γ+I substitution model). Furthermore, they concluded that the Noctuidae (as they were defined then) are paraphyletic as the morphological analyses already suggested. Mitchell *et al*. (2006) acknowledged the lack of support on many key nodes, but expected that improved taxon sampling will support their groupings. Nonetheless, they concluded that the subfamily-level within the Noctuidae is unresolved.
Cho *et al*. (2008) investigated the molecular phylogenetics (using 2635 bp of 71 species, MP&ML) of the heliothine moths (Noctuidae: Heliothinae), with *Spodoptera* as one of the outgroups (partly the same data as in Mitchell *et al*. 2006).
Recent molecular work by Zahiri *et al*. (2011, 2012) on this superfamily confirmed that the Noctuidae are not monophyletic and that the Erebidae should be split off (using one mitochondrial and 7 nuclear genes, total 6407 bp, maximum likelihood analysis, GTR+Γ+I model). For an extended overview of the morphological and molecular studies and the taxonomic history of the Noctuoidea, see Zahiri *et al*. (2012).
Lafontaine and Schmidt (2012) published a new classification of the Noctuoidea wherein they attempted to assimilate all the recent progress in the understanding of the relationships within the Noctuoidea.
As far as I know, no other molecular studies on the (sub)family level in the Noctuoidea are available. The current consensus is a classification of Noctuoidea in five families (Nieukerken *et al*. 2011, Zahiri *et al*. 2011, 2012, Lafontaine and Schmidt 2012), see figure 3. Herein the Noctuidae *sensu* Zahiri has 11,741 species in 1,088 genera (Nieukerken *et al*. 2011).

The position of *Spodoptera* was uncertain in the only molecular family-level study which included *Spodoptera* (Mitchell et al. 2006). In all performed analyses it was included in a 'pest clade', a name given by the authors to a part of the Noctuidae which includes many pest species (but many others as well). *Galgula*, a new world genus comprising 4 species (Poole 1989), was proposed to be the sister genus of *Spodoptera*, based on the maximum likelihood analysis (see above), but this had no bootstrap support. In the same analysis, *Elaphria* was proposed to be sister to these two genera, but again without bootstrap support. *Elaphria* is a cosmopolitan genus comprising 127 species (Poole 1989, Pogue and Sullivan 2003). Unfortunately the studies by Zahiri *et al*. (2011, 2012) did not include *Spodoptera*, thus were not helping in elucidating the place of *Spodoptera* in the Noctuidae. Nevertheless, *Spodoptera* is firmly placed in the family Noctuidae, based on morphological characters (Poole 1989, Pogue 2002, Lafontaine and Schmidt 2012).

Table 1. Total number of described genera and species, according to Nieukerken *et al*. (2011).

	Noctuoidea				**Total**
	Notodontidae	Noctuidae		Other Noctuoids	
		Spodoptera	non-*Spodoptera*		
Genera	704	1	1087	1979	**3771**
Species	3800	31	11741	26835	**42376**

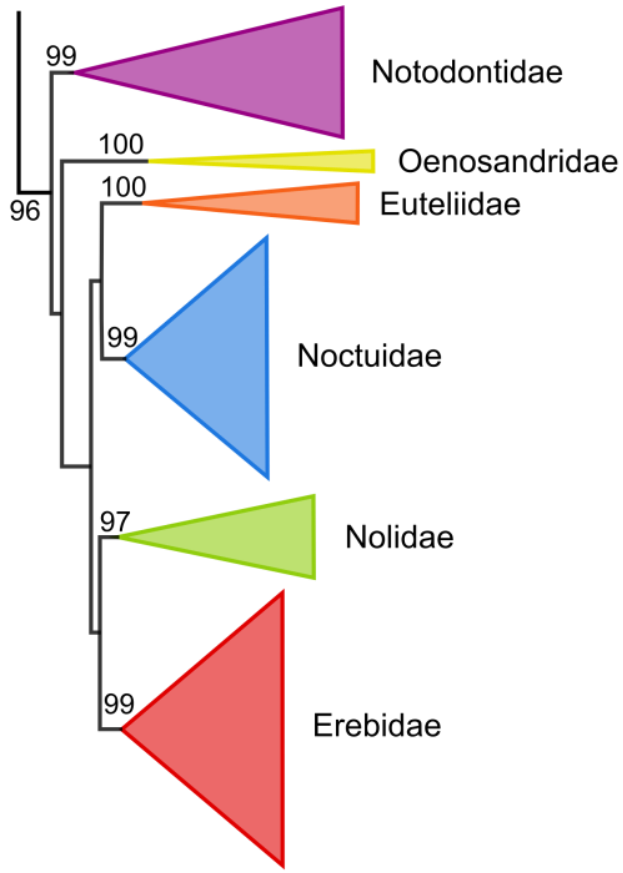


Figure 3. Current consensus of the classification of the Noctuoidea (modified from Zahiri *et al*. 2011)

1.4. Systematics of *Spodoptera*

1.4.1. Review of the non-phylogenetic literature

The genus *Spodoptera* consists currently of 31 valid species (Pogue 2002, 2011, appendix 1) and is of high economic importance, because of its invasive potential and half of them being described as crop pests. Being such an economically important genus, one might expect that much is known about it. Indeed, a lot is known about the control of the important pest species (e.g. Heppner 1998, Pogue 2002, Ellis 2004, Meagher *et al*. 2008), the species identification (e.g. Todd and Poole 1980, Passoa 1991, Heppner 1998, Pogue 2002) and their host preference (e.g. Heppner 1998, Tojo *et al*. 2008). Here I review this part of the literature on *Spodoptera*.

1.4.1.1. *Spodoptera* species

Published lists of valid names of *Spodoptera* species vary in length from 6 to 34 species (see table 2). Unfortunately, the type specimens of many good species are lost. Pogue (2002) solved this taxonomic incongruence in a major world revision of the genus, which still stands today as the monograph for the genus. He also described the history of the species nomenclature and explained the modification and revision thereof. He assigned many novel lectotypes and a neotype, increasing the verifiability of his statements. He argued that 30 species names are valid and provided descriptions for all, some of which were adjusted in a later publication (Pogue 2011). The phylogenetic part of the monograph was, however, solely based on morphologic characters; in contrast to all recent phylogenetic studies, which used at the least in part DNA-sequence data for phylogenetic reconstruction. In a later publication Pogue (2011) added one species, resulting in the current consensus of 31 *Spodoptera* species.
S. frugiperda has two different host strains, which might need to be elevated to the species level (Prowell *et al*. 2004). Hebert *et al*. (2010) described the situation as follows: "the fall armyworm, *Spodoptera frugiperda*, includes two barcode lineages with 1.3 per cent divergence […]. This species consists of two 'host races' that not only have different primary hosts (rice versus corn), but show allozyme and mitochondrial DNA divergence as well as reproductive isolation (Levy et al. 2002), justifying their recognition as distinct species." The host strains are, however, morphologically indistinguishable (Levy et al. 2002). Moreover, these so-called host strains co-exist in the same populations in South America (Juárez *et al*. 2012) and hybridize (Prowell *et al*. 2004). This case might lead to the discussion of species concepts and taxonomic decision circles, evaluated by DeSalle *et al*. (2005).

Table 2. Published numbers of overviews of the number of *Spodoptera* species.

Reference	# species
Guenée 1852	7
Hampson 1906	6
Poole 1989	34
Pogue 2002	30
Pogue 2011	31

1.4.1.2. *Spodoptera* morphological identification

Identification of adults of *Spodoptera* species is fairly straightforward for experts. Some external characters are useful (form of the antenna, length of the hairs on the prothoracic tibia, scaling of the spurs on the mesothoracic tibia), but often the genital structures need to be examined for reliable identification (e.g. Todd and Poole 1980, Passoa 1991, Heppner 1998, Pogue 2002, Van der Straten and Germain: EPPO diagnostic protocols, in publ.). This requires careful dissection of the specimens' abdomen and tedious chemical processes (Pogue 2002). For less experienced lepidopterologists al these steps are time consuming and often verification of the identification by an expert is necessary.
Species from the genus *Spodoptera* have several look-a-likes, either in the immature stage or the adult stage, in other related genera. Examples are *Neogalea sunia* Guenée, 1852, *Trichordesta prodeniformis* (Smith, 1888), *Copitarsia* Hampson, 1906 spp., *Antachara diminuta* (Guenée, 1852) and *Elaphria nucicolora* (Guenée, 1852) (pers. comm. van der Straten). In order to investigate their relationship with *Spodoptera* I also sampled as many as possible of these in my project.

Identification of the immature stages is however more difficult than of the adults. The fully-grown larvae of several species were described, mostly from America (Pogue 2002). However, it needs quite a bit of experience to understand the sometimes very subtle differences between species; experience that is rare amongst entomologists. Further the coloration changes dramatically during development of the larvae, and early instars are hardly described at all. Information on the differences in eggs and pupae of the different species is lacking completely. In the field and especially in international trade it is precisely the early stages which need to be identified. Molecular identification techniques can contribute to solving these problems (pers. comm. van der Straten).

Earlier I stated that morphological identifications are time-consuming and tedious. So why do we bother? In Lepidoptera taxonomy deciding on, and describing of, a new species is traditionally based on the morphology of the genitalia. External morphology differs between *Spodoptera* species as well, but I think it is likely to be constrained by camouflaging capabilities and flight performance of adults. Therefore I predict that we see much less and less distinct variation in external morphology between species. For this reason I did not treat many of those characters in my study. Nonetheless those characters are very important for scientists working in the field, or when a quick analysis of the risk of an intercepted species is necessary.
All information concerning the morphology of *Spodoptera* species is provided by Pogue (2002, 2011), including a matrix of 24 characters for all species. Table 3 lists the characters recognized by Pogue (2002). Van der Straten and Germain (EPPO diagnostic protocols, in publ.) provided a large matrix of many characters for many species, which is a modified version of the one created by Pogue (2002).
S. exigua and *S. hipparis* have a slightly different genital morphology: The apex of the male valve misses an indentation, which is present in all other *Spodoptera* species. This indentation splits the valve in the cucullus (dorsoapical part) and valvula (ventrodistal part). Pogue (2002 p.7) recognised this character, but did not draw conclusions on that. Therefore one might question the validity of calling these two species real *Spodoptera* species (pers. comm. Van der Straten).
This study did not go into the larval morphology, for several reasons. For ten of the species the larvae were not described yet (Heppner 1998, Pogue 2002). Furthermore, the morphology of the larvae varies per stage of development and not every form is available. Because of time constraints for this thesis and the described lack of data, it was not feasible to include larval morphology in this study. Also because of these time constraints no validated analyses on adult morphological data were performed in this project. The main problem in these analyses (cladistics) is character delimitation, which actually is an entire study on its own (Hennig 1966). Pogue (2002) published a morphological cladogram, with which I by visual inspection checked for the congruence with my molecular phylogenetic reconstruction.

Table 3. List of morphological characters used by Pogue (2002).
***depending on interpretation of the character**

Character number	External or genital	Character name	Data type
1	External	Form of male flagellum (antenna)	Multistate, nominal or ordinal*, 3 states
2	External	Length of the lateral scale tufts on the prothoracic tibia	Binary
3	External	Tibial spur scaling on mesothoracic tibia	Multistate, nominal or ordinal*, 5 states
4	Male genital	Width of the uncus	Multistate, nominal or ordinal*, 5 states
5	Male genital	Form and amount of sclerotization of the scaphium	Multistate, nominal or ordinal*, 3 states
6	Male genital	Lateral projections of the tegumen	Multistate, nominal or ordinal*, 3 states
7	Male genital	Shape of the costal process	Multistate, nominal or ordinal*, 9 states
8	Male genital	Location of the costal process on the costa	Multistate, nominal or ordinal*, 4 states
9	Male genital	Form of the cucullus	Multistate, nominal or ordinal*, 6 states
10	Male genital	Form of the ampulla	Multistate, nominal or ordinal*, 10 states
11	Male genital	Form of the basal sclerite of the clasper	Multistate, nominal or ordinal*, 8 states
12	Male genital	Shape of the clavus	Multistate, nominal or ordinal*, 6 states
13	Male genital	Shape of the sacculus	Multistate, nominal or ordinal*, 11 states
14	Male genital	Shape of the ventral margin of the valvula	Binary
15	Male genital	Relative width of the valvula	Binary
16	Male genital	Presence of the valvular indentation	Binary
17	Male genital	Number of lobes on the coremata	Multistate, nominal or ordinal*, 4 states
18	Male genital	Shape of the juxta	Multistate, nominal or ordinal*, 10 states
19	Male genital	Dorsobasal sclerotized patch on the vesica	Multistate, nominal or ordinal*, 9 states
20	Male genital	Distal cornutus on the vesica	Multistate, nominal or ordinal*, 11 states
21	Male genital	Presence of a dense cornutal patch on the vesica	Binary
22	Male genital	Location of the dense cornutal patch on the vesica	Multistate, nominal or ordinal*, 4 states
23	Female genital	Form of the distal margin on the ventral plate of the ostium bursa	Multistate, nominal or ordinal*, 9 states
24	Female genital	Presence of the ventrolateral invaginated pocket of the 8th sternite, lateral to the ventral plate of the ostium bursa	Binary

1.4.1.3. Biogeography (species distribution) and host specificity/polyphagy,

Spodoptera is a cosmopolitan genus. Data on the distribution of all species is available, but not complete (Pogue 2002, appendix 1). But, given all available data, no one had ever published a historic biogeographic analysis on the genus at the start of my project, so there was no theory on the origin and spread of this agriculturally important genus. This can be important for, for instance, understanding the origin of pest species, given that many crops have been shifted around the globe in the last centuries. It might be that *Spodoptera* species shifted their host preference to agricultural crops recently, but we don't know that until we have a dated phylogenetic reconstruction. Human agriculture emerged approximately 40,000 years ago in the Fertile Crescent, while speciation usually is a much slower evolutionary process (Wilson 2010).

Some *Spodoptera* species are specialist on one or few host plants, for example *S. picta* feeds only on Amaryllidaceae and *S. pectinicornis* is monophagous on *Pistia stratiotes* L. (Water lettuce). All other species, of which we know the host plant, are more or less polyphagous (Heppner 1998 for Nearctic and northern Neotropical species; Pogue 2002 for all species). Differences in host preferences within a species are also published (e.g. Tojo *et al*. 2008 for *S. litura*). Of 10 species the host plants are unknown (Pogue 2002).

It is generally assumed that specialist herbivorous insects evolve from generalists (Schoonhoven *et al*. 2005). In most insect clades the majority of species is specialized, not generalist (Janz *et al*. 2001). Generalists are thought to be in a transient situation towards specialism or consisting of populations with different host preferences (Nylin and Janz 2009). Specialized lines are thought to differentiate faster (Nosil 2002, Fordyce 2010). In *Spodoptera* most species are polyphagous, which is one of the reasons why some of them are of quarantine concern: the more host plants, the higher the chance of establishment in a new area. From this point of view this genus is an interesting case.
As in morphologic studies, character delimitation is also an issue here. Host preference can be coded by family, but some polyphagous species have up to 80 recorded host families, while others are poorly known (Pogue 2002). Host specificity can also be coded, for example as monophagous – oligophagous – polyphagous. But then there is still the problem of missing data and the definitions of these phenomena. Due to time constraints in this thesis I did not perform any validated host optimizations on my phylogenetic reconstructions.

1.4.1.4. Agriculture and monitoring

About half of the *Spodoptera* species have a high invasive potential (my judgement, based on this thesis) and have been considered crop pest species (Pogue 2002). Some characteristics of these pest species are: polyphagy (Pogue 2002), good dispersal capabilities (e.g. Saito 2000) and a short generation time (Wilson 1932, Merhkhou *et al*. 2012). For example, *Spodoptera exigua* can have a very short generation time (20 days) and high fecundity (up to 1200 eggs per female) (Wilson 1932) and has many recorded host families (Pogue 2002). When an invasive species becomes established, drastic measures are taken (e.g. Ellis 2004 for the policy in Florida) and continuous monitoring programs (e.g. with pheromone traps) are executed (Meagher *et al*. 2008). In the Netherlands the Plant Protection Service is responsible for the regulation of, and inspection for, EU quarantine-list species. Four *Spodoptera* species are listed on the European quarantine list (http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2000:169:0001:0112:EN:PDF), and the EPPO (http://www.eppo.org/) list, because of the threat they can pose to agriculture and greenhouse crops. These are *S. littoralis* from Africa and Southern Europe, *S. litura* from Asia and Oceania, and *S. frugiperda* and *S. eridania* from America. From these species *S. littoralis* is being intercepted in the EU most frequently, followed by *S. litura* (van der Straten and Bakker 2011). More importantly: nations might place import bans upon countries in which these species have been established, damaging the export of the infested country.

Many plant extracts are being tested for their toxicity and repellence to insects, especially on lab strains of *Spodoptera* spp. One of the tested plants, among many others, is the extracts from *Pelargonium* x *hortorum*, which is tested on *Spodoptera littoralis* (Farag *et al*. 2012).

1.4.2. Phylogenetics of *Spodoptera*

The first cladistic analysis of *Spodoptera* was published by Pogue (2002). He based his analysis on 24 morphological characters (see section1.4.1.2 above) and included the outgroups *Elaphria* and *Galgula*. This yielded 10 equally parsimonious trees, whereof I manually made a strict consensus tree (figure 4). Nagoshi *et al*. (2011) published a COI DNA-barcode neighbour-joining (NJ) tree with bootstrap support values of only seven *Spodoptera* species. Pogue (2011) used a COI-based phylogenetic reconstruction (MP and ML methods) of ten *Spodoptera* species to support his morphology-based hypotheses, that *Leucochlaena hipparis* should be included in *Spodoptera*. However, neither of these publications discussed their phylogenetic hypotheses in detail, nor did they focus on within-*Spodoptera* phylogenetic patterns. Also their analyses might be compromised by their limited taxon sampling.
For the goal of developing a molecular identification test for the four EU quarantine Spodoptera species, Bakker, van der Straten and van de Vossenberg (in prep.) produced a phylogenetic reconstruction based on two mitochondrial genes (COI & cytB).
Later during my project we learned that Kergoat *et al*. (2012) was in the process of publishing a paper on the phylogenetics of *Spodoptera*. The details and consequences of this publication I discuss later in this report.

So far everyone assumed that the genus *Spodoptera* is monophyletic. To our knowledge, no one doubted this, but formal testing has not been published either. The place of *Spodoptera* in the Noctuidae and its monophyly was tested in this research.

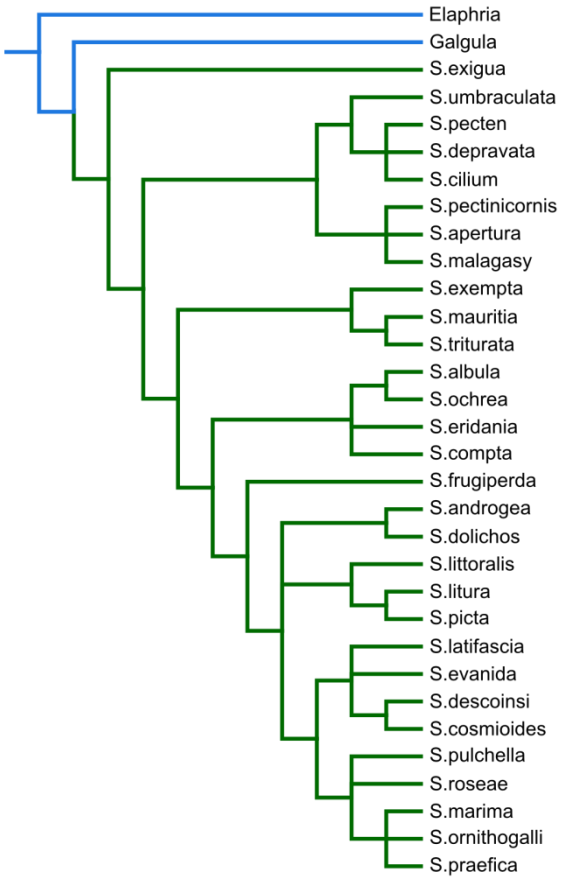


Figure 4. The strict consensus tree created from 10 equally parsimonious trees from Pogue 2002.

1.5. Aim, research questions and hypotheses

In this study I aimed for a better understanding of the phylogenetic relationships within *Spodoptera*. Given all the data which was available to me and the progress in phylogenetic modelling, this was fairly straightforward. Since the current distribution of the *Spodoptera* species is relatively well known, a reconstruction of the historical biogeography was possible. The morphology of adults in this genus is well described. I planned to use this available data to identify morphological synapomorphies for clades within *Spodoptera*.
I formulated several research goals at the start of this project. (1) We wanted to know how a well-build species-level phylogenetic reconstruction of *Spodoptera* looks like and what distinct clades there are within *Spodoptera*. Furthermore, (2) we wanted to know the position of *Spodoptera* within the Noctuidae. Next, (3) we wanted to know the age of this important pest genus and its clades.
Each research question required a different data set. I will describe these requirements in the material and methods section.

I formulated the following hypotheses, which are tested in this MSc thesis:

1) *Galgula* is sister to *Spodoptera*;
2) *S. exigua* and *S. hipparis* are not *Spodoptera*, based on DNA-sequence divergence;
3) The other *Spodoptera* species are monophyletic;
4) *Spodoptera* originated in Asia;
5) The ancestor of *Spodoptera* was polyphagous;
6) Speciation in *Spodoptera* was directed by host plant shifts.

2. Material and methods

2.1. Data collection

2.1.1. Data set from FES Phytosanity

The previous version of the *Spodoptera* data set was generated for the FES Phytosanity 2.2 project (van der Straten and Bakker 2011; van de Vosseberg and van der Straten in prep.). It consisted of a total of 72 specimens from 20 different *Spodoptera* species and a total of 190 sequences for COI, cytB and EF1α, see table 4. The phylogenetic reconstruction from Straten and Bakker (2011) is shown in figure 5.
New information on the identity of the specimens (as determined by Marja van der Straten, pers. comm.) and curating (new insights in data quality standards) of the data set by Freek Bakker (pers. comm.) and by myself, led to a different overview of this dataset. See table 5 for the data set composition as it was in retrospect at the start of this project.
Some of the sequences still had their primer sequences attached to them, I removed those before analysis.

Not all 31 described *Spodoptera* species are in this data set, 13 are missing. These species were the major part of the original wish list as placed in the proposal (see appendix 2).
These species were not available or specimens were too old in the Dutch collections and other collections contacted abroad (van der Straten and Bakker 2011). The missing species are listed in table 6 with some extra remarks on their availability. The specimens listed as present in the MNHN are added based on Kergoat *et al*. (2012), some of which are sequenced and uploaded to GenBank at the end of my project. These were therefore unavailable for my (already finished) analyses.

Table 4. Details of the starting data set.
(COI and EF1α from van der Straten and Bakker 2011; CytB from van de Vosseberg and van der Straten in prep.).

Species	Number of specimens	COI-sequences		EF1α-sequences		CytB-sequences
		generated	GenBank	generated	GenBank	
S. albula	2	2		1		2
S. androgea			8			
S. cilium	3					1
S. dolichos	2	2	10	1		5
S. eridania	3					2
S. exempta	1	4		1		1
S. exigua	10	5	6		1	5
S. frugiperda	3	3	27		1	4
S. latifascia	6	4	8			6
S. littoralis	12	8	2		1	12
S. litura	6	6	4			7
S. malagasy	1					
S. mauritia	8	5				4
S. ochrea	2	2				1
S. ornithogalli	1		5		1	
S. pecten	3	1		1		2
S. pectinicornis	1					
S. picta	2	1				
S. pulchella	1		2			1
S. triturata	5					1
Total: 20 species	72	51	81	4	4	54

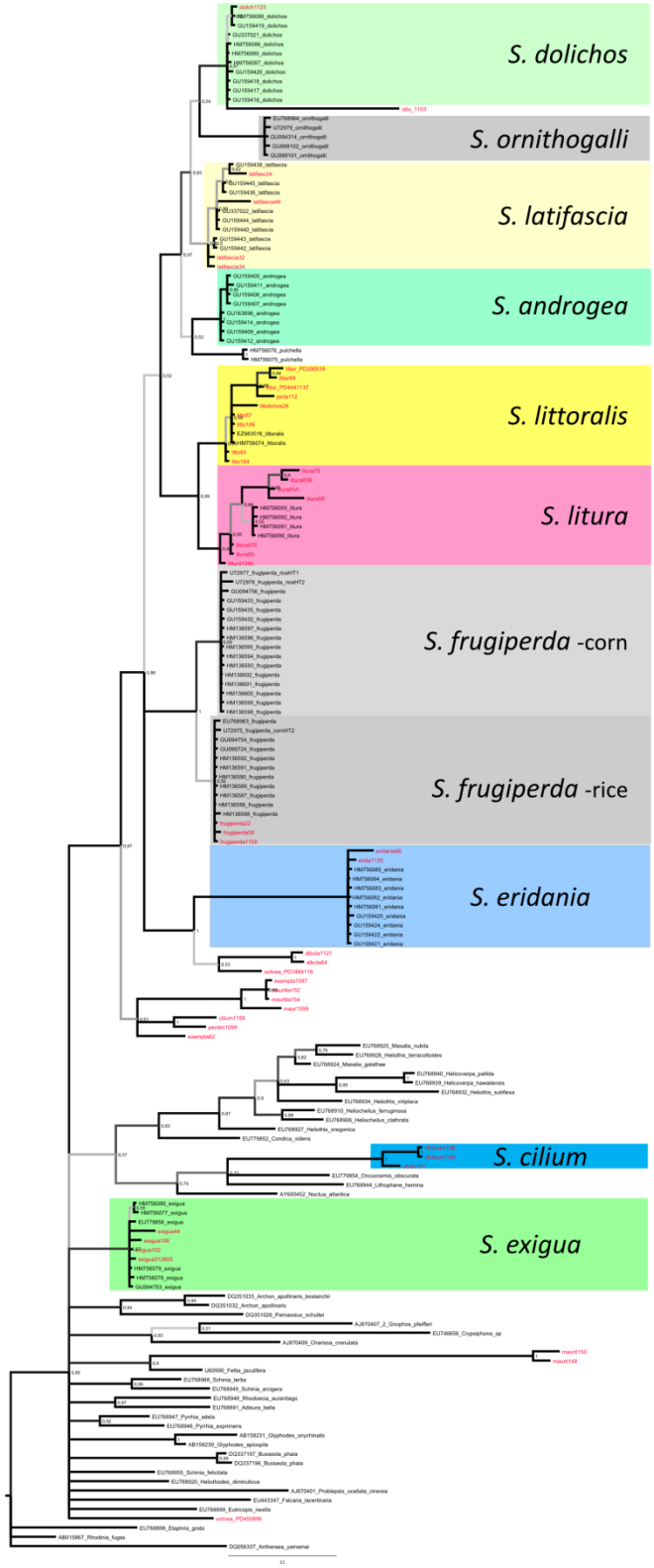


Figure 5. Phylogenetic reconstruction of Straten and Bakker 2011.

Table 5. Overview of the contents in retrospect of the starting data set.
by van der Straten and Bakker 2011; van de Vosseberg and van der Straten in prep.). The GenBank sequences are not reported again.

Species	Number of sequenced specimens	COI-sequences	CytB-sequences	EF1α-sequences
S. albula	4	2	3	1
S. cilium	1	1	1	0
S. dolichos	6	1	5	0
S. eridania	3	2	2	0
S. exempta	1	1	1	0
S. exigua	7	6	3	0
S. frugiperda	6	5	5	0
S. latifascia	7	4	6	0
S. littoralis	13	11	13	1
S. litura	7	7	7	0
S. malagasy	1	1	0	0
S. mauritia	7	7	4	1
S. ochrea	1	1	1	0
S. pecten	1	1	1	1
S. pectinicornis	0	0	0	0
S. picta	1	1	0	0
S. pulchella	3	1	3	0
S. triturata	1	1	1	0
non-Spodoptera	5	5	1	0
Total: 18+x species	75	58	57	4

Table 6. Missing species in the starting data set and remarks on their availability.
MNHN = Muséum national d'histoire naturelle in Paris, France.

Species	Availability
S. apertura	2 old specimens in MNHN
S. compta	no published recent observations of this species
S. cosmioides	7 specimens in MNHN
S. depravata	no published recent collections
S. descoinsi	7 specimens in MNHN
S. evanida	2 specimens in MNHN
S. hipparis	in Pogue 2011
S. marima	2 specimens in MNHN
S. praefica	in Pogue 2011, 4 specimens in MNHN
S. roseae	no published recent observations of this species
S. umbraculata	no published recent collections
Total: 11 species	

2.1.2. Overview of available sequence data in GenBank and BOLD

An overview of the available protein-coding DNA-sequences of the Noctuoidea and several subgroups is listed in table 7, compiled from NCBI GenBank (Benson 2008) on June 26th, 2012, based on name searches (basically: 'gene name' AND 'taxon name'). The reader should keep in mind that this table also includes all misidentified specimens and 'bad' data of GenBank. COI is by far the most available, but the general pattern visible is a low coverage of available data per species (table 7). The data by Kergoat *et al*. (2012) were clearly not uploaded (or not released) yet.
Several nuclear markers had a good availability, mostly uploaded by Zahiri *et al*. (2011, 2012). This made them suitable markers for the extension of my data set across the Noctuoidea. This was necessary for the time-calibration of the phylogenetic reconstruction (see below). I therefore sequenced these markers (CAD, GAPDH, IDH) also from some *Spodoptera* specimens (see below).

Table 7. Available Noctuoid sequences in GenBank on June 26th, 2012.

	Noctuoidea				
	Notodontidae	Noctuidae (incl. Erebidae *partim*)		Other	
Genes		*Spodoptera*	non-*Spodoptera*		Total
COI	11475	408	25982	11130	48587
cytB	3	1	773	15	791
EF1α	26	12	752	128	906
CAD	12	2	186	37	235
GAPDH	7	2	267	28	302
IDH	10	0	206	34	250
other	143	5174	21621	5360	27124
Total	**11686**	**5660**	**50033**	**16732**	**84111**

2.2. Lab protocols

DNA from the selected specimens was isolated using standardized protocols. Either one or two legs were used in the DNA extraction, depending on the size of the species legs, or part of the abdomen when legs were missing. The Qiagen DNeasy kit (Hilden, Germany) was used for extraction, following the manufacturers' instructions. Extracted DNA is then ready for downstream amplification of the gene of interest. See table 8 for the primer details.
The target 658-bp fragment of COI turned out in a test on April 4th 2012 to be amplified better by using primer pair LepF and LepR than using the universal primer pair HCO and LCO. Therefore all amplifications of COI were done with the pair LepF and LepR. Where amplification of COI failed and no PCR product was obtained, two internal primers were added, using primer pairs LepF with mLR and mLF with LepR (Hajibabaei *et al*. 2006a). These primer pairs yield fragments of 311 and 407 bp respectively, which afterwards can be assembled to the full 658-bp sequence. The full EF1α needs to be sequenced in 3 parts. I only sequenced the first part using primers HybStarsky and HybMonicaR. The downloaded sequences from GenBank are usually longer.
CAD is best done in two pieces using CAD743nF/CADmidR and CADmidF/CAD1028R, each giving ca 450 bp (Wahlberg and Wheat 2008). The full CAD-sequence is 849 bp. I only amplified the second half of CAD using primers CADmidF and CAD1028R.
Several primers have a 'hybrid tail' attached to them; see Regier and Shi (2005) for details and explanations of these. PCR was performed in 10 µL, see table 9 for components of the reaction mix. Each PCR was performed in quintuple (5 times 10 µL), in order to produce enough amplicon for the sequencing reaction. These amplicons were pooled into one tube for the purification step. PCR products were purified using GeneJet PCR purification kit (Thermo Fisher Scientific, Vilnius, Lithuania) following protocols provided by the manufacturer and eluted in 20 µL elution buffer. Of the purified fragments both strands were sequenced directly using a fluorescent dye-labelled sequencing reaction (DYEnamic™ ET Terminator Cycle Sequencing Kit; Amersham Biosciences, see table 10 for the components of the reaction mix) and the high-throughput ABI sequencing facilities at Greenomics™, Wageningen (the Greenomics sequence service was terminated at March 14th, 2013). Returned sequence tracers were analysed and assembled in CodonCode Aligner software v.3.7.1 (CodonCode Corporation, available online at http://www.codoncode.com/). The assembled contigs were added to the data set.

Table 8. Overview of all primers used in this study.
Primers with a hybrid tail attached have the tail printed in bold.

Gene region	Fragment length	Direction	Primer	Primer sequence	Annealing temperature	Reference
COI	648 bp	Forward	LepF	5'-attcaaccaatcataaagatattgg-3'	47 °C	Hebert *et al.* 2004
		Reverse	LepR	5'-taaacttctggatgtccaaaaaatca-3'	47 °C	Hebert *et al.* 2004
	710 bp	Forward	HCO	5'-taaacttcagggtgaccaaaaaatca-3'	47 °C	Folmer *et al.* 1994
		Reverse	LCO	5'-ggtcaacaaatcataaagatattgg-3'	47 °C	Folmer *et al.* 1994
	see text	Forward	mLR	5'-gctttcccacgaataaataata-3'	47 °C	Hajibabaei *et al.* 2006a
		Reverse	mLF	5'- cctgttccagctccattttc-3'	47 °C	Hajibabaei *et al.* 2006a
CytB	609 bp	Forward	CB-J-10933	5'-tatgtactaccatgaggacaaatatc-3'	55 °C	Simon *et al.* 1994
		Reverse	CB-J-11545	5'-acatgaattggagctgaccagt-3'	55 °C	Simon *et al.* 1994
EF1α	541 bp	Forward	HybStarsky	5'-**taatacgactcactataggg**cacatyaacattgtcgtsatygg-3'	56 °C	Wahlberg and Wheat 2008
		Reverse	HybMonicaR	5'-**attaaccctcactaaag**catrttgtckccgtgccarcc-3'	56 °C	Wahlberg and Wheat 2008
CAD	849 bp	Forward	CAD743nF	5'-**taatacgactcactataggggg**ngntnacnacngcntgyttygarcc-3'	55 °C	Wahlberg and Wheat 2008
		Reverse	CAD1028R	5'-**attaaccctcactaaag**ttrttnggnarytgnccncccat-3'	55 °C	Wahlberg and Wheat 2008
		Forward	CADmidF	5'-**taatacgactcactataggg**kggattytcngayaaacaaatngc-3'	55 °C	Wahlberg and Wheat 2008
		Reverse	CADmidR	5'-**attaaccctcactaaag**cattcwgckgcwactgtatc-3'	55 °C	Wahlberg and Wheat 2008
GAPDH	690 bp	Forward	HybFrigga	5'-**taatacgactcactataggg**aargctggrgctgaatatgt-3'	55 °C	Wahlberg and Wheat 2008
		Reverse	HybBurre	5'-**attaaccctcactaaag**gwttgaatgtacttgatragrtc-3'	55 °C	Wahlberg and Wheat 2008
IDH	713 bp	Forward	IDHdeg27F	5'-**taatacgactcactataggggg**swgaygaratgacnagrathathtgg-3'	55 °C	Wahlberg and Wheat 2008
		Reverse	IDHdegR	5'-**attaaccctcactaaag**ttyttrcaigcccanacraanccncc-3'	55 °C	Wahlberg and Wheat 2008

Table 9. Components of the PCR-mix.

Reagent	Concentration (if applicable)	Amount
10x buffer		1 µL
dNTPs	10 µM/µL	0.4 µL
BSA	10 µM/µL	1 µL
Primer Forward	10 µM/µL	0.35 µL
Primer Reverse	10 µM/µL	0.35 µL
mQ Water		6.82 µL minus amount of template
Taq Polymerase	5U/µL	0.08 µL
DNA Template		1 µL or 2 µL, depending on DNA-concentration
Total		**10 µL**

Table 10. Components of the cyclesequence reaction mix.

Reagent	Concentration (if applicable)	Amount
dilution buffer		2 µL
Big Dye RRM		2 µL
Primer	10 µM/µL	0.5 µL
mQ Water		5.5 µL minus amount of purified amplicon
Purified amplicon		1 to 5.5 µL, depending in the strength of the PCR-product on gel
Total		**10 µL**

2.3. Data set

The starting data set was provided by Freek Bakker (313SpodFusedB). The first step was to remove all data I knew I wasn't going to use, basically all non-*Spodoptera* (86 sequences). Next, I used the program RAxML (see section 2.6 'Phylogenetic analyses') in order to remove 43 identical sequences. In this step dolichos1155 and mauritia156 were also unintentionally removed. Manually removed from the data set were also: 'cf_ochrea/exigua_PD450896' because the voucher is an unidentified caterpillar (blasting the sequence suggests a close relationship with species from the genus *Callopistria*); exigua3126055 and exigua3738297 because the voucher was not verifiable (their abdomens are missing); nonSpodoptera1106, nonSpodoptera1105 and nonSpodoptera1101 because the vouchers are unidentified (non-*Spodoptera*) species.

In order to reconstruct the Noctuoid phylogeny and assess the monophyly of *Spodoptera*, part of the data sets of Zahiri *et al*. (2011: 151 sequences of 5 genes of 32 species; 2012: 70 sequences of 5 genes of 16 species), Mitchell *et al*. (2006: 3 sequences of EF1α of 3 species) and Cho *et al*. (2008: 18 sequences of 2 genes of 9 species) were downloaded from GenBank (see table 11). The data of Zahiri *et al*. 2012 was released well after I analysed the COI and EF1α alignments. Therefore these taxa are not included in all results of these one-gene analyses.
On request, Michael G. Pogue sent me 9 COI-sequences of 3 species which were generated for his study on the taxonomy of *Leucochlaena hipparis* (Pogue 2011). Two CytB sequences were downloaded from GenBank in order to have an outgroup in the analysis of the one-gene CytB analysis (see sections 2.4 'Alignments' and 2.6 'Phylogenetic analyses'). Lastly, as many *Elaphria* sequences as possible (14 COI sequences) were downloaded from GenBank, because of their proposed sister status to *Spodoptera*+*Galgula*.

In the research proposal I estimated to be able to generate 20 sequences in the time given. In total I generated 37 new sequences, of which 18 of COI, 7 of EF1α, 6 of CAD and 6 of GAPDH. 5 sequences were reassembled, in 4 cases after re-sequencing of the same specimens; see section 2.3.1 'Re-sequencing and (re-)assembly' below. Attempts to amplify IDH failed for all tested specimens; see section 2.3.3 'IDH problems' below. Other failed amplification/sequencing attempts were Diar1204 (*Diarsia* sp., from Kenya) and Anic1261 (*Anicla infestans*, from USA), which failed for all attempted genes, so they are therefore not in the data set. I also tried to sequence other genes of exigua102, but this failed, probably because the specimen was too old.
This resulted in a total data set of 511 sequences from 280 specimens (including all 6 specimens of which we failed to generate any sequence) and 100 species. For an overview of the total data set, see table 11.

Table 11. All used specimens in this study, grouped by the analyses where they are used in.
ML = sequence generated by Mark Lammers; COI = COI sequence was provided in the starting data set; ½COI = COI sequence is only half the length of a full barcode; EF1a = EF1α sequence provided in the starting data set; cytB = CytB sequence provided in the starting data set; ML (in bold) = this sequence is newly generated by me; ML (in italics) = this sequence is assembled by me from already existing sequence tracers; ML (in bold and italics) = I resequenced and re-assembled this specimen.

Group	Reference	voucher	Genus	epithet	COI	EF1a -1	EF1a -2	CAD	GAPDH	IDH	cytB
CA 1 & 2	Zahiri *et al.* 2011	MM01529	*Acronicta*	*rumicis*	GU828666	GU828997	GU829280	GU828163	GU829792	GU830053	-
	Zahiri *et al.* 2011	MM06745	*Craniophora*	*ligustri*	HQ006148	HQ006246	HQ006341	HQ006948	HQ006432	HQ006498	-
	Zahiri *et al.* 2011	MM00005	*Panemeria*	*tenebrata*	HQ006157	HQ006254	HQ006349	HQ006956	HQ006437	HQ006506	-
	Zahiri *et al.* 2011	MM01542	*Brachionycha*	*nubeculosa*	GU828667	GU828998	GU829281	GU828164	GU829793	GU830054	-
	Zahiri *et al.* 2011	MM01162	*Amphipyra*	*perflua*	GU828660	GU828991	GU829275	GU828157	GU829787	GU830047	-
	Zahiri *et al.* 2011	MM07669	*Periscepta*	*polysticta*	GU828820	GU829125	GU829400	GU828289	GU829892	GU830201	-
	Zahiri *et al.* 2011	MM05114	*Pyrrhia*	*umbra*	GU828712	GU829034	GU829324	GU828200	GU829825	GU830098	-
	Zahiri *et al.* 2011	MM04919	*Cryphia*	*raptricula*	GU828708	GU829031	GU829320	GU828196	GU829822	GU830094	-
	Zahiri *et al.* 2011	MF-05-0053	*Diaphone*	*capillamentum*	GU828571	GU828913	GU829206	GU828076	GU829738	GU829960	-
	Zahiri *et al.* 2011	MM01651	*Hoplodrina*	*octogenaria*	HQ006153	HQ006251	HQ006346	HQ006953	HQ006434	HQ006503	-
	Zahiri *et al.* 2011	MM05153	*Actinotia*	*polyodon*	GU828714	-	GU829326	GU828202	GU829827	GU830100	-
	Zahiri *et al.* 2011	MM04752	*Noctua*	*fimbriata*	GU828705	GU829028	GU829028	GU828194	GU829820	GU830091	-
	Zahiri *et al.* 2011	MM01170	*Apamea*	*crenata*	GU828661	GU828992	GU829276	GU828158	GU829788	GU830048	-
	Zahiri *et al.* 2011	RR-98-0914	*Ufeus*	*faunus*	GU828860	GU829163	GU829425	GU828320	GU829911	GU830238	-
	Zahiri *et al.* 2011	MM00328	*Autographa*	*gamma*	GU828636	HQ006502	GU829256	GU828135	-	GU830023	-
	Zahiri *et al.* 2011	CWM-95-0471	*Condica*	*vecors*	GU828550	GU828895	GU829194	GU828061	-	GU829941	-
	Zahiri *et al.* 2011	RZ25	*Ecpathia*	*longiqua*	HQ006190	HQ006286	HQ006380	HQ006985	-	HQ006532	-
	Zahiri *et al.* 2012	RZ341	*Hemicephalis*	*alesa*	JN401251	JN401368	JN401479	JN401049	JN401581	JN401682	-
	Zahiri *et al.* 2012	RZ277	*Aedia*	*leucomelas*	JN401250	JN401367	JN401478	-	JN401580	JN401681	-
	This study	doli1123	*Spodoptera*	*dolichos*	**ML**	**ML**	-	**ML**	**ML**	-	cytB
	This study	exig1160	*Spodoptera*	*exigua*	**ML**	-	-	**ML**	**ML**	-	cytB
	This study	Eagr1270	*Elaphria*	*agrotina*	**ML**	**ML**	-	**ML**	**ML**	-	-
	This study	frugi1267	*Spodoptera*	*frugiperda*	**ML**	**ML**	-	**ML**	**ML**	-	-
	This study	litt1268	*Spodoptera*	*littoralis*	**ML**	**ML**	-	**ML**	**ML**	-	-
	This study	litu1262	*Spodoptera*	*litura*	**ML**	**ML**	-	**ML**	**ML**	-	-
	Cho *et al.* 2008	see_GenBank	*Elaphria*	*grata*	EU768898	U85697	U85697	-	-	-	-
	Cho *et al.* 2008	see_GenBank	*Galgula*	*partita*	AF549719	AF151626	AF151626	-	-	-	-
	Cho *et al.* 2008	S.ex	*Spodoptera*	*exigua*	EU779856	AF151624	AF151624	-	-	-	-
	Cho *et al.* 2008	Sfr	*Spodoptera*	*frugiperda*	U72976/ U72975	U20139	U20139	-	-	-	-
	Cho *et al.* 2008	Sor	*Spodoptera*	*ornithogalli*	EU768964	AF151623	AF151623	-	-	-	-
	Cho *et al.* 2008	Ham	*Helicoverpa*	*armigera*	EU768935	U20129	U20129	-	-	-	-
	GenBank	"Reg"	*Spodoptera*	*frugiperda*	-	-	-	EU032751	-	-	-
CA 1 only	This study	pecten1098	*Spodoptera*	*pecten*	COI	EF1a	-	-	-	-	cytB
	This study	mauritia1099	*Spodoptera*	*mauritia*	COI	EF1a	-	-	-	-	cytB
	This study	albula1121	*Spodoptera*	*albula*	COI	EF1a	-	-	-	-	cytB
	This study	frugiperda1216	*Spodoptera*	*frugiperda*	COI	-	-	-	-	-	cytB
	This study	eridania1135	*Spodoptera*	*eridania*	COI	-	-	-	-	-	cytB
	This study	pulchella1133	*Spodoptera*	*pulchella*	COI	-	-	-	-	-	cytB
	This study	cilium1156	*Spodoptera*	*cilium*	COI	-	-	-	-	-	cytB
	This study	exempta62	*Spodoptera*	*exempta*	COI	-	-	-	-	-	cytB
	This study	ochrea60	*Spodoptera*	*ochrea*	½COI	-	-	-	-	-	cytB
	This study	latifascia32	*Spodoptera*	*latifascia*	½COI	-	-	-	-	-	cytB

	This study	litura1096	*Spodoptera*	*litura*	COI	-	-	-	-	-	cytB
	This study	triturata1152	*Spodoptera*	*triturata*	COI	-	-	-	-	-	cytB
	This study	picta112	*Spodoptera*	*picta*	***ML***	**ML**	-	-	-	-	-
	Pogue 2011	S_praefica1	*Spodoptera*	*praefica*	COI	-	-	-	-	-	-
	Pogue 2011	L_hipparis1	*Spodoptera*	*hipparis*	COI	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*androgea*	GU159412	-	-	-	-	-	-
	This study	malagasy1103	*Spodoptera*	*malagasy*	COI	-	-	-	-	-	-
CA 2 only	Zahiri *et al*. 2011	MM00027	*Thyatira*	*batis*	GU828580	GU828919	GU829212	GU828083	GU829743	GU829969	-
	Zahiri *et al*. 2011	NW141-12	*Sphinx*	*ligustri*	EU141358	EU136665	EU136665	EU141313	EU141494	EU141550	-
	Zahiri *et al*. 2011	NW149-1	*Bombyx*	*mori*	EU141360	EU136667	EU136667	EU141315	EU141495	EU141552	-
	Zahiri *et al*. 2011	MM01005	*Clostera*	*pigra*	GU828654	GU828985	GU829269	GU828151	GU829782	GU830041	-
	Zahiri *et al*. 2011	MM00998	*Notodonta*	*dromedarius*	GU828653	GU828984	GU829268	GU828150	GU82978	GU830040	-
	Zahiri *et al*. 2011	RZ10	*Dryops*	*chromatophila*	HQ006158	HQ006255	HQ006350	HQ006957	HQ006438	HQ006507	-
	Zahiri *et al*. 2011	MM04601	*Deltote*	*uncula*	GU828703	GU829026	GU829315	GU828192	GU829818	GU830089	-
	Zahiri *et al*. 2011	MM04543	*Cucullia*	*umbratica*	GU828701	GU829024	GU829313	GU828190	GU829817	GU830087	-
	Zahiri *et al*. 2011	CWM-94-0372	*Raphia*	*abrupta*	GU828548	GU828893	GU829193	GU828059	GU829728	GU829939	-
	Zahiri *et al*. 2011	MM00152	*Acontia*	*lucida*	GU828617	GU828952	GU829243	GU828118	GU829763	GU830006	-
	Zahiri *et al*. 2011	RZ54	*Arcte*	*modesta*	HQ006226	HQ006321	HQ006413	HQ007015	-	HQ006559	-
	Zahiri *et al*. 2011	MM05132	*Abrostola*	*tripartita*	HQ006152	HQ006250	HQ006345	HQ006952	-	HQ006502	-
	Zahiri *et al*. 2011	MM09894	*Xanthodes*	*albago*	GU828844	GU829145	GU829412	GU828308	-	GU830224	-
	Zahiri *et al*. 2011	MM04583	*Panthea*	*coenobita*	GU828702	GU829025	GU829314	GU828191	-	GU830088	-
	Zahiri *et al*. 2011	MM09893	*Emmelia*	*trabealis*	HQ006147	HQ006245	HQ006340	HQ006947	-	-	-
	Zahiri *et al*. 2012	RZ147	*Pseudoarcte*	*melanis*	JN401235	JN401353	JN401465	JN401037	JN401567	JN401669	-
	Zahiri *et al*. 2012	RZ281	*Sosxetra*	*grata*	JN401236	JN401354	JN401466	JN401038	JN401568	JN401670	-
	Zahiri *et al*. 2012	RZ384	*Belciana*	*biformis*	JN401237	JN401355	JN401467	JN401039	JN401569	JN401671	-
	Zahiri *et al*. 2012	RZ416	*Belciana*	*kala*	JN401238	JN401356	JN401468	JN401040	JN401570	JN401672	-
	Zahiri *et al*. 2012	RZ395	*Dyrzela*	*plagiata*	JN401240	JN401358	JN401469	JN401042	JN401572	JN401673	-
	Zahiri *et al*. 2012	RZ50	*Amyna*	*octo*	JN401242	JN401360	JN401471	JN401043	JN401574	JN401675	-
	Zahiri *et al*. 2012	RZ351	*Encruphion*	*leena*	JN401243	JN401361	JN401472	JN401044	JN401575	JN401676	-
	Zahiri *et al*. 2012	RZ388	*Antitrisuloides*	*catocalina*	JN401248	JN401365	JN401476	JN401048	JN401578	JN401679	-
	Zahiri *et al*. 2012	RZ472	*Diopa*	*corone*	JN401239	JN401357	-	JN401041	JN401571	-	-
	Zahiri *et al*. 2012	RZ464	*Parangitia*	*mosaica*	JN401245	JN401363	JN401474	JN401046	-	JN401677	-
	Zahiri *et al*. 2012	RZ87	*Eucocytia*	*meeki*	JN401247	-	-	JN401047	-	-	-
	Zahiri *et al*. 2012	RZ382	*Ramadasa*	*pavo*	JN401241	JN401359	JN401470	-	JN401573	JN401674	-
	Zahiri *et al*. 2012	RZ459	*Thiacidas*	*sp.*	JN401249	JN401366	JN401477	-	JN401579	JN401680	-
	Zahiri *et al*. 2012	MM09267	*Diloba*	*caeruleocephala*	JN401246	JN401364	JN401475	-	-	JN401678	-
One-gene alignment only	This study	littoralis26	*Spodoptera*	*littoralis*	½COI	EF1a	-	-	-	-	cytB
	This study	exigua44	*Spodoptera*	*exigua*	½COI	-	-	-	-	-	cytB
	This study	exigua102	*Spodoptera*	*exigua*	COI	-	-	-	-	-	cytB
	This study	frugiperda1159	*Spodoptera*	*frugiperda*	COI	-	-	-	-	-	cytB
	This study	frugiperda76	*Spodoptera*	*frugiperda*	COI	-	-	-	-	-	cytB
	This study	latifascia46	*Spodoptera*	*latifascia*	½COI	-	-	-	-	-	cytB
	This study	littoralis68	*Spodoptera*	*littoralis*	COI	-	-	-	-	-	cytB
	This study	littoralis1215	*Spodoptera*	*littoralis*	½COI	-	-	-	-	-	cytB
	This study	littoralis69	*Spodoptera*	*littoralis*	½COI	-	-	-	-	-	cytB
	This study	littoralis87	*Spodoptera*	*littoralis*	COI	-	-	-	-	-	cytB
	This study	littoralis93	*Spodoptera*	*littoralis*	COI	-	-	-	-	-	cytB
	This study	littoralis104	*Spodoptera*	*littoralis*	COI	-	-	-	-	-	cytB

	This study	littoralis106	*Spodoptera*	*littoralis*	COI	-	-	-	-	-	cytB
	This study	littoralis108	*Spodoptera*	*littoralis*	COI	-	-	-	-	-	cytB
	This study	litura79	*Spodoptera*	*litura*	½COI	-	-	-	-	-	cytB
	This study	litura970	*Spodoptera*	*litura*	COI	-	-	-	-	-	cytB
	This study	litura83	*Spodoptera*	*litura*	½COI	-	-	-	-	-	cytB
	This study	litura58	*Spodoptera*	*litura*	½COI	-	-	-	-	-	cytB
	This study	litura70	*Spodoptera*	*litura*	COI	-	-	-	-	-	cytB
	This study	litura95	*Spodoptera*	*litura*	COI	-	-	-	-	-	cytB
	This study	mauritia152	*Spodoptera*	*mauritia*	COI	-	-	-	-	-	cytB
	This study	mauritia154	*Spodoptera*	*mauritia*	COI	-	-	-	-	-	cytB
	This study	mauritia1097	*Spodoptera*	*mauritia*	COI	-	-	-	-	-	cytB
	This study	Agro1205	*Agrotis*	*segetum*	**ML**	**ML**	-	-	-	-	-
	This study	Evenus	*Elaphria*	*venustula*	**ML**	-	-	-	-	-	-
	This study	triturata1208	*Spodoptera*	*triturata*	*ML*	-	-	-	-	-	-
	This study	Aposp1269	*Apospasta*	*fuscirufa*	**ML**	-	-	-	-	-	-
	This study	Call1203	*Callopistria*	*maillardi*	**ML**	-	-	-	-	-	-
	This study	Copi1263	*Copitarsia*	*spec.*	**ML**	-	-	-	-	-	-
	This study	Ment1207	*Mentaxya*	*albifrons*	**ML**	-	-	-	-	-	-
	This study	Enuc1206	*Elaphria*	*nucicolora*	**ML**	-	-	-	-	-	-
	This study	Harm1264	*Helicoverpa*	*armigera*	**ML**	-	-	-	-	-	-
	This study	Perid1265	*Peridromia*	*saucia*	**ML**	-	-	-	-	-	-
	This study	exigua100	*Spodoptera*	*exigua*	COI	-	-	-	-	-	-
	This study	exigua1158	*Spodoptera*	*exigua*	COI	-	-	-	-	-	-
	This study	frugiperda22	*Spodoptera*	*frugiperda*	COI	-	-	-	-	-	-
	This study	frugiperda58	*Spodoptera*	*frugiperda*	½COI	-	-	-	-	-	-
	This study	latifascia24	*Spodoptera*	*latifascia*	½COI	-	-	-	-	-	-
	This study	eridania66	*Spodoptera*	*eridania*	COI	-	-	-	-	-	-
	This study	albula64	*Spodoptera*	*albula*	COI	-	-	-	-	-	-
	Pogue 2011	L_hipparis2	*Spodoptera*	*hipparis*	COI	-	-	-	-	-	-
	Pogue 2011	S_albula1	*Spodoptera*	*albula*	COI	-	-	-	-	-	-
	Pogue 2011	S_albula2	*Spodoptera*	*albula*	COI	-	-	-	-	-	-
	Pogue 2011	S_albula3	*Spodoptera*	*albula*	COI	-	-	-	-	-	-
	Pogue 2011	S_albula4	*Spodoptera*	*albula*	COI	-	-	-	-	-	-
	Pogue 2011	S_praefica2	*Spodoptera*	*praefica*	COI	-	-	-	-	-	-
	Pogue 2011	S_praefica3	*Spodoptera*	*praefica*	COI	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*latifascia*	GU337022	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*perigeana*	GU163190	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*versicolor*	GU438191	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*alapallida*	GU438183	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*cornutinus*	GU087784	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*hyposcota*	JQ559447	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*agrotina*	JQ559197	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*subobliqua*	JQ563860	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*rubripicta*	JQ563848	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*devara*	JQ563288	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*perigeana*	JQ562626	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*marmorata*	JQ561702	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*niveopis*	JQ559604	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*venustula*	JF860086	-	-	-	-	-	-
	GenBank	see_GenBank	*Elaphria*	*venustula*	HQ563361	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exigua*	GU094753	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exigua*	GU707393	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exigua*	JF415658	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exigua*	HM756080	-	-	-	-	-	-

	GenBank	see_GenBank	*Spodoptera*	*exigua*	HM756079	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exigua*	HM756078	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exigua*	HM756077	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exigua*	HM914242	-	-	-	-	-	-
	GenBank	PHMO36203	*Spodoptera*	*exigua*	GU094753	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*ornithogalli*	U72979	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*ornithogalli*	GU094314	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*ornithogalli*	GU088102	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*ornithogalli*	GU088101	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*ornithogalli*	MJMSL004	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*ornithogalli*	XAB64304	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*ornithogalli*	LOT32604	-	-	-	-	-	-
	GenBank	riceHT1	*Spodoptera*	*frugiperda*	U72977	-	-	-	-	-	-
	GenBank	riceHT2	*Spodoptera*	*frugiperda*	U72978	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	EU768963	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	GU094754	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	GU094756	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	GU159435	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	GU090724	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HM136589	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HM136587	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HM136586	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HM136588	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HM136594	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HM136593	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HM136602	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	ACLB001	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	GU159431	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	GU439151	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	GU439147	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	XAD265	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	XAD490	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HQ964527	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*frugiperda*	HM136598	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*latifascia*	GU159445	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*latifascia*	GU159444	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*latifascia*	GU159443	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*latifascia*	GU159442	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*latifascia*	GU159440	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*latifascia*	GU159438	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*latifascia*	GU159436	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*eridania*	HM756085	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*eridania*	HM756082	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*eridania*	HM756081	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*eridania*	GU159425	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*eridania*	GU159424	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*eridania*	GU159422	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*eridania*	GU159421	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	GU337021	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	HM756086	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	HM756089	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	HM756088	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	HM756087	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	GU159420	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	GU159419	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	GU159418	-	-	-	-	-	-

	GenBank	see_GenBank	*Spodoptera*	*dolichos*	GU159417	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*dolichos*	GU159416	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*androgea*	GU159405	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*androgea*	GU159411	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*androgea*	GU159408	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*androgea*	GU163696	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*androgea*	GU159414	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*androgea*	GU159409	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*androgea*	GU159407	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*littoralis*	EZ983516	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*littoralis*	HM756074	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*litura*	HM756093	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*litura*	HM756092	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*litura*	HM756091	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*pulchella*	HM756076	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*pulchella*	HM756075	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*triturata*	HM892616	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*triturata*	HM892940	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exempta*	DQ092374	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exempta*	DQ092375	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exempta*	DQ092371	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exempta*	DQ092372	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exempta*	DQ092373	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exempta*	DQ092376	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exempta*	DQ092370	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*exempta*	HM893111	-	-	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*praefica*	HM867882	-	-	-	-	-	-
	GenBank	see_GenBank	*Galgula*	*partita*	HQ964391	-	-	-	-	-	-
	GenBank	see_GenBank	*Peridromia*	*saucia*	GU094228	-	-	-	-	-	-
	This study	exigua1162	*Spodoptera*	*exigua*	-	-	-	-	-	-	cytB
	This study	frugiperda56	*Spodoptera*	*frugiperda*	-	-	-	-	-	-	cytB
	This study	latifascia34	*Spodoptera*	*latifascia*	½COI	-	-	-	-	-	cytB
	This study	latifascia1128	*Spodoptera*	*latifascia*	-	-	-	-	-	-	cytB
	This study	latifascia1129	*Spodoptera*	*latifascia*	-	-	-	-	-	-	cytB
	This study	latifascia1130	*Spodoptera*	*latifascia*	-	-	-	-	-	-	cytB
	This study	eridania1136	*Spodoptera*	*eridania*	-	-	-	-	-	-	cytB
	This study	dolichos1124	*Spodoptera*	*dolichos*	-	-	-	-	-	-	cytB
	This study	dolichos1125	*Spodoptera*	*dolichos*	-	-	-	-	-	-	cytB
	This study	dolichos1126	*Spodoptera*	*dolichos*	-	-	-	-	-	-	cytB
	This study	dolichos1127	*Spodoptera*	*dolichos*	-	-	-	-	-	-	cytB
	This study	littoralis52	*Spodoptera*	*littoralis*	½COI	-	-	-	-	-	cytB
	This study	littoralis48	*Spodoptera*	*littoralis*	½COI	-	-	-	-	-	cytB
	This study	littoralis57	*Spodoptera*	*littoralis*	-	-	-	-	-	-	cytB
	This study	littoralis77	*Spodoptera*	*littoralis*	-	-	-	-	-	-	cytB
	This study	albula1120	*Spodoptera*	*albula*	-	-	-	-	-	-	cytB
	This study	albula1122	*Spodoptera*	*albula*	-	-	-	-	-	-	cytB
	This study	nonSpodoptera1100	*indet*	*sp*	COI	-	-	-	-	-	cytB
	This study	pulchella1131	*Spodoptera*	*pulchella*	-	-	-	-	-	-	cytB
	This study	pulchella1132	*Spodoptera*	*pulchella*	-	-	-	-	-	-	cytB
	GenBank	see_GenBank	*Bombyx*	*mori*	-	-	-	-	-	-	GU966623
	GenBank	see_GenBank	*Autographa*	*gamma*	-	-	-	-	-	-	AB125678
	GenBank	see_GenBank	*Spodoptera*	*litura*	-	-	-	-	-	-	AB125946
	Mitchell *et al.* 2006	see_GenBank	*Anicla*	*infecta*	-	U85703	U85703	-	-	-	-
	Mitchell *et al.* 2006	see_GenBank	*Diarsia*	*rosaria*	-	AY952613	AY952613	-	-	-	-

	Mitchell *et al.* 2006	see_GenBank	*Peridromia*	*saucia*	-	AY952614	AY952614	-	-	-	-
	GenBank	see_GenBank	*Spodoptera*	*litura*	-	DQ192234	DQ192234	-	-	-	-
	Cho *et al.* 2008	see_GenBank	*Helicoverpa*	*assulta*	EU768937	EU769062	EU769062	-	-	-	-
	Cho *et al.* 2008	see_GenBank	*Helicoverpa*	*gelotopoeon*	EU768938	U20132	U20132	-	-	-	-
	Cho *et al.* 2008	see_GenBank	*Helicoverpa*	*hawaiiensis*	EU768939	EU769063	EU769063	-	-	-	-
	Cho *et al.* 2008	see_GenBank	*Helicoverpa*	*pallida*	EU768940	EU769047	EU769047	-	-	-	-
	Cho *et al.* 2008	see_GenBank	*Helicoverpa*	*punctigera*	EU768941	EU769064	EU769064	-	-	-	-
	Cho *et al.* 2008	see_GenBank	*Helicoverpa*	*zea*	EU768942	U20136	U20136	-	-	-	-
only used in prelimi-nary analyses	This study	exigua3126055	*Spodoptera*	*exigua*	½COI	-	-	-	-	-	-
	This study	exigua3738297	*Spodoptera*	*exigua*	½COI	-	-	-	-	-	-
	This study	dolichos1155	*Spodoptera*	*dolichos*	COI	-	-	-	-	-	-
	This study	mauritia156	*Spodoptera*	*mauritia*	COI	-	-	-	-	-	-
	This study	mauritia148FJI	*Spodoptera*	*mauritia*	***ML***	-	-	-	-	-	-
	This study	mauritia150FJI	*Spodoptera*	*mauritia*	***ML***	-	-	-	-	-	-
	This study	maur148	*Spodoptera*	*mauritia*	**ML**	-	-	-	-	-	-
	This study	maur150	*Spodoptera*	*mauritia*	**ML**	-	-	-	-	-	-
	This study	cf_ochrea/exigua_ PD450896	*Spodoptera?*	*ochrea?*	½COI	-	-	-	-	-	-
	This study	nonSpodoptera1106	*indet*	*sp*	COI	-	-	-	-	-	-
	This study	nonSpodoptera1105	*indet*	*sp*	COI	-	-	-	-	-	-
	This study	nonSpodoptera1101	*indet*	*sp*	COI	-	-	-	-	-	-
	This study	exigua1161	*Spodoptera*	*exigua*	-	-	-	-	-	-	-
	This study	littoralis138	*Spodoptera*	*littoralis*	-	-	-	-	-	-	-
	This study	litura142	*Spodoptera*	*litura*	-	-	-	-	-	-	-
	This study	pecten146	*Spodoptera*	*pecten*	-	-	-	-	-	-	-
	This study	Diar1204	*Diarsia*	*sp*	-	-	-	-	-	-	-
	This study	Anic1261	*Anicla*	*infestans*	-	-	-	-	-	-	-
Totals					248	71	60	49	41	43	58

2.3.1. Re-sequencing and (re-)assembly

The COI sequences of maur0148, maur0150 and picta112 in the starting data set were low in data quality, see figure 6 for one example. I re-sequenced these specimens and reassembled their tracers into new contigs. I overlooked the presence of the COI sequence of doli1123, so the sequence I generated thereof was superfluous.

I wasn't satisfied with the quality of the COI contig of malagasy1103. I re-analysed the sequence tracers and reassembled them into a more satisfactory contig.

The tracres of triturata1208 were not assembled yet into a contig, which I therefore did.

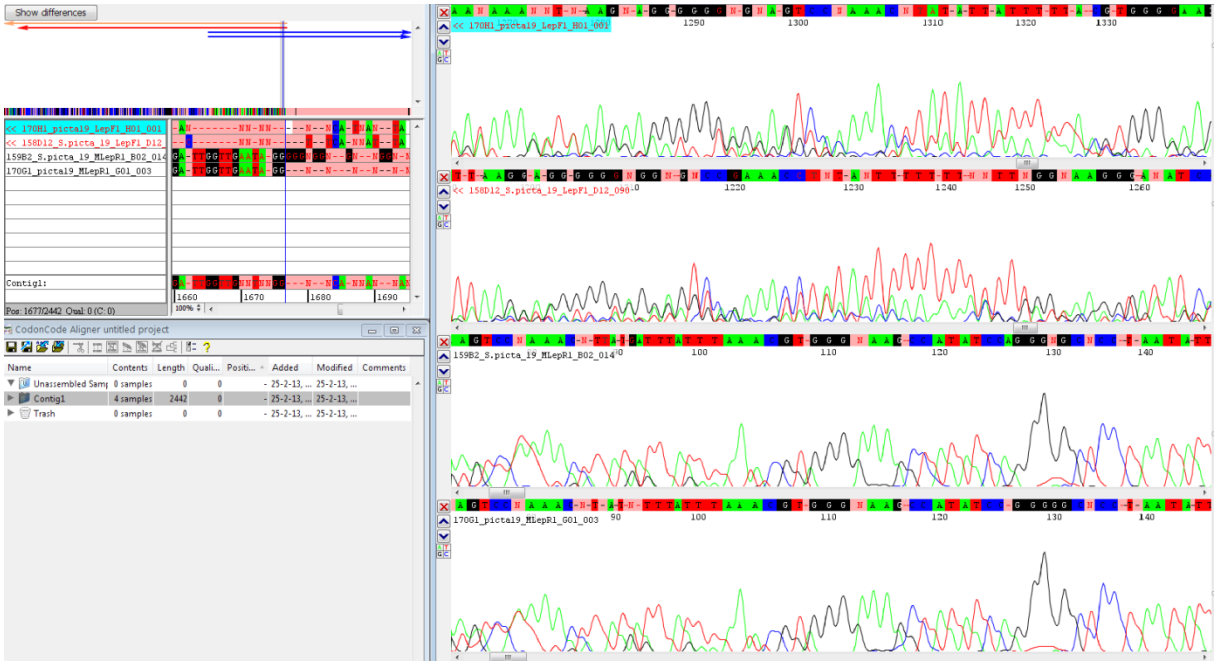


Figure 6. Low quality sequence tracers of *Spodoptera picta*.
Many double peaks and peak shifts are visible, resulting in a low quality assembly with many gaps.

Figure 7. Assembly of the tracers of the re-sequenced *Spodoptera picta*.
Every site has 2x coverage.

2.3.2. IDH-problems

I had no success sequencing IDH for any specimen. On May 22th, I attempted to amplify CAD and IDH for exig102 and doli1123 using 1 µL of both undiluted and 1:10-diluted template DNA for the first time. IDH wasn't successfully amplified; CAD was only amplified for doli1123. On May 23th I also failed to amplify COI for exig102. Therefore I decided not to use this specimens' DNA extract anymore. I made new IDH-primer aliquots, in order to rule out the primers as cause of the problems. On May 31st I tried again to amplify IDH for the second time, as well as CAD and GAPDH, for 6 specimens: litu1262, frugi1267, litt1268, doli1123, exig1160 and Eagr1270. CAD was successfully amplified for 4 of the 6 specimens, GAPDH was successful for all. For IDH vague bands were visible, which I purified and cyclesequenced. To improve the chance of sequence success I added 2 or 3 µL, instead of the usual 1 µL, to the cyclesequence mix depending on the strength of the PCR-product on the gel. Returned sequence tracers were mostly useless. Some showed minor results, like figure 8a, where up to 100bp was amplified, but then still this part wasn't useable because of the high number of double peaks. On June 15th I cyclesequenced the IDH amplicons of the 6 specimens again, this time using 4.5-5.5 µL of amplicon in the cyclesequence reaction. This also had not the anticipated effect: all trace files seemed to show a positive result on the first look. However, on closer inspection, most peaks weren't single peaks, but often double or even triple peaks (see figure 8b). Therefore these results were discarded.

One can only speculate on the causes of the failure. What I do know, is that the all the reagents in the PCR reaction mix (except the primers) were OK, because most other sequencing attempts succeeded. Perhaps there was something wrong with the primer aliquots. It could even be that the primer stock isn't OK, because I made a new aliquot, which still had only minor result. Perhaps *Spodoptera* species share important substitutions in the primer binding sites, or maybe *Spodoptera* has multiple primer binding sites explaining the double peaks in the tracers.
If someone in the future would like to try again to obtain a sequence of IDH, I recommend sending the specimens to the Wahlberg lab and letting them try it. I feel confident that I tried whatever I could in the time given. The only thing I haven't tried, is ordering new stocks of the primers, because there wasn't any time left to use them anyway. See the discussion for other remarks on this marker.

IDH has great potential as a useful gene in phylogenetic reconstruction (Wahlberg and Wheat 2008, Zahiri *et al*. 2011, 2012). I have no sequences for *Spodoptera* for this gene. Therefore this marker did not contribute to the position of *Spodoptera* in the Noctuidae. However, the other nuclear markers did provide this information. Nonetheless, the IDH-part of the data set was retained because it would contribute to resolving the Noctuid backbone.

a)

b)

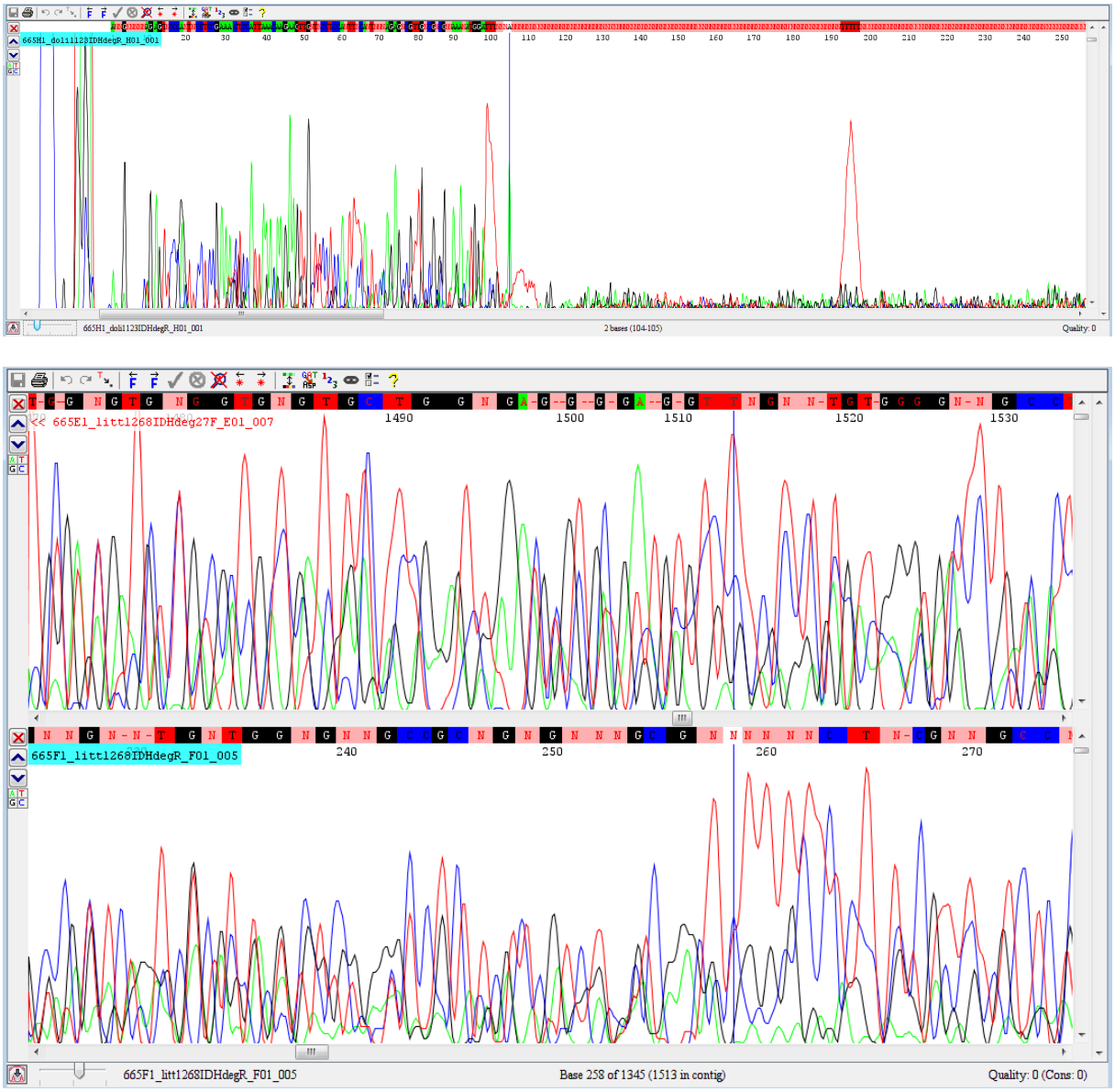


Figure 8. IDH sequence tracers of the second sequencing attempt.
a) Trace of the sequence of doli1123 using the reverse-primer. The first ~100 bases are amplified (but see text).
b) Tracers of the sequence of litt1268, forward (top) and reverse (bottom). The quality is extremely low, because of all the overlapping peaks and all double/triple peaks.

2.4. Alignments

Aligning the DNA-sequence data was done by hand in Mesquite version 2.74 (Maddison and Maddison 2007). In all cases the alignment of sequences was trivial, because all sequenced genes were free of gaps, except IDH which has 3 bp of indels (one amino acid).
The codon frame was visually checked in Mesquite by defining codon positions in the 'List of characters'-tab and next colouring the character matrix by amino acid. This first confused me, because COI and CytB were coding for a lot of stop-codons, no matter what codon frame was chosen. This was solved by setting the genetic code for this part of the alignment to 'insect mitochondrial', which has a different coding from all other cell types.
Non-silent substitutions (substitutions which led to an amino acid change) and substitutions resulting in stop-codons were double-checked in CodonCodeAligner. No stop-codons were present in the final alignments.

One alignment per gene was made, see table 12 for their details. Mesquite saves alignments in Nexus-format. Other programs require other formats; each of these formats is mentioned at the relevant sections.
All alignments were trimmed when necessary in order to make the first nucleotide in the alignment a first coding position nucleotide: in the COI and GAPDH alignments the first character and in the IDH alignment the first two characters were removed.
Every one-gene alignment was first analysed separately in order to inspect the data quality. For example, from the alignment of COI (221SpodCOI.nex) were 10 sequences removed because of the reliability of the data.

Congruence of the phylogenetic reconstructions was inspected visually. See also section 2.6 'Phylogenetic analyses' for the specification of the analyses performed. Not all branching patterns were the same, but there was not sufficient data to accept hard cytonuclear incongruence.

The average percentage of data coverage (the amount of sites in a sequence not missing) of the alignment, was calculated manually with this formula:

$$Data\ coverage = \frac{\sum(number\ of\ sites\ with\ A,\ C,\ T,\ G)}{total\ number\ of\ sites} * 100\%$$

using the 'Matrix>Show Selection Summary Strip'-option in Mesquite version 2.74 (Maddison and Maddison 2007). The nucleotide composition is calculated over the non-missing part, following this formula:

$$Percentage\ of\ nucleotide\ i = \frac{number\ of\ sites\ with\ nucleotide\ i}{\sum(number\ of\ sites\ with\ A,\ C,\ T,\ G)} * 100\%$$

Table 12. Size, length, data coverage and nucleotide composition of the one-gene alignments.

Gene	Name	Number of terminals	Length	Data coverage	Nucleotide composition			
					A	C	G	T
COI	210SpodCOI	211	657 bp	91.6%	29.8%	15.4%	14.5%	40.3%
CytB	60SpodCytB	60	602 bp	93,0%	34.1%	14.5%	9.6%	41.8%
EF1α	59SpodEF1a	59	1032 bp	75.0%	25.6%	28.4%	25.0%	21.0%
CAD	51SpodCAD	51	849 bp	76.8%	35.6%	14.0%	21.1%	29.4%
GAPDH	43SpodGAPDH	43	690 bp	90.0%	22.6%	29.6%	23.6%	24.2%
IDH	45SpodIDH	45	711 bp	96.7%	30.5%	21.1%	22.0%	26.4%

Next, two subsets of the specimens with high data coverage were chosen in order to answer the different research questions asked earlier. Not all data was relevant or required for the question at hand. Several considerations play a role: A smaller taxonomic sampling increases analytical speed, while a denser taxon sampling in a clade tends to increase that clades' age (Pirie *et al*. 2005), an effect later named the Node Density Effect artefact (Hugall and Lee 2007), which potential I tried to minimize by only including the nodes of interest. Their chosen specimens' sequences were concatenated into two different supermatrices. These were named 'Concatenated Alignment 1' (CA1) for research goal one (*Spodoptera* species level phylogenetic reconstruction) and 'Concatenated Alignment 2' (CA2) for research goals two (position of *Spodoptera* in the Noctuidae) and three (Age of *Spodoptera*). These alignments are described below.

Concatenated Alignment 1 (49SpodFused.nex)**:** Contains all 6 genes and 49 taxa of which 26 *Spodoptera*. The other 23 taxa are selected Noctuids based on Cho *et al*. (2008) and Zahiri *et al*. (2011, 2012). The outgroup in this alignment is *Autographa gamma*.

Concatenated Alignment 2 (61SpodFused.nex)**:** Contains 5 genes (all except CytB), 61 taxa of which 9 *Spodoptera*. The other 52 taxa are all Noctuids, two Notodontids and three outgroups from Zahiri *et al*. (2011, 2012) and the same data from Cho *et al*. (2008) as in Concatenated Alignment 1.
First this alignment was analysed in MrBayes (see section 2.6 'Phylogenetic analyses') in order to confidently position *Spodoptera* in a subclade of the Noctuidae. Next this alignment was analysed in BEAST (see section 2.6 'Phylogenetic analyses') in order to date the age of the Noctuidae and *Spodoptera*.

In these Concatenated Alignments there are several specimens of which several markers are missing. This problem is well-investigated and usually had no consequences (see Wiens and Morrill 2011 and references therein). Therefore I do not expect this to have a major effect in my analyses as well.

2.5. Substitution model selection

Selection of the proper substitution model per gene was done with jModelTest version 0.1.1 (Posada 2008). The alignments for input were exported from Mesquite in FASTA-format as required for jModelTest. Settings were left to the defaults, except that the I-parameter (the estimation of the amount of invariant characters) was excluded because it is correlated with, and inseparable from, the gamma-parameter (Yang 1996, Wadall *et al*. 1997, Sullivan and Swofford 2001, Ren *et al*. 2005; Kelchner & Thomas 2007). Selection of the best fitting model was based on lowest Akaike information criterion (AIC). The available (and thus tested) models in jModelTest are listed in table 16. Models were first tested on the entire one-gene data sets (see table 13) and next on the ingroup only (see table 14), which is the main part the model of evolution should fit to. Comparison of the tables shows that this choice can influence the selection of the model: For GAPDH another model is better fitting the data.
Summary of the best fitting models' statistics are listed in table 14. Not all models are supported by MrBayes version 3.1.2 (Huelsenback and Ronquist 2001): only JC69, F81, K80, HKY85, SYM and GTR are supported (not the models having separate parameters for both transitions but not for transversions). BEAST version 1.7 (Drummond and Rambaut 2007) theoretically supports all models, however, this requires manual editing of the XML-input file, which was beyond my capabilities. Therefore for CAD, GAPDH, IDH and CytB the second best model is chosen (see table 15). This resulted in all cases in choosing the most parameter-rich model (GTR+gamma). I think this is OK, because in simulation studies (e.g. Lemmon and Moriarty 2004) overparameterization has little effect on bipartition accuracy and branch length estimation, in contrast to the major negative effects of underparameterization. Overparameterization may lead to lower precision (Lemmon and Moriarty 2004), therefore I assume a high threshold for a well-supported node: pp ≥ 0.95.

Table 13. Summary of the best fitting substitution models per gene.
fA, fC, fG & fT = the calculated base frequencies of A, C, G & T respectively.
R(ij) = the calculated relative substitution rate of nucleotide i to nucleotide j.

Gene	Best model	fA	fC	fG	fT	R(AC)	R(AG)	R(AT)	R(CG)	R(CT)	R(GT)
COI	GTR+G	0.265	0.138	0.156	0.441	6.595	11.763	23.934	1.236	62.150	1
EF1α	GTR+G	0.263	0.286	0.236	0.216	3.388	13.688	7.038	2.459	29.264	1
CAD	TIM3+G	0.421	0.105	0.168	0.306	12.906	14.441	1	12.906	117.919	1
GAPDH	SYM+G	-	-	-	-	1.314	4.957	1.670	0.861	9.214	1
IDH	TIM2+G	0.287	0.208	0.204	0.301	1.745	6.336	1.745	1	10.085	1
CytB	TIM2+G	0.32	0.11	0.1	0.47	33.52	74.81	33.52	1	310.39	1

Table 14. Summary of the best fitting substitution models per gene on the ingroup only.
fA, fC, fG & fT = the calculated base frequencies of A, C, G & T respectively.
R(ij) = the calculated relative substitution rate of nucleotide i to nucleotide j.

Gene	Best model	fA	fC	fG	fT	R(AC)	R(AG)	R(AT)	R(CG)	R(CT)	R(GT)
COI	GTR+G	0.255	0.143	0.159	0.442	7.563	12.181	24.024	0.966	60.601	1
EF1α	GTR+G	0.257	0.285	0.242	0.215	3.610	13.597	7.457	2.298	29.702	1
CAD	TIM3+G	0.425	0.100	0.169	0.306	15.631	16.150	1	15.631	147.494	1
GAPDH	TPM2uf+G	0.200	0.303	0.221	0.276	1.826	7.651	1.826	1	7.651	1
IDH	TIM2+G	0.283	0.214	0.203	0.300	1.699	6.479	1.699	1	9.987	1
CytB	TIM2+G	0.31	0.12	0.1	0.48	40.06	99.72	40.06	1	361.72	1

Table 15. Summary of the selected substitution models per gene on the ingroup only.
fA, fC, fG & fT = the calculated base frequencies of A, C, G & T respectively.
R(ij) = the calculated relative substitution rate of nucleotide i to nucleotide j.

Gene	Selected model	fA	fC	fG	fT	R(AC)	R(AG)	R(AT)	R(CG)	R(CT)	R(GT)
COI	GTR+G	0.255	0.143	0.159	0.442	7.563	12.18	24.02	0.966	60.601	1
EF1α	GTR+G	0.257	0.285	0.242	0.215	3.61	13.6	7.457	2.298	29.702	1
CAD	GTR+G	0.360	0.161	0.168	0.311	2.492	9.459	1.991	1.425	13.286	1
GAPDH	GTR+G	0.224	0.290	0.236	0.251	1.165	4.83	1.833	0.762	7.3447	1
IDH	GTR+G	0.283	0.216	0.203	0.298	1.473	5.746	1.531	0.759	8.8653	1
CytB	GTR+G	0.321	0.115	0.081	0.482	3E+05	1E+06	4E+05	8E+05	4E+06	1

Table 16. Models of nucleotide substitution.

Model	Reference	Free parameters	Base frequencies	Substitution rates
JC69	Jukes and Cantor 1969	0	equal	AC=AG=AT=CG=CT=GT
F81	Felsenstein 1981	3	unequal	AC=AG=AT=CG=CT=GT
K80	Kimura 1980	1	equal	AC=AT=CG=GT; AG=CT
HKY85	Hasegawa, Kishino, and Yano 1985	4	unequal	AC=AT=CG=GT; AG=CT
TNef	Tamura and Nei 1993	2	equal	AC=AT=CG=GT; AG; CT
TN	Tamura and Nei 1993	5	unequal	AC=AT=CG=GT; AG; CT
TPM1 =K81	Kimura 1981	2	equal	AC=GT; AT=CG; AG=CT
TPM1uf	Kimura 1981	5	unequal	AC=GT; AT=CG; AG=CT
TPM2	Kimura 1981	2	equal	AC=AT; CG=GT; AG=CT
TPM2uf	Kimura 1981	5	unequal	AC=AT; CG=GT; AG=CT
TPM3	Kimura 1981	2	equal	AC=CG; AT=GT; AG=CT
TPM3uf	Kimura 1981	5	unequal	AC=CG; AT=GT; AG=CT
TIM1ef	Posada 2003	3	equal	AC=GT; AT=CG; AG; CT
TIM1	Posada 2003	6	unequal	AC=GT; AT=CG; AG; CT
TIM2ef	Posada 2003	3	equal	AC=AT; CG=GT; AG; CT
TIM2	Posada 2003	6	unequal	AC=AT; CG=GT; AG; CT
TIM3ef	Posada 2003	3	equal	AC=CG; AT=GT; AG; CT
TIM3	Posada 2003	6	unequal	AC=CG; AT=GT; AG; CT
TVMef	Posada 2003	4	equal	AC; AT; CG; GT; AG=CT
TVM	Posada 2003	7	unequal	AC; AT; CG; GT; AG=CT
SYM	Zharkikh 1994	5	equal	AC; AG; AT; CG; CT; GT
GTR =REV	Tavaré 1986	8	unequal	AC; AG; AT; CG; CT; GT

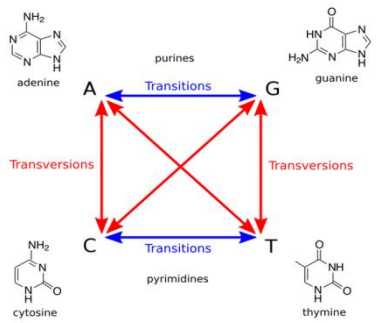


Figure 9. Transitions and transversions.
Source: http://en.wikipedia.org/wiki/File:Transitions-transversions-v3.png

2.6. Phylogenetic analyses

All analyses were either run on a local computer (named 'Henk' and 'Ingrid' at the Biosystematics group), or at the CIPRES TeraGrid in San Diego (the web portal is accessible on http://www.phylo.org/portal2/home.action).

2.6.1. Discerning *Spodoptera* clades: Maximum Likelihood

All one-gene alignments and the concatenated alignments were analysed using maximum likelihood (ML) methods using the program in RAxML version 7.3 (Stamatakis 2006, Stamatakis *et al*. 2008). Each alignment was converted from Nexus to relaxed Phylip-format, as required for RAxML. All RAxML analyses were performed on the CIPRES TeraGrid (see below). Support for the nodes was calculated using 1000 bootstrap iterations. The 50%-majority rule consensus tree (RAxML_bipartitions.result) is the chosen result for interpretation. The resulting phylogenetic reconstructions of all ML-analyses are put in the Appendix. Summaries of the analyses are reported in the results section.
A node will be interpreted as well supported when the bootstrap value is above 95.

2.6.2. Discerning *Spodoptera* clades: Bayesian inference

All alignments were analysed using Bayesian Inference methods in MrBayes version 3.1.2 (Huelsenback and Ronquist 2001) or BEAST (Bayesian Evolutionary Analysis by Sampling Trees) version 1.7 (Drummond and Rambaut 2007), depending on which software package had to be mastered by the author at the time of analysis.
For MrBayes, the alignments were exported from Mesquite to the 'Nexus for MrBayes'-format. BEAST requires alignments in 'simplified Nexus' format.
Data was partitioned into genes because of the variation in nucleotide frequencies and substitution rates as estimated by jModelTest (see section 2.5 above). Next the data was partitioned across genes and codon positions, where the first and second codon positions are taken together and the third codon position separate. The GTR+gamma model was set for all partitions. The nucleotide frequencies, substitution rates, shape of the gamma-parameter were unlinked for all partitions. Site-specific rates were allowed to vary in all partitions. All other priors were left to default. Two independent runs of four MCMC chains with a temperature of 0.05 were performed for a different number of generations in each analysis. The number of generations per run is specified in the results section. A relative burn-in fraction of the first quarter was discarded.
A node was interpreted as well supported when the posterior probability is above 0.95. Resulting phylogenetic reconstructions of the COI one-gene alignment (barcode tree) and both concatenated alignments are reported in the results section. The results of the other one-gene alignments were put in the appendix; summaries of these analyses are reported in the results section.

2.6.3. Dating the clades

In order to estimate the age of *Spodoptera* clades a calibrated phylogenetic reconstruction in BEAST (Drummond and Rambaut 2007) was made. The one-gene alignments were exported from Mesquite in simplified Nexus format, as required for BEAUTi (Bayesian Evolutionary Analysis Utility) version 1.7 (Drummond *et al*. 2012), the input program for BEAST. This basically reconstructs the same supermatrix as Concatenated Alignment 2. Using BEAUTi the XML-files as required for BEAST are easily generated. An uncorrelated lognormal relaxed molecular clock was selected, following advice from Drummond *et al*. (2006). The data was partitioned into genes and codon positions, where the first and second codon positions are taken together and the third codon position separate. All genes were analysed under the GTR+gamma model. As in the MrBayes-analyses, all nucleotide frequencies and substitution rates were unlinked across all partitions and the clock models were unlinked between genes. The ingroup (Noctuoidea), Notodontidae and Noctuidae (note that *Spodoptera* wasn't defined as a taxon set) was set as monophyletic. The chosen calibrations and the resulting priors (prior probability density distributions) are described below.
The tree prior was set to the Yule process, with a uniform prior between [0,1] and an initial value of 0.5. A starting tree with node heights was included. This tree was the best tree created in RAxML using the same data (Concatenated Alignment 2) and the node heights were rescaled to absolute ages in order to remove conflict between the starting tree and the priors. This step is necessary when working with (multiple) calibrations in BEAST, because otherwise the chance is very small that the automatically generated starting tree is congruent with the other specified priors, resulting in termination of the program. The MCMC chain length was 50,000,000 generations and sampled once every 5000 generations. Other settings were left to default.
Before the actual analysis a run was performed with an empty alignment (thus sampling from the prior only) in every case, in order to get an indication of possible conflict between priors and to infer the relative influence of the prior settings on the results. No conflict was found on visual inspection.

In TreeAnnotator version 1.6 (part of the BEAST package, Drummond and Rambaut 2007) the first 10% of trees was discarded as burn-in, resulting in 9000 trees per run, which were summarised by the program into one target tree with posterior probabilities at the nodes. One of these trees and the statistics of all trees are reported in the results section.

2.6.3.1. Calibrations

The first calibration point I used was that of a fossil Notodontid moth (Prokop 2003, Kvaček *et al*. 2004, figure 10). The authors describing the fossil recognised its trifid venation (Prokop 2003, Kvaček *et al*. 2004, pers. comm. J. Prokop), which is a synapomorphy with the Notodontidae (Zahiri *et al*. 2012). The fossil was found in the Most Formation in the Czech Republic in a layer dated at approximately 20 million years old (Kvaček *et al*. 2004, Prokop *et al*. 2010). This dating was based on the Mollusc and Mammal fossils found in the layer, a common way of dating sedimentary strata (pers. comm. J Prokop).
Fossils only provide minimum ages for nodes (Ho and Phillips 2009). A lognormal distribution is usually the prior probability distribution shape of choice (e.g. Ho and Phillips 2009). This shape, however, has some limitations in the specification of its exact shape. The gamma distribution, however, can take a similar shape as the lognormal distribution, with the added benefit of manipulating its shape through the alpha-parameter. I therefore chose to use a gamma distribution with an alpha shape parameter of 2, which results in roughly the same shape as a lognormal distribution. The offset was put on 20 million years, the hard minimum age for this family based on the age of the fossil.

Figure 10. The 20 million year old Notodontid fossil from the Most Formation in the Czech Republic. This photograph was sent to me by the leading investigator, J. Prokop from the Charles University in Prague.

The second calibration point was a secondary calibration derived from a dating study on the 'butterflies' (Heikkilä *et al*. 2012). They estimated the ancestor of the Papilionoidea+Hesperoidea to be at most 110 million years old. In the across-Lepidoptera phylogenetic reconstructions (Mutanen *et al*. 2010, Cho and Zwick *et al*. 2011, Regier *et al*. 2013) the Noctuoidea were recovered in a far more derived position relative to the 'butterflies'. Therefore I could put this age of 110 million years old as a conservative maximum age on the tree root.

For the third calibration point, I introduce here what I think is a new, unpublished, concept: 'ecological' calibration. Some morphological traits evolved in a reaction to the emergence of a selective pressure in an ecologically related (e.g. predator) taxon. In this case I used the independent emergence of tympanal organs ('ears') in many Lepidopteran superfamilies, including the metathoracic tympanal organ of the Noctuoidea (figure 11), which supposedly evolved in a response to the emergence of bat echo-location (Kristensen *et al*. 2007). Corroborating evidence for this hypothesis comes from different directions: Noctuoid tympanal organ hearing frequencies are similar to those of bat echo-locating (Yack *et al*. 1999) and Noctuoids respond directly to bat echo-location calls (Nakano *et al*. 2010).
The molecular phylogenetic reconstruction of bats (Chiroptera) is resolved (Meredith *et al*. 2011, Teeling *et al*. 2005, figures 12 and 13): Bats are a monophyletic group, having flight and echo-location as a synapomorphy (Jones and Teeling 2006, Jones and Holderied 2007). Although, like the Lepidoptera, the bat fossil record is very poor (Eiting and Gunnell 2009), Teeling *et al*. (2005) managed to date the inferred ancestral age (using secondary calibrations and 6 fossils) of the crown all bats at 64 Mya (95% CI: 58-71 Mya) and Eick *et al*. (2005) dated them (also using secondary calibrations and 6 fossils) at 62 Mya (95% CI: 56-70 Mya). These two studies both have applied some slightly outdated methods in the application of the priors on the calibrated node ages, but nonetheless they only differ slightly in their final estimates for their crown age.
This most recent common ancestor of all bats was both flying and using echo-location, but at one point these traits must have evolved. We don't exactly know when, but we do know that flight came before echo-location, because a fossil of a bat was found, which was capable of flight but not capable of echo-location (Simmons *et al*. 2008, Speakman 2008).
Because the emergence of the metathoracic tympanal organ (a synapomorphy for all Noctuoidea) evolved in a response to bat echo-location, we can assume that the Noctuoidea are about 64 million years old, with an unknown error margin.

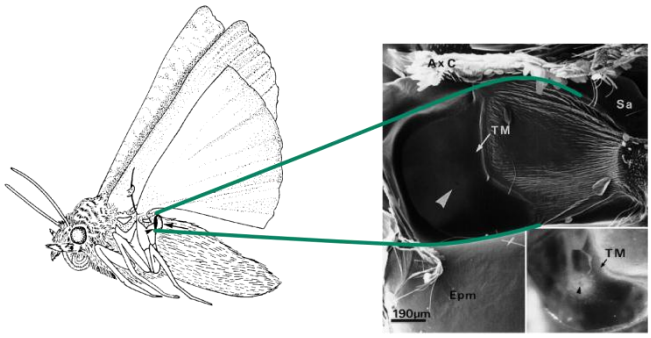


Figure 11. Location (left) and SEM photograph (right) of the Noctuoid tympanal organ. (modified from Yack *et al*. 1999).

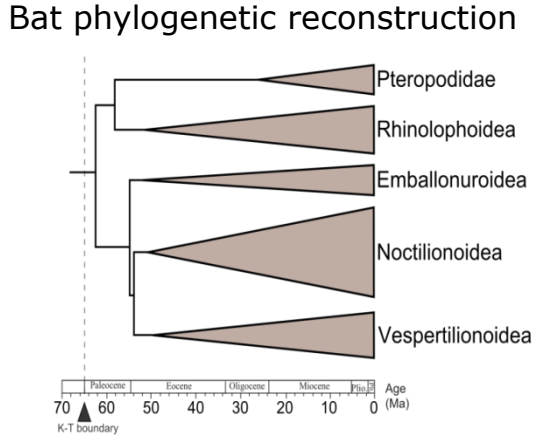


Figure 12. Dated phylogenetic reconstruction of the order Chiroptera (bats). All depicted nodes had a support of 100 (modified from Teeling *et al*. 2005).

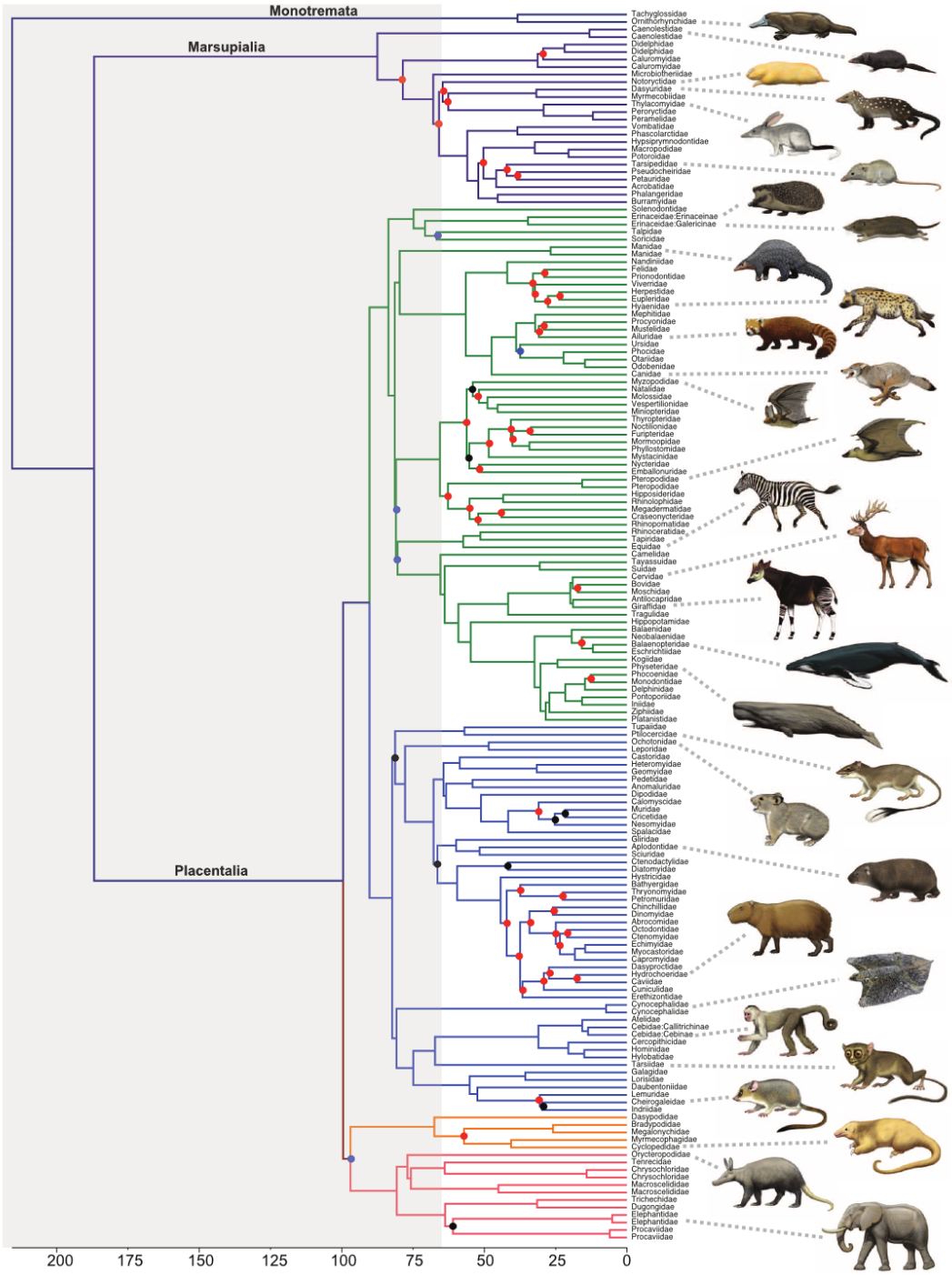


Figure 13. Time-calibrated phylogenetic reconstruction of the Mammalia by Meredith *et al*. 2011. The bats are in the middle of the green coloured part of the tree.

The chosen shape of the prior probability density distributions should reflect the information and uncertainty provided by the evidence (Hedges and Kumar 2004, Drummond *et al*. 2006, Ho and Phillips 2009). Because the three described calibrations have quite some uncertainty to them, I also chose priors with large error margins. These are the settings I used for the calibration points:

- Notodontid fossil: Gamma distribution, shape = 2, scale = 5 , offset = 20, on the crown of the tmrca.Notodontidae (thus excluding the stem);
- Bottom-up (butterfly) calibration point: Normal distribution, μ = 105, σ = 10, in [0; 110] on the treeModel.rootHeight-parameter;
- Bat echo-location calibration point: Normal distribution, μ = 62, σ = 4 on the tmrca.Noctuoidea including the stem.

Figure 14 visualises these prior shapes.

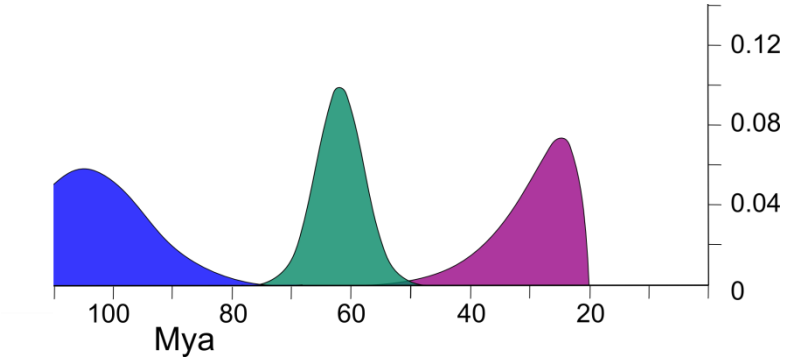


Figure 14. Graphical visualisation of the chosen prior probability density distribution shapes. Purple: Notodontid fossil calibration point. Blue: Secondary calibration point from Heikkilä *et al*. 2012. Green: Ecological calibration point based on the emergence of bat echo-location.

2.6.4. Biogeography

Elaborate analyses and packages are available for reconstructing ancestral areas and dispersal-vicariance patterns (e.g. RASP, Yu *et al*. 2013, which is the new version of S-DIVA, Yu *et al*. 2010). Due to time constraints I could not do any extensive analyses. Nevertheless, I used a quick method to visualise the biogeographical pattern within *Spodoptera*. I imported the phylogenetic reconstruction of Concatenated Alignment 1 in Mesquite. This step loses branch lengths, but I don't need those for this visualisation. I created a character matrix for the tree, wherein I coded species having an Old World distribution as 1 and a New World distribution as 2. *S. exigua* is a cosmopolitan species and is therefore coded as '1,2'. Next, I made a maximum parsimony reconstruction of the ancestral states of the species' distribution, using the Analysis > 'Trace Character History' function with 'Parsimony Ancestral States' method, using the 'Stored Characters', i.e. the coded distribution data. I exported this modified tree as pdf, which I opened in InkScape (see below) to remove the outgroups. This visualisation of the biogeographic pattern is reported in the results section.

2.6.5. Host plant analysis

The host preference of most *Spodoptera* species is not well known (Pogue 2002). Of all the sampled species only one is confidently characterised as oligophagous (*S. picta*). All others are polyphagous or in the grey area between poly- and oligophagous (e.g. *S. cilium*), or unknown. Therefore character delimitation is a major issue and because of time constraints I chose not to optimise these traits in any way.

2.6.6. Interpreting Markov chain results

The program Tracer (Rambaut and Drummond 2009) was used to analyse convergence of runs of all MrBayes and BEAST analyses. The estimated sample size (ESS) had to be >100 (Drummond *et al*. 2007) and both runs had to give the same result (equal LnL, TL) to infer convergence of the runs and to call an analysis successful and accept its results.
I explored the utility of the Are We There Yet (AWTY) (Wilgenbush *et al*. 2004) system for graphical exploration of MCMC convergence for some analyses of COI. The program gives good insight in the quality of the MCMC runs, but there are also problems with it. Unfortunately, the program is only available online. Therefore it requires uploading of the files containing all sampled trees, which typically are huge. This takes a lot of time. Moreover, the program needs a lot of time to analyse the data and present the results. Because of these constraints I decided not to use this program anymore. I'd like to recommend to the authors of AWTY to create a (downloadable) stand-alone version of the program. I expect that this would greatly increase its use.

Accepted resulting trees produced by phylogenetic software will be visualized in the program FigTree version 1.3.1 (Rambaut 2006) and included in the thesis report. Some figures were manually adjusted for presentation purposes in InkScape version 0.48.4 (http://inkscape.org/), without changing any actual information.

3. Results

3.1. Properties of the data set

Fifteen specimens, containing seventeen sequences in the data set, should have been discarded from the dataset before analysis:

- littoralis26 (2 sequences) for the unlikely background (*S. littoralis* doesn't occur in Surinam, where the specimen supposedly comes from);
- frugiperda58 (1 sequence) for the non-traceability of the used voucher (this COI is most likely to have been frugiperda56 because that one is lost and the number is similar, but because of this ambiguity it should have been excluded);
- frugiperda22 (1 sequence) because the voucher is not verifiable (the voucher number belongs to a *S. latifascia* specimen, but the sequence is not matching with other *S. latifascia* specimens, but with specimens of *S. frugiperda*);
- latifascia1128, latifascia1129, latifascia1130, eridania1136, dolichos1124, dolichos1125, dolichos1126, dolichos1127, albula1120, albula1122 (1 sequence each) should all have been removed before the analyses for varying reasons, usually because the voucher is not morphologically identified since another specimen of the same series (specimens from the same laboratory strain or specimens collected at the same time at the same location) was already included in the analysis (pers. comm. Marja van der Straten);
- nonSpodoptera1100 (2 sequences) because the specimen is unidentified;
- doli1123 was included twice in the COI alignment, this was caused by the (unnecessary) resequencing of this specimen for this gene, because I overlooked its presence;
- GU094753_exigua was included twice in the COI alignment, once under this name and once under the label PHMO36203_exigua (the Guelph voucher number + species epithet), where GU094753 is the GenBank ID.

I discovered this after analysing the COI and EF1α data sets, so they are still in the results of those analyses. The reader should bear in mind that these should be excluded. I expect that this error has no influence on the final phylogenetic reconstruction, because their sequences are very similar to their conspecifics.

I tried to make data set summaries (percentage invariant sites, percentage parsimony informative sites) of the alignments in DnaSP version 5.10 (Librado and Rozas 2009). This did not work and it never will (with the current version), because the program does not accept ambiguity in the data set: a site can either have an A, C, G T (or U) or be a gap/missing. My data sets contain many ambiguous sites which are IUPAC coded.
To overcome this problem, I used the software package MEGA version 5.05 (Tamura *et al*. 2011). The program requires input in MEGA-format, to which it can convert the alignment itself. The interface of the program has an easy-to-use function for highlighting (in)variant sites and other types of sites.

Table 17. Statistics of the alignments.
Statistics were calculated in MEGA5. Invariant sites are the same as conserved sites.

Gene	Alignment name	Number of terminals	Length (bp)	Invariant sites		Variable sites		Parsimony informative sites	
				#	%	#	%	#	%
COI	210strictSpodCOI	211	657	397	60.4	260	39.6	225	34.2
CytB	60SpodCytB	60	602	395	65.6	207	34.4	162	26.9
EF1α	59SpodEF1a	59	1032	755	73.2	277	26.8	196	19.0
CAD	51SpodCAD	51	849	439	51.7	410	48.3	359	42.3
GAPDH	43SpodGAPDH	43	690	411	59.6	279	40.4	232	33.6
IDH	45SpodIDH	45	711	377	53.0	334	47.0	286	40.2

The taxonomic coverage for *Spodoptera* of my study is 67.7% (21 species out of a total of 31 species). For the Noctuidae my taxonomic coverage is 0.81% (95 out of a total of 11772 species). Figure 15 compares these data between mine and three other relevant studies: Mitchell *et al*. 2006, Zahiri *et al*. 2011 and Kergoat *et al*. 2012. The raw data is listed in table 18.

These differences in taxonomic sampling can largely be explained by the different focus of the studies: Mitchell *et al*. (2006) and Zahiri *et al*. (2011) aimed at resolving the Noctuid (sub)family relationships, while Kergoat *et al*. (2012) and myself aimed at recovering the species level relationships within *Spodoptera*.

Table 18. Taxonomic sampling of three relevant published studies in comparison with the total number of species and our study.

Taxonomic sampling	# species	Mitchell *et al*. 2006	Zahiri *et al*. 2011	Kergoat *et al*. 2012	Lammers
Noctuoidea	42407	146	148	29	97
Notodontidae	3800	7	9	0	2
Noctuidae	11772	106	139	29	95
Spodoptera	31	3	0	24	21

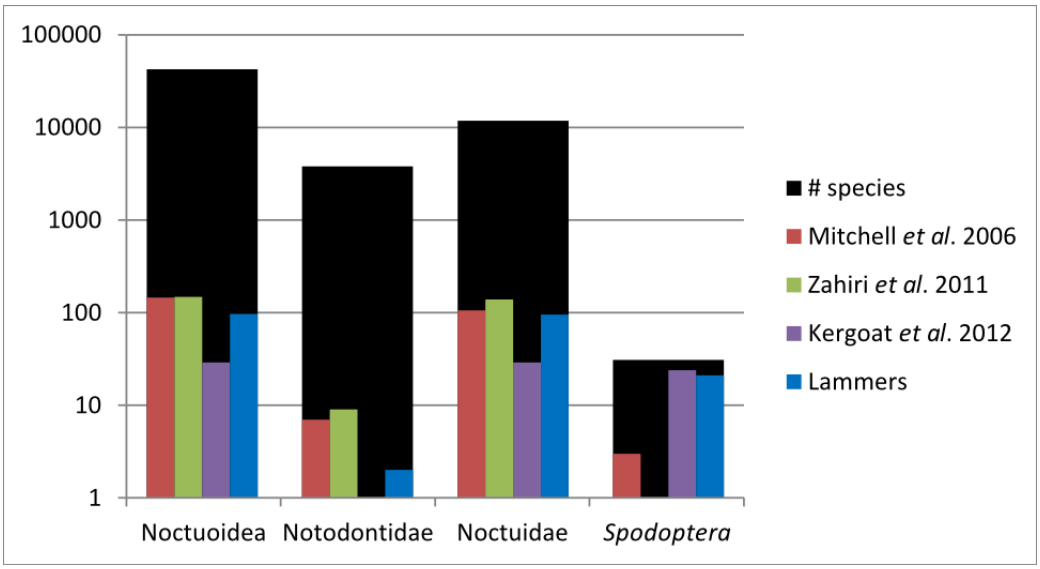


Figure 15. Comparison of taxonomic sampling between two recent studies and our study.
The y-axis is on an logarithmic scale. The black part is the total number of species. The raw data is listed in table 18.

3.2. Gene trees and concatenated analyses

All results from the one-gene RAxML and MrBayes analyses are reported in the appendix, except the COI barcode tree, which is reported here (see figure 16). The results from the analyses of the concatenated alignments created for the research questions are also reported here.
Spodoptera is consistently recovered as monophyletic, except in the ML analysis of EF1α in RAxML. COI does recover *Spodoptera* as monophyletic (when excluding the *S. mauritia* specimens from Fiji, see below), but with low to mediocre support. This is not surprising, for COI has a high number of silent substitutions reaching saturation at the genus level and above that (Hebert *et al*. 2003, Ren *et al*. 2005).
To investigate this problem, I also performed a codon-based analysis in MrBayes on a reduced COI alignment. This indeed results in a higher accuracy at the subfamily level and above, but at a loss of accuracy at the species level, see figure 17 and compare this one to figure 16.
Interestingly, in the COI-barcode tree including all data (see figure 16) *Elaphria* spp. are positioned in various places in the Noctuidae, not as the sister of *Spodoptera+Galgula* and, most importantly, not monophyletic. This tree also shows the problems that arise from the saturation of substitutions: The two Notodontidae species are nested in the Noctuidae. The specimens of *S. mauritia* from Fiji are, based on the DNA sequence divergence, not closely related to the other *Spodoptera* species.

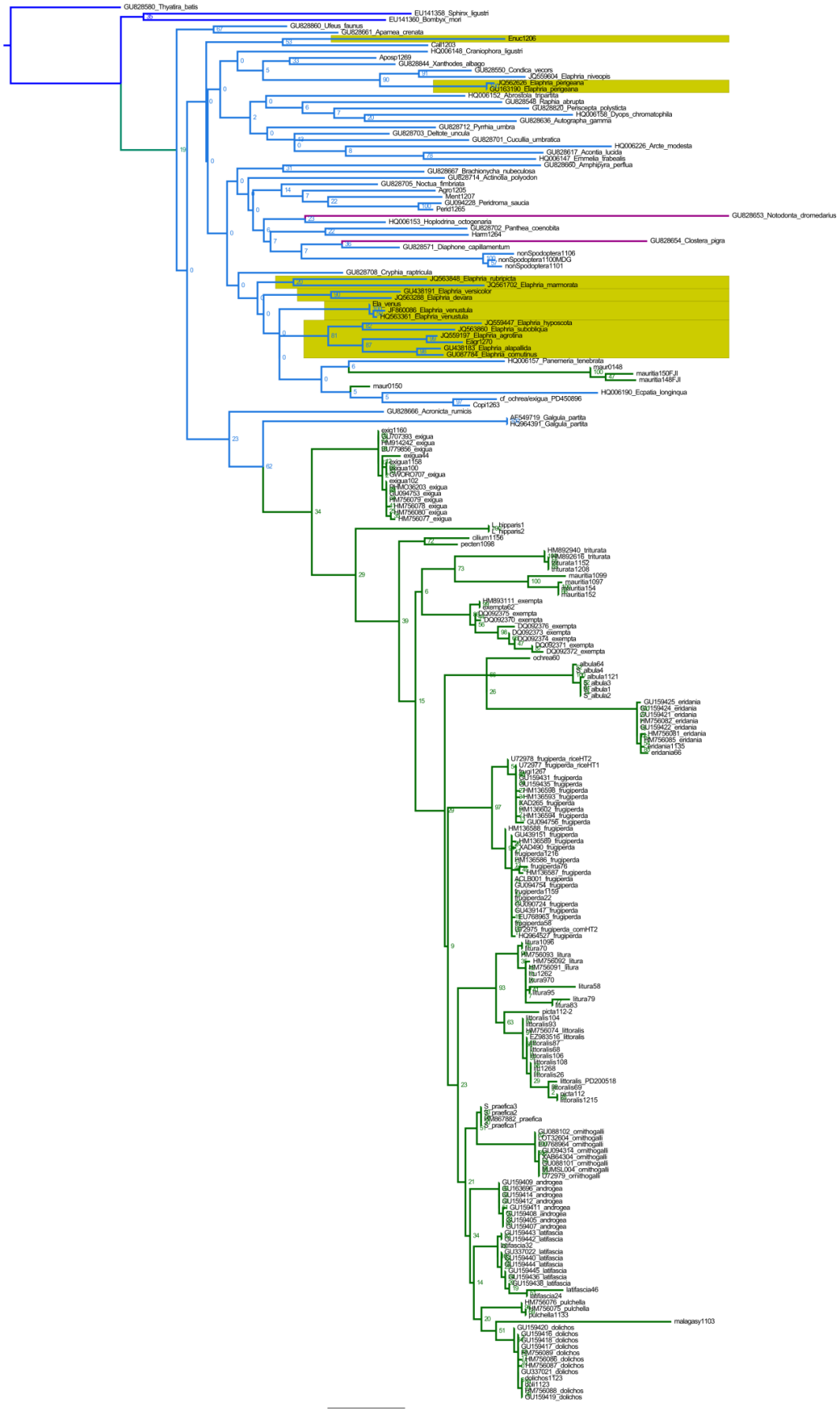


Figure 16. Maximum Likelihood COI-barcode tree. All *Elaphria* specimens are highlighted in yellow.

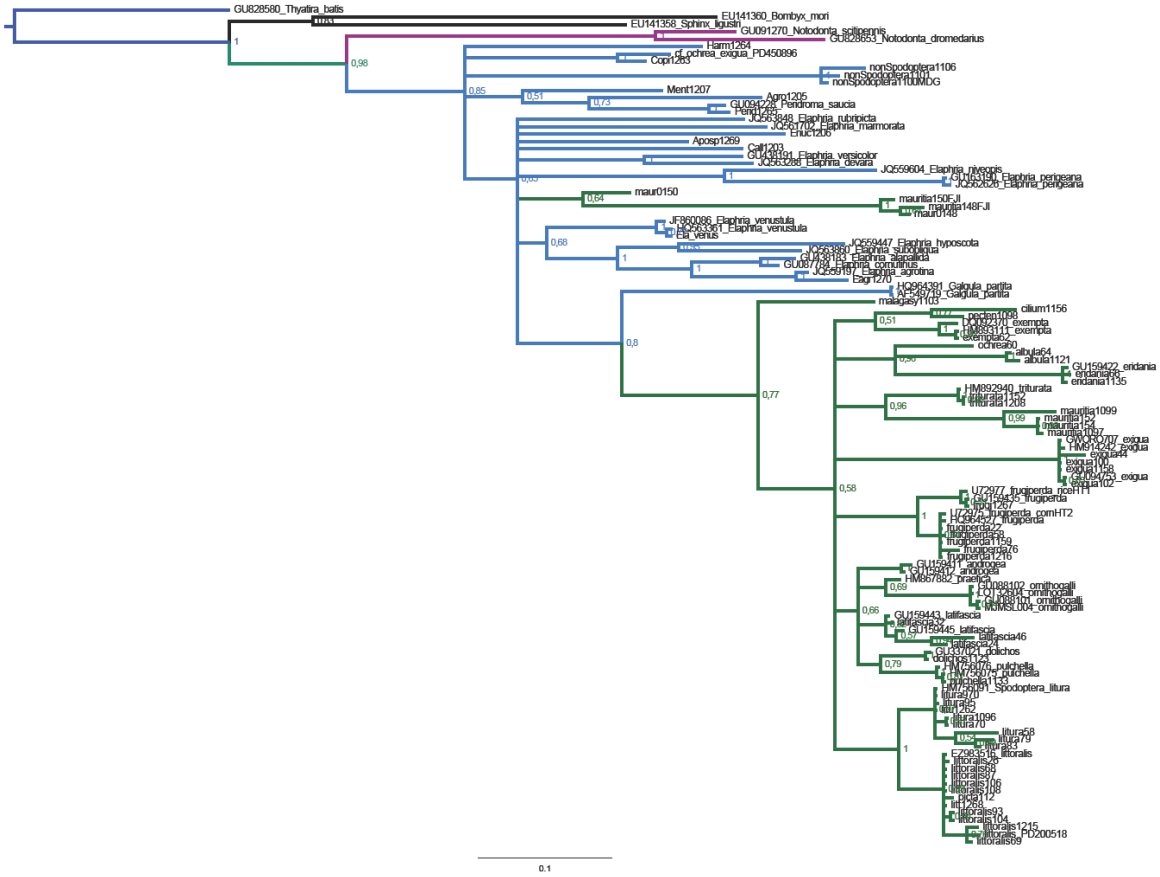


Figure 17. Phylogenetic reconstruction using a codon-based model.
The used alignment was a 116-taxa COI alignment and it was analysed in MrBayes for 190 million generations.

In order to answer the first research question (what clades within *Spodoptera* are discerned) the Concatenated Alignment 1 (49SpodFused.nex) was analysed in MrBayes for 197 million generations.
I discerned clear clades within *Spodoptera*, which I named according to the sampled species in the clades. Figure 18 shows the result of the analysis and defines the clades. Table 19 summarises the appearance and support of these consistently found clades in all performed analyses, sorted from top to bottom by the character sampling. To summarise: POLDA is not always recovered as monophyletic. In those cases LPL is placed within POLDA. The different analyses differ in their power: When an alignment is analysed in RAxML generally fewer clades and lower support is recovered in comparison with analyses in MrBayes/BEAST.

Concatenated Alignment 2 was also analysed in MrBayes for 100 million generations in order to confidently find the position of *Spodoptera* in the Noctuidae, see figure 19. This recovers the same position of *Spodoptera* in the Noctuidae as in Concatenated Alignment 1.

Figure 18. The phylogenetic reconstruction of Concatenated Alignment 1.
a) Base of the tree; b) The *Spodoptera*-part of the tree, defining the clades within *Spodoptera*.

Table 19. Support for the clades of figure 18 in all performed analyses.
***taxonomic sampling is expressed in 2 numbers: [total number of taxa in the alignment] / [number of *Spodoptera* specimens]; low = 1 to few specimens per species; dense = few to much specimens per species**
Posterior probabilities of BI are converted to numbers between 0 and 100.
- = not monophyletic
[number] = monophyletic and with the reported support
n/a = not applicable, meaning that the clade was insufficiently sampled in that gene/analysis

Character sampling	Gene	Taxonomic sampling*	Software	Clades						
				Spod.	EH	MT	PEC	OEA	LPL	POLDA
low: 1 gene	**COI:** 657 bp	low: 56 / 21	BEAST	68	99	-	73	24	100	47
		dense: 221 / 162	RAxML	34	-	73	-	56	93	21
		dense: 116 / 75	Codon in MrBayes	77	n/a	96	51	96	100	66
		dense: 210 / 151	RAxML	55	-	75	-	49	94	23
		dense: 210 / 151	BEAST	90	62	92	100	96	100	-
	CytB: 602 bp	dense: 60 / 57	RAxML	80	n/a	44	93	88	100	21
		dense: 60 / 57	MrBayes	100	n/a	100	100	100	100	-
	EF1α: 1032 bp	low: 59 / 13	RAxML	-	n/a	n/a	n/a	n/a	65	92
		low: 59 / 13	BEAST	94	n/a	n/a	n/a	n/a	83	100
	CAD: 849 bp	low: 39 / 6	RAxML	100	n/a	n/a	n/a	n/a	100	n/a
		low: 39 / 6	BEAST	100	n/a	n/a	n/a	n/a	100	n/a
	GAPDH: 690 bp	low: 32 / 7	RAxML	99	n/a	n/a	n/a	n/a	96	n/a
		low: 32 / 7	BEAST	100	n/a	n/a	n/a	n/a	100	n/a
	IDH: 713 bp	low: 31 / 0	RAxML	n/a	n/a	n/a	n/a	n/a	n/a	n/a
		low: 31 / 0	MrBayes	n/a	n/a	n/a	n/a	n/a	n/a	n/a
medium: 2 genes	**MT:** 1257 bp	low: 32 / 25	MrBayes	85	96	60	97	100	100	81
medium: 4 genes	**NC:** 3282 bp	low: 16 / 9	MrBayes	100	n/a	n/a	n/a	n/a	100	100
high: 5 genes	**NC+COI** 3939 bp	low: 61 / 9	RAxML	100	n/a	n/a	n/a	n/a	100	87
		low: 61 / 9	MrBayes	100	n/a	n/a	n/a	n/a	100	100
		low: 61 / 9	BEAST	100	n/a	n/a	n/a	n/a	100	100
high: 6 genes	**All:** 4539 bp	low: 49 / 26	RAxML	97	53	80	57	98	98	25
		low: 49 / 26	MrBayes	100	87	78	88	96	100	88
		low: 49 / 26	BEAST	100	97	87	100	100	100	100

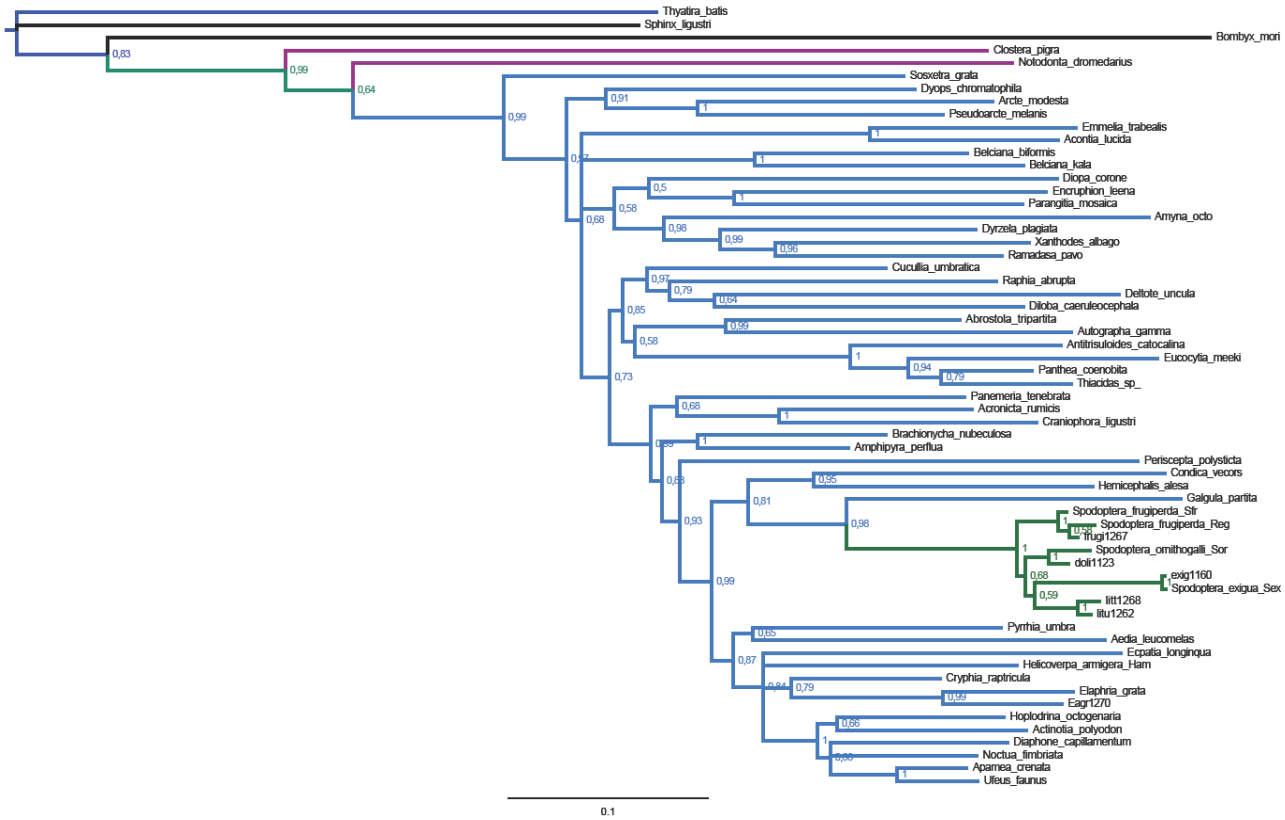


Figure 19. Phylogenetic reconstruction of Concatenated Alignment 2 in MrBayes.

3.3. Interpretation of phylogenetic patterns

3.3.1. *Spodoptera* species relationships

Spodoptera hipparis (former *Leucochlaena hipparis*, see Pogue 2011) is consistently recovered as sister to *S. exigua*. In some analyses (e.g. some analyses of COI), this clade (EH) is recovered as the most basal clade, but in other analyses (e.g. all results of Concatenated Alignment 1) it is sister to MT+PEC, although with low support.
In all analyses performed the LPL clade was well supported, with the exception of analyses of EF1α. In that analysis the support was lower, but nonetheless the clade is still monophyletic. This might be caused by the low amount of variable and parsimony informative sites in this gene (see table 17), especially within *Spodoptera*: Only 16 variable sites, of which 7 parsimony informative, are present in this *Spodoptera* EF1α alignment.

3.3.2. *Spodoptera* genus level and subfamily level relationships

Galgula is consistently recovered as sister of *Spodoptera* in the BI results of both the concatenated alignments (pp 0.95-0.98). *Elaphria* is recovered as far more distantly related from *Spodoptera* than was expected (given Pogue 2002). *Spodoptera+Galgula* is often placed in a clade with *Condica* and *Hemicephalis*.

3.4. Age of Noctuidae and *Spodoptera*

Concatenated Alignment 2 was analysed in BEAST using the calibration points described, in order to infer the age of the Noctuidae and, most importantly, *Spodoptera*. Figure 20 shows one resulting time-calibrated phylogenetic reconstruction. The phylogenetic relationships are consistent with the ones found in the results of the analyses on Concatenated Alignment 1 and the MrBayes-analysis of Concatenated Alignment 2. As far as I know, this is the first across-Noctuidae time-calibrated phylogenetic reconstruction ever created.

Box 1. *Spodoptera* clades

Clade	Species
EH	*S. exigua* *S. hipparis*
MT	*S. mauritia* *S. triturata*
PEC	*S. pecten* *S. exempta* *S. cilium*
OEA	*S. ochrea* *S. eridania* *S. albula*
LPL	*S. littoralis* *S. picta* *S. litura*
POLDA	*S. praefica* *S. ornithogalli* *S. latifascia* *S. dolichos* *S. androgea* *S. pulchella*

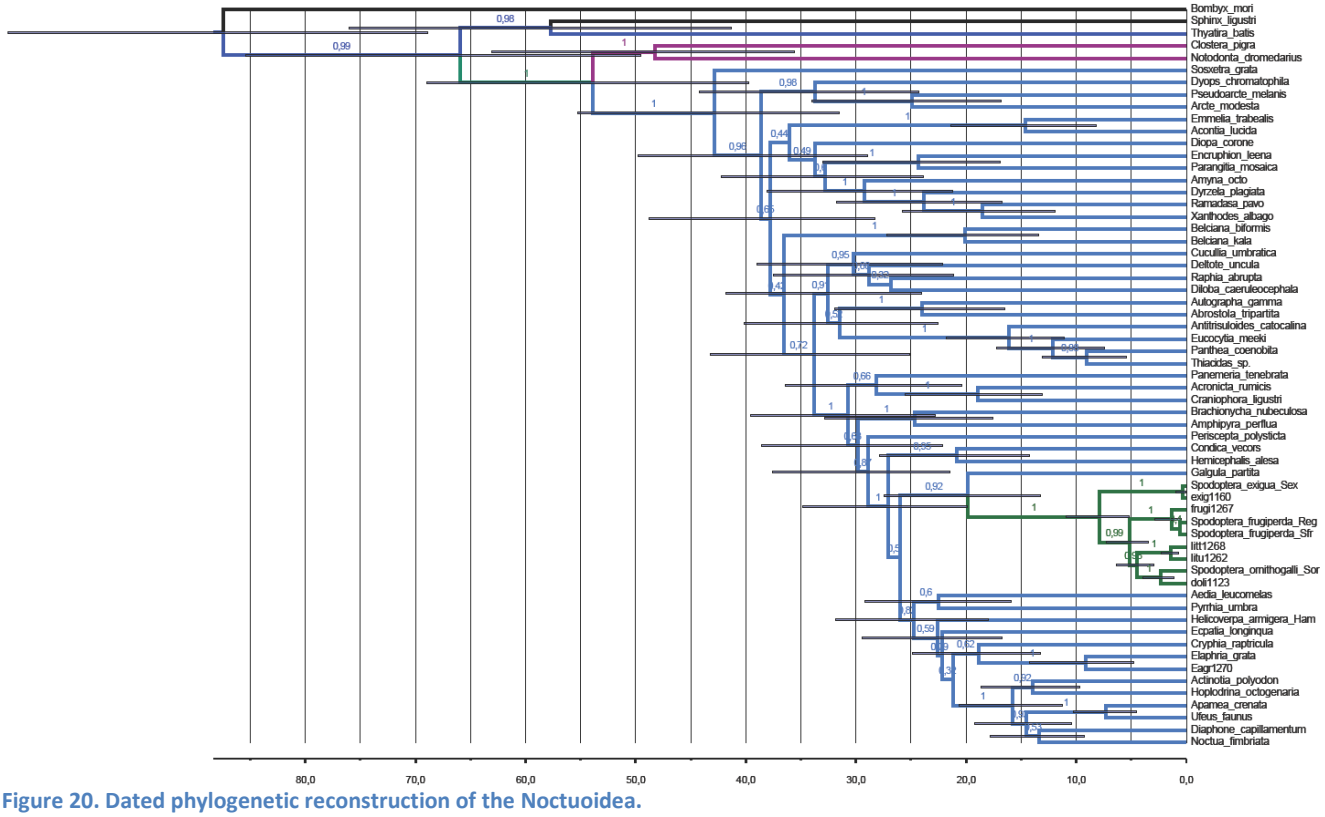


Figure 20. Dated phylogenetic reconstruction of the Noctuoidea.
Concatenated Alignment 2, calibrated with the Notodontid fossil and the (butterfly) bottom-up dating, run for 50 million generations.

Different combinations of calibration points led to different age estimates of the clades. The differences, however, are relatively small. Only when I solely use the Notodontid-fossil calibration point the age estimates are much lower than in any other combination (*Spodoptera* crown age between 2.43 and 5.26 million years). This is logical, because this fossil only provides a minimum age, an age apparently much younger than was expected from the other calibration points. This perfectly explains the younger age estimate from that analysis. Therefore these different age estimates are not in conflict with each other.
The age of the crown of the Noctuidae is estimated between 30.44 and 55.25 million years old when using calibration point combinations. The crown age of *Spodoptera* is estimated between 4.78 and 10.91 million years old. The effect of different calibration combinations on the estimated age of *Spodoptera* is shown in figure 21. Similarly this can be done for the estimated age of the clades within *Spodoptera*, for example the LPL-clade. Figure 22 shows this.

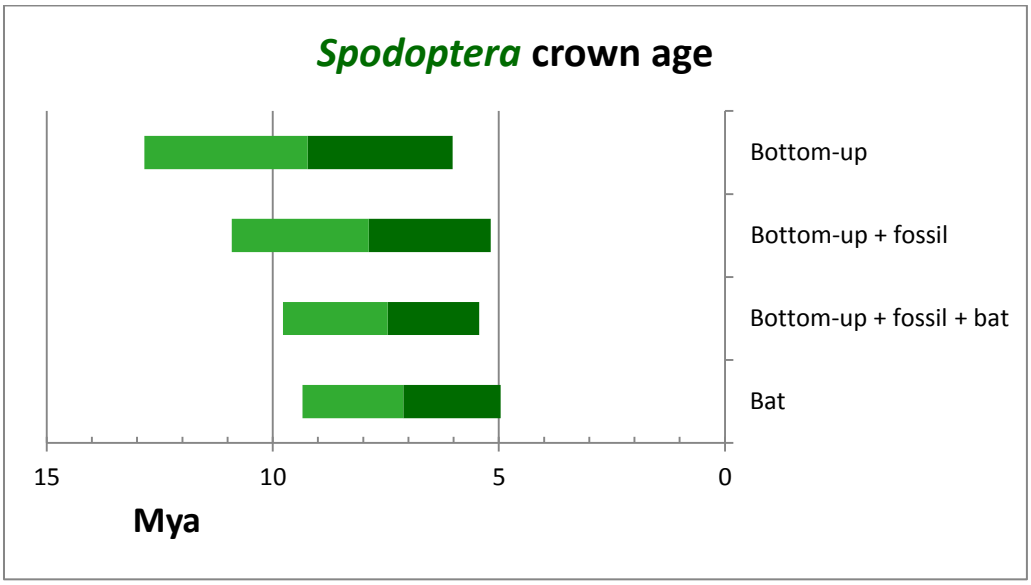


Figure 21. Comparison of the results of different calibration combinations on the estimated age of *Spodoptera*. The bar represents the 95% HPD and has a colour change at the median.

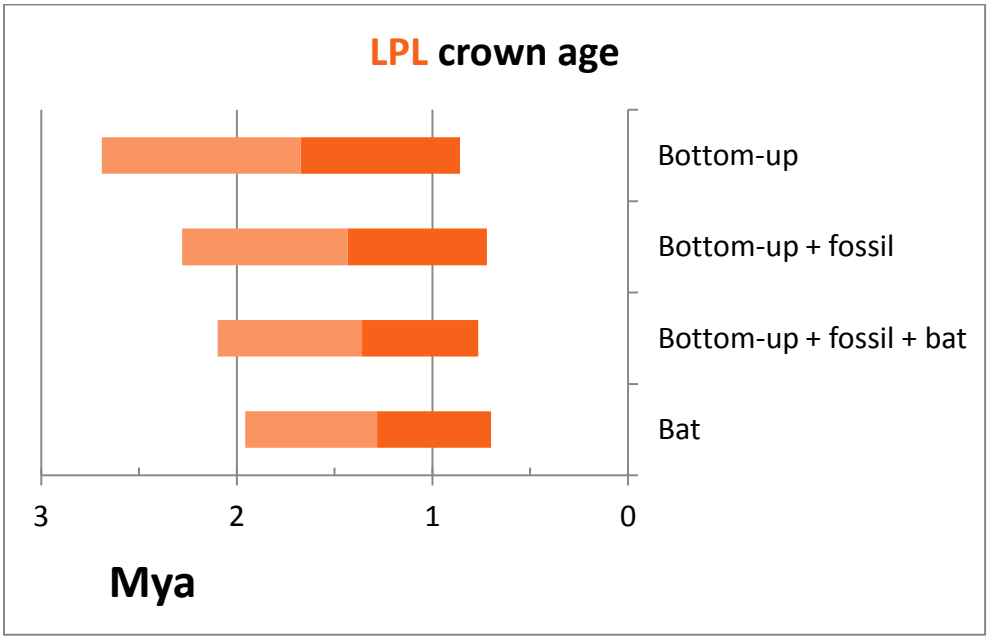


Figure 22. Comparison of the results of different calibration combinations on the estimated age of the LPL clade. The bar represents the 95% HPD and has a colour change at the median

3.5. Biogeographic pattern

I did a maximum parsimony reconstruction in Mesquite, coding for New World versus Old World. I imported the tree resulting from the analysis of Concatenated Alignment 1 in MrBayes in Mesquite, which was then automatically converted to a cladogram-style tree. The common ancestor is estimated to be of New World origin and two clear Old World clades emerge, both nested in this New World backbone, see figure 23. This means most likely that two transoceanic dispersal events brought *Spodoptera* to the Old World, where both radiated independently of their New World relatives. Spodoptera species have good flying capabilities (e.g. Saito 2000) and therefore these dispersal events are plausible and not limited to overland routes (e.g. Beringia, Lafontaine and Wood 1988).

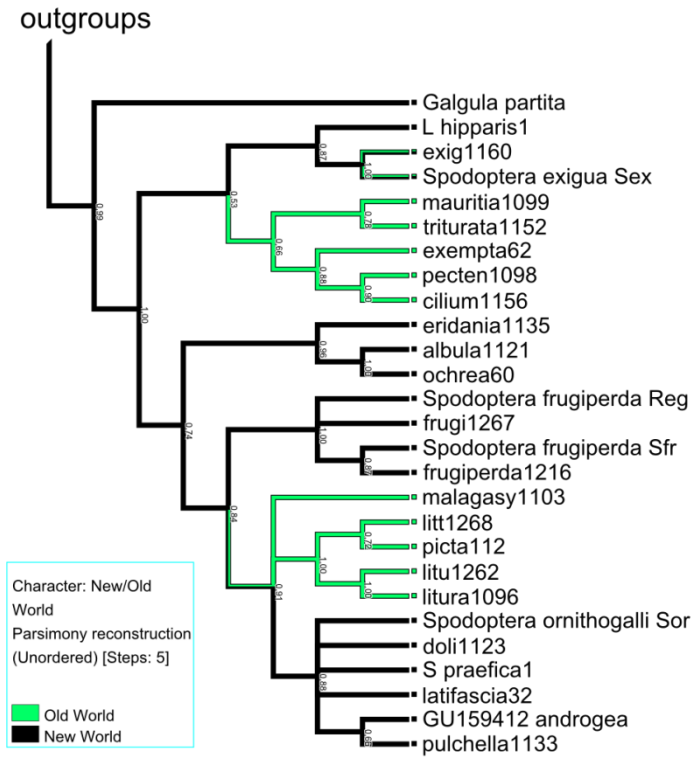


Figure 23. Maximum parsimony reconstruction of the historical biogeography of the lineages of *Spodoptera*. Two Old World lineages are nested in a mostly New World genus. The phylogenetic reconstruction of Concatenated Alignment 1 is imported in Mesquite for this historical biogeographic optimisation.

4. Discussion

At the start of this project I formulated several research goals. (1) We wanted to know what a well-build species-level phylogenetic reconstruction of *Spodoptera* looks like and what distinct clades there are within *Spodoptera*. Furthermore, (2) we wanted to know where to place *Spodoptera* within the Noctuidae. Next, (3) we wanted to know the age of this important pest genus and its clades. In this chapter I reflect on these goals and the hypotheses.

In this study I aimed for a better understanding of the phylogenetic relationships within *Spodoptera*. This was mostly successful, see below. Unfortunately, there was no time left for (proper) historical biogeographic, ecological and morphological analyses. Therefore the three hypotheses on this part are neither rejected nor accepted.

4.1. Phylogenetic reconstructions

Among the results of all ML and BI analyses using the same alignment I found major differences in the support of the nodes. BI usually finds higher support for the same node than ML does. In figure 24 I plotted the calculated support for the nodes from the analyses of the same data set under both ML and BI methods. See table 19 in the results section for the raw data. Every dot represents one node. This clearly visualises the differences: Most nodes lie above the y=0,01x line, meaning that they get a (much) higher support in BI than in ML. I interpret this as that BI has higher analytical power than ML. However, bootstrapping and posterior probabilities are entirely differently calculated, and small differences between the values do not necessarily mean a difference in support (Alfaro *et al*. 2003, Kolaczkowsky and Thornton 2009, Yang and Rannala 2012).

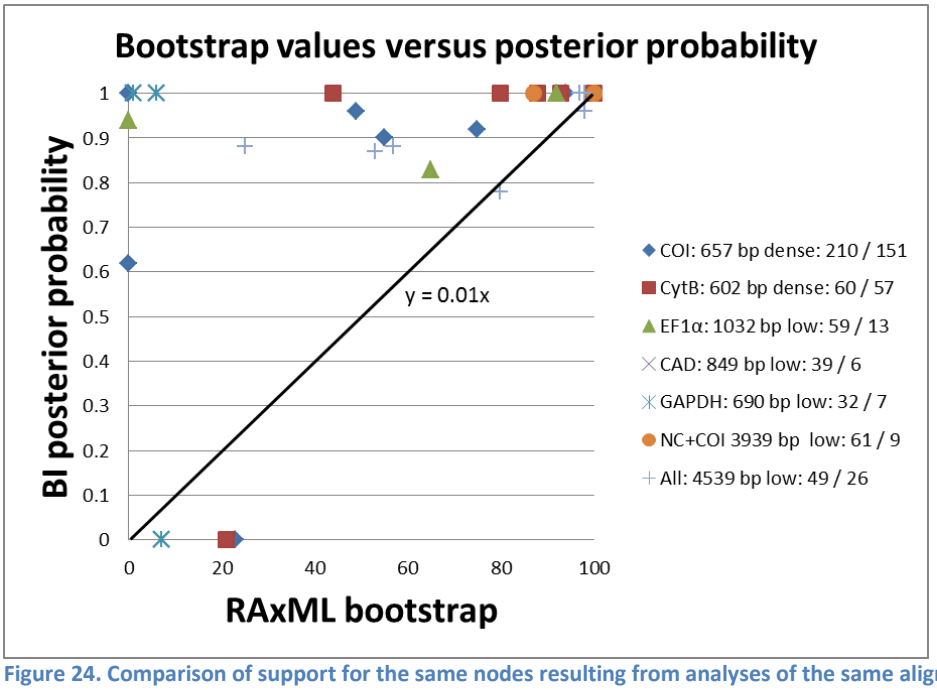


Figure 24. Comparison of support for the same nodes resulting from analyses of the same alignments (see legend) using different methods. Table 19 shows the raw data. Every dot represents one node.

I hypothesised that *S*. *exigua* and *S*. *hipparis* are not *Spodoptera* species, based on DNA-sequence divergence (hypothesis 2). This hypothesis is not supported by my results: *Spodoptera* is recovered monophyletic, including these two sister species. That also rejects hypothesis 3: The *Spodoptera* species excluding *S*. *exigua* and *S*. *hipparis* form a paraphyletic clade; including these two species forms a monophyletic genus. This is consistent with the results of Pogue (2011).

As was noted by previous authors (Pogue 2002, Levy 2002, Prowell *et al*. 2004, Juárez *et al*. 201) I also recovered two strains of the species *S*. *frugiperda* (e.g. see figure 16). These strains are commonly called the rice and corn strains, referring to the host plants where there are often found on. Our sampled specimens are all found on dicotyledons (pers. comm. Marja van der Straten). This species is polyphagous (Pogue 2002) and therefore this is not unexpected. These strains co-exist in the same populations in South America (Juárez *et al*. 2012) and also hybridize (Prowell *et al*. 2004). Whether these strains should be split will mostly depend on the species concept applied (DeSalle *et al*. 2005) and therefore I will not decide on this matter here, because I think that only sequence divergence and putative difference in host preference is not enough.
Since both strains are highly polyphagous I expect them to be of equal harm to economically grown host plants and of equal potential of establishment in Europe. Moreover, only a DNA-test can reliably differentiate between both strains, which currently cannot be performed in the short amount of time available to the relevant authorities (e.g. customs officers, plant protection services). I therefore recommend equal treatment of both strains of this EPPO quarantine species.

My results confirm the close phylogenetic relationship between the LPL clade and the POLDA clade (e.g. Pogue 2002). *S*. *litura* and *S*. *littoralis*, both from the LPL-clade, are on the EPPO quarantine list and intercepted regularly in the Netherlands (van der Straten and Bakker 2011). Neither species from the POLDA clade is on any quarantine list. Two species of this clade, *S*. *dolichos* and *S*. *latifascia*, have been intercepted several times in the last years in the Netherlands (pers. comm. Marja van der Straten). I think that the close phylogenetic relationship of these species with the LPL clade, combined with their similar life history characteristics, would justify the addition of these POLDA clade species to the EPPO quarantine list.

4.1.1. Position of *Spodoptera* in Noctuidae

The first hypothesis, '*Galgula* is sister to *Spodoptera'*, is supported by this study. This confirms the position it got based on morphology (Pogue 2002) and in the publication of Mitchell *et al*. (2006).
The phylogenetic position of *Spodoptera* in the Noctuidae is not fully resolved yet. *Spodoptera+Galgula* groups with *Condica+Hemicephalis* in the phylogenetic reconstructions, but this node has only mediocre support: 0.75 posterior probability in concatenated alignment 1 and 0.81 posterior probability in concatenated alignment 2. This I find is not high enough to support a classification of *Spodoptera* into the Condicinae *sensu* Mitchell *et al*. 2006. Moreover, remember that my taxonomic sampling of the Noctuidae is just 0.81 per cent of all described species. Therefore it is likely that some other closer relatives of *Spodoptera* are missing in my data set. Adding these taxa (see Lafontaine and Schmidt 2010 for likely candidates) will probably resolve these subfamily-level relationships.

4.1.2. Calibrations

Perhaps it was better to place the Notodontid fossil calibration at them stem of the Notodontid family, because that normally is the appropriate way of placing fossils with synapomorphies for a certain clade (pers. comm. Lars Chatrou, Ho and Philips 2009). This would, however, not have made a great difference, because the branch leading to the Notodontidae is very short (see figure 20). This means that the Notodontidae would only be estimated to be slightly younger.
In figure 25 I visualise the differences between the prior probability distribution shapes and the posterior distribution of the sampled ages for the same three calibrated nodes. They do not fully overlap, as would be expected in a perfect congruence between calibration points and the age of real phylogeny. But because the bottom-up dating point provided only a conservative maximum age and the Notodontid fossil provided only a minimum age for that family, I conclude that there is no conflict between the prior and posterior probability distributions.

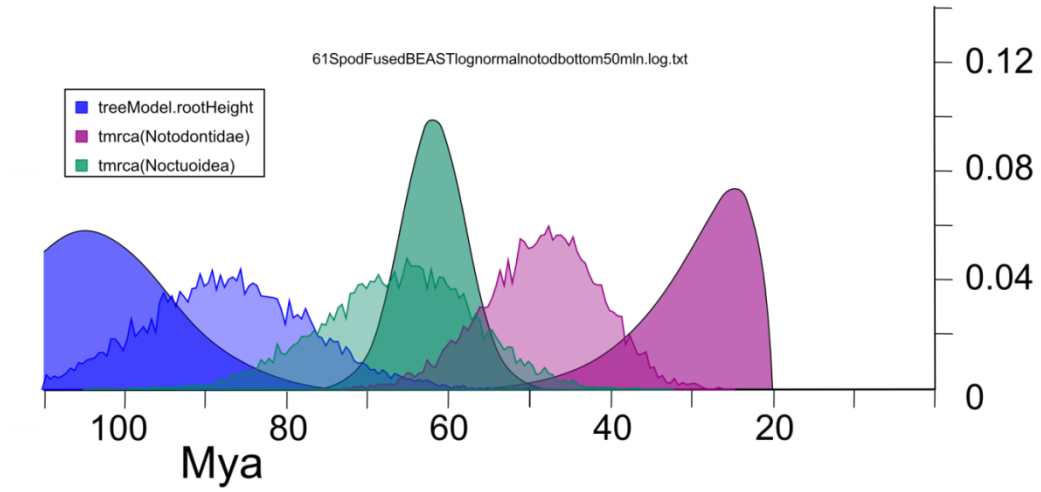


Figure 25. Differences between prior probability distribution shapes and posterior probability distribution shapes of the three calibrated nodes.

4.1.3. Comparison with the literature

A comparison between my results and Kergoat *et al*. (2012) deserves a separate chapter, because these two studies cover mostly the same subjects. See chapter 5 for a thorough comparison of the results.

4.1.3.1. Morphological cladogram and my molecular phylogenetic reconstructions

In my analyses I recovered mostly the same clades as in the morphological cladistics study by Pogue (2002). Clades EH, OEA and LPL are exactly the same. POLDA is slightly different, but this might be an artefact caused by the wider taxonomic sampling of Pogue (2002). MT is also present in Pogue (2002), but its sister species is *S. exempta*, a member of my PEC clade. This clade is not present in the cladogram, these species take polyphyletic positions. Also the position of *S. malagasy* is very different in Pogue (2002): There it is sister to part of the PEC-clade, while in my molecular phylogenetic reconstruction it is sister to LPL+POLDA. In my phylogenetic reconstruction the most plausible explanation of the relationship of *S. malagasy* would be that its ancestor was in the same transoceanic dispersal event as the LPL clade, although they are in a polytomy with POLDA. In the tree topology of Pogue is *S. malagasy* nested in the group of the other transoceanic dispersal event.

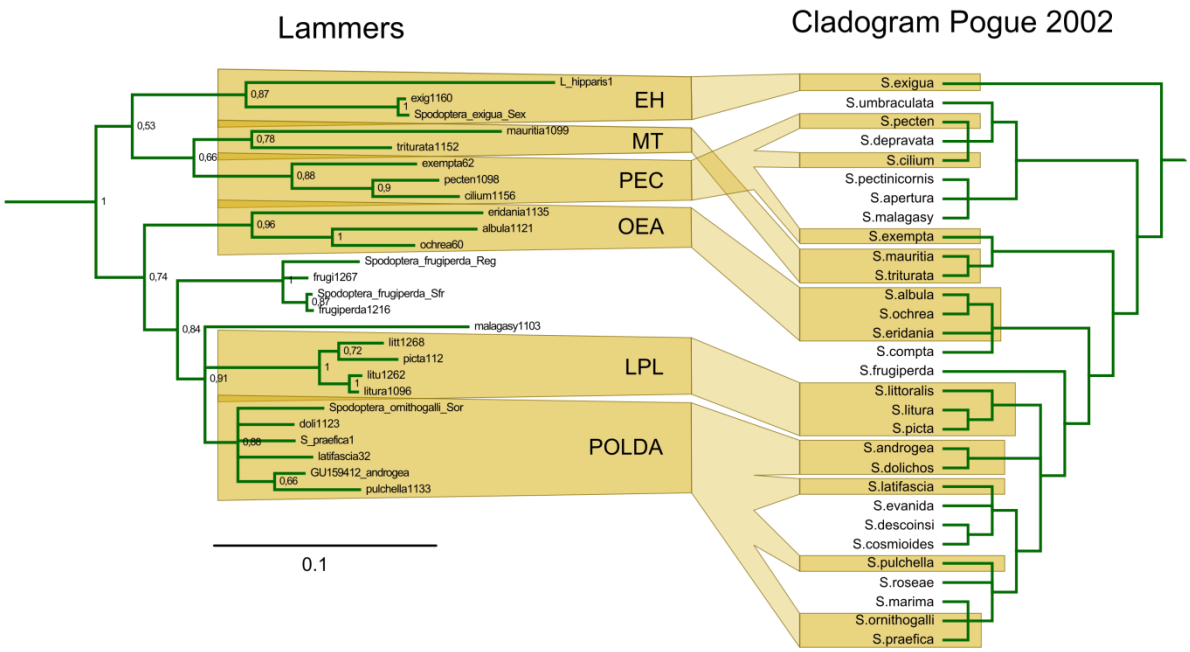


Figure 26. Comparison of the tree topology from my phylogenetic reconstruction of Concatenated Alignment 1 with the strict consensus cladogram of the morphological MP analysis of Pogue (2002).

4.1.3.2. Published alignment compositions and substitution rates

In table 20 I compare the number and percentage of variable and parsimony informative sites with those found in Wahlberg and West Wheat (2008). They investigated the usefulness of these markers for phylogenetic reconstructions. My recovered percentages are not entirely similar, but nonetheless not very different. These differences are the results of the taxonomic sampling of the data set: Whether a site is classified as variable and / or parsimony informative depends on the taxa sampled and their reciprocal (taxonomic and genetic) distance. Sampling more distantly related taxa results in a sampling of a longer evolutionary history and therefore a longer time for accumulation of substitutions, leading to a higher percentage of variable sites.

Table 20. Comparison of the number and percentage of variable and parsimony informative sites with the original publication of the markers.

	Variable sites		Parsimony informative sites	
Gene	**Wahlberg and Wheat 2008**	**Lammers**	**Wahlberg and Wheat 2008**	**Lammers**
COI	712 (48%)	260 (39.6%)	571 (38%)	225 (34.2%)
CAD	458 (54%)	410 (48.3%)	401 (47%)	359 (42.3%)
EF1α	482 (39%)	277 (26.8%)	390 (31%)	196 (19.0%)
GAPDH	272 (39%)	279 (40.4%)	248 (36%)	232 (33.6%)
IDH	337 (47%)	334 (47.0%)	284 (40%)	286 (40.2%)

The recovered rates of substitution per million years per site over the tree were between 0.00173 (EF1α) and 0.0122 (COI), see table 21. This is, for all genes except COI, lower than the assumed average insect mitochondrial molecular clock of 0.02 substitutions per site per million years (Hebert *et al*. 2003). However, many different estimates are published: Brower (1994) found an average rate of 0.0115 for mitochondrial DNA; Simonsen *et al*. (2011) reported an average rate of substitution of 0.01909 (over 5 mitochondrial plus 2 nuclear genes); Quek *et al*. (2004) assumed a COI substitution rate of between 0.0137 and 0.0153 per site per million years based on a review of several previous studies (see the references there). Todisco *et al*. 2012 calculated an average rate of 0.086 substitutions per site per million years for COI, which is much higher than all the previous estimates. A major difference between this study and the previously listed is, however, that their study was at the population level, instead of above the species level. Papadopoulou *et al*. (2010) nuanced the differences between all published estimated substitution rates in insects: the estimates are dependent on which marker is sequenced, the partitioning of the data, the mixing of interspecific and intraspecific data, the substitution model, the reliability of the calibration and other factors. Their calculations resulted in an estimation of the substitution rate per site per million years of 0.0131 ± 0.0013 for mitochondrial DNA and 0.0012 ± 0.0003 for nuclear DNA. These rates are congruent with the rates I found, see table 21 and figure 27.
This would imply that the ages I estimated for the clades of *Spodoptera* are likely to be close to the true clades' ages.
Different rates of substitution per marker are actually very useful. They will lead to different resolutions at different phylogenetic depth (Wahlberg and Wheat 2008, Wilson 2010).

Table 21. Rates of substitution in my time-calibrated phylogenetic reconstruction of Concatenated Alignment 2. These statistics are calculated for the analysis using the fossil and bottom-up calibration points.

Summary statistic	**61SpodCOI .meanrate**	**61SpodCAD .meanrate**	**61SpodEF1a .meanrate**	**61SpodGAPDH .meanrate**	**61SpodIDH .meanrate**
mean	$1.22 * 10^{-2}$	$7.29 * 10^{-3}$	$1.73 * 10^{-3}$	$3.28 * 10^{-3}$	$4.45 * 10^{-3}$
stderr of mean	$2.37 * 10^{-4}$	$4.61 * 10^{-5}$	$6.13 * 10^{-6}$	$1.21 * 10^{-}$	$1.66 * 10^{-5}$
median	$1.18 * 10^{-2}$	$7.15 * 10^{-3}$	$1.71 * 10^{-3}$	$3.23 * 10^{-3}$	$4.39 * 10^{-3}$
geometric mean	$1.20 * 10^{-2}$	$7.19 * 10^{-3}$	$1.72 * 10^{-3}$	$3.24 * 10^{-3}$	$4.41 * 10^{-3}$
95% HPD lower	$7.77 * 10^{-3}$	$5.07 * 10^{-3}$	$1.27 * 10^{-3}$	$2.34 * 10^{-3}$	$3.25 * 10^{-3}$
95% HPD upper	$1.79 * 10^{-2}$	$9.69 * 10^{-3}$	$2.25 * 10^{-3}$	$4.19 * 10^{-3}$	$5.76 * 10^{-3}$
effective sample size (ESS)	136.63	707.85	1775.63	1662.67	1587.59

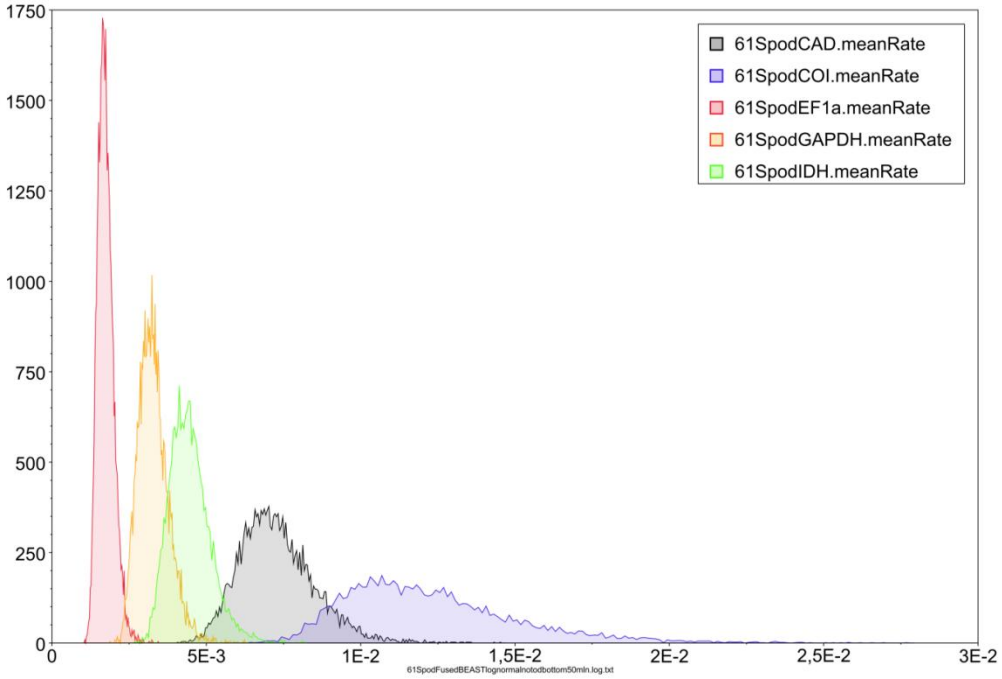


Figure 27. Marginal densities of the five markers used in the time-calibrated phylogenetic reconstruction of the Noctuoidea.

4.1.4. Unexpected COI patterns

4.1.4.1. *Elaphria* not monophyletic

Elaphria, the proposed sister of *Spodoptera+Galgula* (e.g. Pogue 2002, Mitchell *et al*. 2006, Pogue 2011) is, based on DNA sequence divergence, far more distantly related from *Spodoptera* than expected. Moreover, *Elaphria* is probably not monophyletic (see figure 16 in the Results section). Pogue and Sullivan (2003) already noted that the genital morphology is also very dissimilar across *Elaphria* species and therefore they expected *Elaphria* to be polyphyletic. This seems indeed to be the case. Therefore I recommend a full revision of this genus.

4.1.4.2. *S. mauritia* population of Fiji

The specimens of *Spodoptera mauritia* from Fiji show very high DNA sequence divergence from the rest of the *Spodoptera* species in COI, much more sequence divergence than expected from their morphology (based on their external and genital morphology they are indeed *S. mauritia*). See figure 16 for the position of these specimens in the COI barcode tree.

Several causes might explain this pattern. It could be that for some reason this population has an elevated ds/dn-ratio, meaning that silent substitutions occur more frequently than in other populations. Another explanation could be that the sequenced gene is actually a type of pseudogene, also known as Nuclear Mitochondrial DNA (NUMT) (Lopez *et al*. 1994, Bensasson *et al*. 2001), but these seem to be rare in insects (Richly and Leister 2004). If this is the case, it would certainly explain the high amount of substitutions because of the lack of selection on such a non-functional gene.

The double peaks observed in many sites in the sequence tracers of these specimens could also be explained with heteroplasmy: the occurrence of the genotypically different mitochondria in one cell (e.g. Chinnery *et al*. 2000). This phenomenon is not well investigated and seems to be rare, because of the maternally inheritance and a bottleneck effect in the early oogenesis (Chinnery *et al*. 2000, Goto *et al*. 2011), but at least one exception to this rule is described: a 30 million year old stably inherited heteroplasmy in isopods (Doublet *et al*. 2008). I don't expect heteroplasmy to be a likely explanation in the *S. mauritia* population of Fiji. The sheer number of heterozygoteous sites is not likely to have evolved in the few millions of years that the species is present. That, combined with the knowledge of the bottleneck in oogenesis and thereby the transmission of only few mitochondria, makes it highly unlikely that this number of polymorphic sites remain present in a population of a species.

Another possible explanation could be that in this population the interaction between the subunits of the Cox-protein works slightly differently, resulting in a different quaternary structure and thus other physical constraints on protein shape (Puslednik *et al*. 2012). This would be reflected in a different sequence of the COI (barcode) region, coding for different amino acids than in other populations or species. Indeed, the sequences of these specimens show between four to six amino acid changes (maybe even more in the ambiguously sequenced regions) compared to their conspecifics from other populations. However, I think it is unlikely that an ancient, functional electron transporting protein like Cox can evolve a different, still functional, quaternary structure. Much more specimens from this and other populations need to be sampled to test this hypothesis. Of all the previously described possible explanations I think the most likely one is that the sequenced gene is actual a pseudogene and not the functional mitochondrial gene. This could be tested by isolating the mitochondria from the specimens and sequencing them. And then simultaneously sequence its nuclear genome counterpart. If both give a positive result, with a different sequence, it is most likely that our COI sequences of the *S*. *mauritia* specimens from Fiji represent a NUMT.

4.1.5. *Wolbachia*

Recently *Wolbachia*-infected specimens of *S*. *exempta* were discovered (Graham and Wilson 2012, 2013). This is not surprising, because 66% of all insect species is estimated to be infected with this alpha-proteobacteria (Hilgenboecker *et al*. 2008). *Wolbachia* are maternally transmitted and are able to influence host reproduction and sex determination (Cordaux *et al*. 2011). This has impact on the evolutionary history of insects: parthenogenesis and other modifications of gender determination (Cordaux *et al*. 2011), enhanced or decreased susceptibility to viruses (Graham *et al*. 2012) and selective sweeps (e.g. Jiggins and Hurst 2011, Graham and Wilson 2012, 2013). This might have an influence on the usefulness of COI as a barcode region, because of the co-transmission of the COI-haplotype with the *Wolbachia*-strain to the offspring of an infected mother (Smith *et al*. 2012). Therefore the recovered phylogenetic pattern from mitochondrial markers might not directly reflect the evolutionary history of the species (Ballard and Whitlock 2004).
I do not know whether our sequenced specimens were infected. This would be worth investigating in future research because of the impact it might have had on the evolutionary history of *Spodoptera*.

4.2. Evolutionary trends in *Spodoptera*

It is important to note that all my calibration points in the dating analysis are outside of the Noctuidae. Therefore the only information provided to the model for the estimated node depth is the number of substitutions over the tree topology in the Noctuidae, while there is no internal calibration of changes of substitution rates over time in this family. This means that the substitution rate on all branches in the Noctuidae will be estimated in the same range.

4.2.1. Echo-location of moth predators

Not only bats use echo-location; also some (modern day) insectivorous shrews are capable of doing so (Siemers *et al*. 2009). However, shrews turn out to be an entirely separate lineage, the Eulipotyphla, which are sister to the Chiroptera + Fereuungulata (which is a new name for Cetartiodactyla + Perissodactyla + Carnivora + Pholidota) (Meredith *et al*. 2011, Zhou *et al*. 2011). They are an older lineage in Mammalia than the bats are (Zhou *et al*. 2011). This would mean that the evolution of echo-location is possibly older than the crown age of all bats. However, the shrews probably do not use their independently evolved echo-location for detection of prey, but nonetheless the moths could have adapted to avoid the calls of shrews. Moreover, I do not expect the Lepidopteran tympanal organs to have evolved immediately after the emergence echo-location. Instead, I expect a major time lag between these events. Therefore there is a considerable margin of error associated with this type of dating and a dating study should never rely on ecological calibration points alone.

Similar to the used ecological calibration here, with an extended data set, the emergence of the proboscis at the crown of the Glossata could be used as a calibration point, assuming that this tongue evolved in a response to the emergence of angiosperms (pers. comm. Don Lafontaine). This method could similarly be applied to other insect orders with tympanal organs, for example the Orthoptera (e.g. Plotnick and Smith 2012).

4.2.2. Biogeography and human agriculture

With my time-calibrated phylogenetic reconstruction I estimated that the crown of the Noctuidae is between 30.44 and 55.25 million years old and that the crown of *Spodoptera* is between 4.78 and 10.91 million years old. This means that the Noctuidae started radiating in the late Eocene or early Oligocene. *Spodoptera* radiated in the late Miocene. At that time the configuration of the continents was similar as it is today. Therefore continental segregation was not a factor in the speciation of *Spodoptera* or any other Noctuid genus.

Human agriculture emerged approximately 40,000 years ago, while the youngest splits between crop pest species in my data set are at least 750,000 years old. Therefore we can reject the influence of human agriculture on *Spodoptera* speciation. This does not mean that the species later cannot have switched their host preference to crops and thereby possibly expanded their distribution. This hypothesis would require a different data set, sampling many populations of the same species, testing their host preference and sequencing them for markers suitable for below the species level.

One possible explanation of a factor involved in *Spodoptera* speciation is revealed in the biogeographic pattern in the genus. There must have been at least two transoceanic dispersal events between the New World and the Old World. After these events the new clades independently diversified in their new residence.
Transoceanic dispersals are not uncommon in Lepidoptera. Many butterfly genera, for example, have a Holarctic distribution: *Papilio* (Sperling and Harrison 1994), *Parnassius* (Todisco *et al*. 2012), *Polygonia* (Wahlberg *et al.* 2009), *Vanessa* (Wahlberg and Rubinoff 2011), *Boloria* (Simonsen *et al*. 2010), *Lycaena* (van Dorp 2004) and *Polyommatus* s.l. (Vila *et al*. 2011) are all well-described cases. This means that *Spodoptera* is not an exception in their distribution pattern.

4.2.3. A world of grasslands

In the Eocene the world was rapidly cooling (Zachos *et al*. 2001). In this Epoch the first major open grasslands appeared (Strömberg 2011). At this time I estimated the radiation of the Noctuidae. Many Noctuidae feed on grasses or herbs of open grassland (see Mitchell *et al*. 2006 and references therein). This, to me, doesn't seem to be a coincidence: When more potential food becomes available to herbivorous caterpillars, their populations are likely to expand (Schoonhoven *et al*. 2005). This will likely lead to a (host plant mediated) radiation of these herbivores and possibly an evolutionary trend towards monophagy and/or co-evolution with their hosts (Schoonhoven *et al*. 2005).
These high numbers of moths were likely an attractive source of protein and fat for flying insectivores like bats. Indeed, the major radiation of bats also occurred in the Eocene and is hypothesised to be in a response to

increased prey diversity (Teeling *et al*. 2005, Simmons *et al*. 2005). This is congruent with my estimation of the age of the radiation of the Noctuidae. This tritrophic model involving bottom-up and top-down interactions could explain the large diversity of each trophic level involved.

In the Miocene the C4-grasses radiated and expanded, largely replacing C3-grasses in many grasslands (Zachos *et al*. 2001, Sage 2004, Edwards *et al*. 2010). C4-grasses are a preferred host plant for many *Spodoptera* species (Pogue 2002). This emergence of C4-grasses is contemporary with the radiation of *Spodoptera* in the Miocene. I therefore hypothesise that the emergence of the ubiquitous C4-grasses played an important role in *Spodoptera* speciation.

4.3. The future and the role of *Spodoptera*

The global human population grows rapidly, leading to increased pressure on agriculture to produce enough food. Any cause of a dip in food production will be quelled as soon as possible. One of the challenges for agriculture is to deal with herbivorous insects competing for food with us. Of these, *Spodoptera* species play an important role on a global level. The excellent dispersal capability (both as larvae and as adults) of many species, combined with an extremely high degree of polyphagy, a high fecundity and short generation time, creates the perfect crop pest species.

These crop pests encounter many different environments and the general pattern seems to be that they cope well with that. From an evolutionary perspective they are a successful Noctuid lineage. The major questions are: How do they do it? What changes in their ancestors made these adaptations possible? How can polyphagous *Spodoptera* species cope with all the different poisons they encounter?
Possibly it has many pathways for detoxification, or a method of preventing uptake of harmful compounds, or perhaps it just resorbs little from its food and defecate the rest. Much more future research needs to be done here.

Elucidating this problem will require a multidisciplinary approach in order to understand the functional genomics of *Spodoptera* crop pest species. It seems likely that the gene expression will differ between species and also between individuals of the same species in a different environment. By sequencing the total mRNA of species of interest in different stages of development is a way of measuring gene expression (RNA-seq, also called transcriptomes, Wang *et al*. 2009). This has great potential for answering the above-mentioned questions.

Comparing the transciptomes of the larvae of a polyphagous species with a related mono- or oligophagous species will provide insight in the relative importance of the expression of key genes. Candidate key genes will be the genes relating to detoxification of poisons and genes related to the formation of digestive enzymes. Ultimately I hope that this will lead to a thorough understanding of the host specificity and invasive potential of these crop pests in our changing world where food for man is likely to become scarce. This will be the focus of a PhD project for which I hope to get funding from the Experimental Plant Sciences Graduate School and other sources.

5. Kergoat *et al*. 2012

When I started my thesis, no publication of a properly tested (i.e. using several independent molecular markers) phylogenetic hypothesis for *Spodoptera* was available. While investigating the availability of the data in GenBank I found some unpublished 16S rRNA sequences of several *Spodoptera* species. The listed authors were J.F. Silvain, D. Vautrin and M. Solignac. Through e-mail contact between Freek Bakker and the first author we learned that Kergoat *et al*. were in the process of submitting a paper containing a phylogenetic reconstruction of *Spodoptera*, a biogeographic analysis and evolutionary relationships with host plants. Later we also learned that they estimated divergence times. The paper was accepted in August 2012 and shortly thereafter published online. In this chapter I review this paper and compare its results with mine.

5.1. Review

The paper treats four different subjects: phylogenetic reconstructions using maximum parsimony and Bayesian Inference methods, a host plant optimization using the maximum likelihood model Mk1 (Lewis 2001), dating of their phylogenetic reconstruction using Bayesian relaxed clock (BRC, Drummond and Rambaut 2007) and Penalized Likelihood (PL, Sanderson 2002, 2003) methods, and historical biogeography analyses using the dispersal-extinction-cladogenesis maximum likelihood method (Ree *et al*. 2005, Ree and Smith 2008). Here I treat these subjects in this order and I'll end with some concluding remarks.

Taxonomic sampling

Kergoat *et al*. (2012) seem unaware of the study published by Pogue (2011): nowadays there are 31 known *Spodoptera* species, instead of the figure of 30 they mention throughout the paper. This figure was based on Pogue 2002, see this thesis's Introduction. This explains why the paper doesn't treat the relationships of *S. hipparis*.

The molecular taxonomic sampling of Kergoat *et al*. (2012) includes 24 species. Of three other species they failed to amplify any gene. They achieved their proclaimed complete taxon sampling by introducing the morphological characters of Pogue (2002, appendix 1). However, I showed that there is incongruence between the morphological cladogram of Pogue (2002) and my molecular phylogenetic reconstruction (e.g. the position of *S. malagasy* and the PEC-clade). I question the validity of this method, since the position of the representatives of the PEC-clade, most notably *S. exempta*, is the same in the molecular phylogenetic reconstruction of Kergoat *et al*. (2012) as in mine. In many other cases it might results in a total 'true' phylogeny, but in this case some taxa might end up in the wrong position.

They place *Spodoptera* in the subfamily Amphipyrinae. It is not specified whereon this is based. This would not be important, if they had not based their outgroup sampling on this assumption. The assumed position of a genus influences which close relatives will be sampled, and which close relatives are sampled influence the possibility of testing the genus' monophyly. They state that *Spodoptera* is recovered monophyletic, which is without doubt in this data set, but they do not have the proper taxon sampling to answer that question.

Character sampling, choices and consequences

The authors claim that they overcome possible mitochondrial introgression bias by sequencing two nuclear genes. However, their character sampling is mitochondrial marker biased: four of the six sequenced markers. Mitochondrial markers are not independent sources of data, because of the sole maternally inheritance and lack of recombination. This means that the majority of the data set represents solely the evolution of the mitochondrial genome and thus not a balanced data set representing the evolutionary history of the species.

Of the nuclear markers, the marker EF1α has only a very low sequence divergence: I downloaded the *Spodoptera*-part of the EF1α alignment of Kergoat *et al*. and calculated the percentage parsimony informative sites, which is 5.1% (0.7% parsimony informative sites in the *Spodoptera* part of my EF1α alignment). I therefore concluded that it is an unsuitable marker for Lepidoptera species level systematics (see chapter 6: Recommendations). They sequenced one other nuclear marker (28s rRNA), which has only 3.2% parsimony informative sites in the *Spodoptera*-part of the alignment, thus less than EF1α. These two nuclear markers, with their few parsimony informative sites, has to counterweigh the overwhelming phylogenetic signal from the mitochondrial genome. Therefore they actually only have good data from 'one' independent marker (the mitochondrial genome), which is not enough in species delimitation (Dupuis *et al*. 2012).

This results in few conflicts between the markers, which likely explains the estimated >0.98 posterior probability on all interspecific nodes.

The authors decided "to generate a molecular dataset with only one representative per *Spodoptera* species. Instead of randomly picking up a specimen per species, we used consensus sequences to have a better

representation of species genetic variations." I think this approach is fundamentally wrong. The terminals in the phylogeny should unambiguously represent observed data. A consensus sequence for a species is not observed data, it is a data summary, it is metadata. The construction of these consensus sequences will be uninformative when a specimen is misidentified and is included in the construction of a consensus sequence of the wrong species. Moreover, the resulting consensus sequences suffer from a sampling bias: When only part of the geographic distribution of a species is sampled, only that part of the species will be represented by the consensus sequence. All these issues could have been avoided by picking one specimen to represent the species, or by including all specimens in the analysis. They state that the latter should not be done in a dating study because that would lead to a systematic overestimation of recent divergence times. However, this should not be a problem when a proper model of evolution is used.

Phylogenetics

The authors used the Incongruence Length Difference test (ILD test) for checking for cytonuclear incongruence (Farris *et al*. 1994, 1995). This is a much used (Campbell *et al*. 2011), parsimony-based test, but this method has been shown to fail when the average rate of substitutions differs between sites and/or partitions and/or over time (Barker and Lutzoni 2002, Darlu and Lecointre 2002). The rate of substitution differs greatly between the partitions of Kergoat *et al*. and therefore this test was inappropriate for investigating cytonuclear incongruence in this data set.

The selection of best partitioning strategy and model selection was done on the total data set. Preferably, this should be tested on the ingroup only, because that is the part you're most interested in for the phylogenetic reconstruction. I've shown that excluding the outgroups from the jModelTest runs can lead to the selection of a different evolutionary model (see section 2.5: Substitution model selection).

The authors report an extensive list of optimal evolutionary models in table 4. Of these, they only used the part from their optimal partitioning strategy (P_F), as determined by the Bayes Factors (Kass and Raftery 1995) of comparisons between runs with different partitioning strategies. But how they implemented this is unclear, because not all of these models are supported by MrBayes (see section 2.5: Substitution model selection). They also omit reporting the temperature of their "incrementally heated chains." In the results they confuse posterior probabilities with bootstrap values, using the latter term for the former.

Kergoat *et al*. argue that *S*. *exigua* exhibits "several distinctive plesiomorphic features" consistent with the position of this species in their molecular phylogenetic reconstruction. The given example hereof, the circular orbicular spot on the wings of this species, as opposed to a elongated oval spot in other Spodoptera species, is not as stable as claimed by the authors: also in *S*. *exigua* the orbicular spot is often quite oval (pers. comm. Marja van der Straten).

They found an genetic distance of 2,34% between the two proposed subspecies of *S*. *mauritia*. However, the specimens of *S*. *m*. *mauritia* come from the island Réunion, while the *S*. *m*. *acronyctoides* specimen comes from Papua New Guinea. This geographic distance between the populations could also (partly) explain the calculated genetic distance between these populations. Therefore they justly state in the discussion: "additional samples are needed to resolve this preliminary finding."

Host plants

They classified the post plants according to the APG2 phylogenetic hypothesis (see the supplementary table A2). Why, anno 2012, not use APG3 for the classification of the host plants? I think that this might lead to future confusion on the host preferences of *Spodoptera* species.

Although the paper goes into larval morphology (mostly their jaws), the paper does not specify of which species they examined the larvae, nor of which they have larvae in the collection.

They hypothesize that mandibles with serrate-like processes are derived in Spodoptera. However, grass-feeders are in the derived position in *Spodoptera* and have the chisel-like mandibles, therefore the more likely hypothesis would be that the chisel-like mandibles are derived in *Spodoptera*. This is exactly what they conclude in the results section on host-plant optimizations. They also conclude that specialization on monocots evolved independently three times. Only once this is accompanied by a modification of the mandibles to the chisel-type, which is an even stronger support for the hypothesis that the chisel-type mandible is derived in *Spodoptera* (unlike the opposite, hypothesized by Kergoat *et al*.).

In the host plant optimizations they make some fundamental errors. Firstly, they speak of "surpressing the branches leading to the terminal taxa with no host plant data." What they mean with this and how they did that, is unclear. I think what they mean here is 'pruning'. Secondly, they have imported their tree topology into Mesquite. In this step the branch lengths are lost and the tree is converted to a cladogram style tree. This is very important, because the Mk1 model they used for the character optimization uses branch lengths as a

proxy for evolutionary time (Lewis 2001). Ideally, you would supply the model with an ultrametric tree, not a cladogram-style tree in Mesquite like Kergoat *et al*. did. Lastly, significance is not a relevant question in a Mk1 model character optimisation: It calculates the chance of having a certain character state at a certain node (Lewis 2001), so comparing the chances of each character state by their likelihoods (like Kergoat *et al*. did) is superfluous. And when likelihoods are compared, it should be done with the Likelihood Ratio Test, not by assuming that a difference in 2 log-likelihood units is sufficient to call a certain difference significant (Pagel 1999).

Dating

The authors argue that they consider the two fossils as representatives for the entire superfamily, instead of assigning them to a lineage within the subfamily. That I feel would be a justified choice if their taxonomic sampling would also cover the entire superfamily. Especially because one calibration points represents the Notodontidae, which is in a much more basal position relative to their sampled taxa. Therefore I think that placing this date (48 Mya) at the root node of their tree topology is an overestimation of this nodes' age.
Next they choose an arbitrary age of 200 Mya as a maximum age on the root. They call that "likely very conservative," which I think is an understatement for the improbability of this age for the crown of the Erebidae+Noctuidae. They used both an uniform and an exponential (without reporting the mean and offset) prior on the tree root with a minimum of 48 and a maximum of 200 million years. Because these dates are likely to overestimate the age of the crown age of their tree, I don't see why a uniform prior would be suitable to use. The authors also note this: using an exponential distribution on the root, it has "less age discrepancies" than the uniform distribution.
The other priors in BEAST they left to the default settings, which should better be given informative priors (Drummond *et al*. 2006, 2012). They used a sample frequency of every 100^{th} generation from a run lasting 50 million generations. This means that their results contain 50,000,000 / 100 = 500,000 trees, which is an unlikely, although not impossible, high number.
The resulting calculated "very conservative age estimates" for all nodes from both the uniform and exponential prior they next transferred to a PL analysis (Sanderson 2002, 2003). In the results they note that "both BRC and PL analyses provide similar estimates", which is trivial given the methods applied.

Biogeography

Kergoat *et al*. state that *Spodoptera exigua* "colonized the Western Hemisphere as a result of an accidental introduction in North America in 1876", referring here to Wilson (1932). This is not stated in Wilson's paper. What is stated is only this: "The first record of *L*. *exigua* in the United States is that given by Harvey (1876) in the Canadian Entomologist". Indeed Harvey (1876) reports this species under the name *Caradrina flavomaculata* collected in Oregon and California, without any comment on its likely origin or possible introduction. This part of the United States was not colonised by many European settlers for a long time yet. Therefore this species was very unlikely to be collected before this date. And most importantly, the absence of a record of a species does not mean that it is not present. Also Pogue (2002) calls this species "the only cosmopolitan species in the genus." Therefore I do not support the conclusion Kergoat *et al*. that *S*. *exigua* is introduced to the western hemisphere.
This flaw in their reasoning has many consequences. Firstly, they set in their DEC analysis (Ree *et al*. 2005, Ree and Smith 2008) the maximum number of ancestral areas to four, because supposedly the extant taxa are also not wider spread than that. However, *S*. *exigua* is spread wider than they assumed and therefore this choice is incorrect. Secondly, their model reconstructs the ancestor of *Spodoptera* to be of Afrotropical+Oriental origin. But the most basal species in any historical biogeographic analysis will have the greatest influence (relative to the other species) on the ancestral state. In their tree topology the most basal lineage is *S*. *exigua*, whose area of distribution is erroneously coded. When this coding would be adapted the inferred ancestral state will likely be different and probably unresolved. This also influences the estimation of the states of the other nodes between the clades in *Spodoptera*. They concluded that two long distance dispersal events have occurred, with the second being a reverse colonization of the Old World. A different coding for the distribution of *S*. *exigua* will also be likely to lead to another conclusion here.
Another problem is their overestimation of the age of *Spodoptera*. For example, they estimate a Neotropic to Nearctic dispersal event to have occurred about 5 million years before the formation of the Isthmus of Panama. But this age difference would at least be smaller if the age of *Spodoptera* would not be overestimated.
Lastly, they concluded that the radiation of the grass-feeding clade is estimated to have happened contemporary with the transition from woodlands to open grasslands on many continents. Thereby they refer to Mitchell *et al*. (2006), a publication indeed saying this, but not investigating nor measuring this.

Concluding remarks

In the introduction the authors state that the high fecundity and short life cycle of many *Spodoptera* species are the main causes of their pest status. Although partly true, I think that this is an understatement of the complexity of this issue. Other factors involved are, among others, polyphagy and a good flying capability. They also aimed at clarifying the status of several species, but there are no hard conclusions on this subject in the paper. They do say that in their phylogenetic reconstructions *S*. *cosmioides* and *S*. *ornithogalli* are recovered paraphyletic: "these findings suggest that the status of *S*. *marima* should be reconsidered." But apparently they leave it to other authors to synonymise this species with *S*. *ornithogalli*. Similarly for the species pair *S*. *descoinsi* and *S*. *cosmioides*, which do not form monophyletic clades in their phylogenetic reconstruction: "it seems difficult to put the two taxa into synonymy because they exhibit marked differences in wing ornamentation and genitalia in addition to the differences in pheromones component ratio and calling period." In the section on character sampling I described that the study effectively only used 'one' marker, which is not enough in species delimitation (Dupuis *et al*. 2012). Therefore I think that these two species pairs should remain separate entities until a molecular study samples these species for at least three independent markers.

They stress that "Though our most probable scenarios are only partially statistically supported […], our analyses rely on *several* sets of calibrations, a *well-resolved* phylogenetic hypothesis, and *complete* taxonomic sampling." All things in italics I have contested in this review.

5.2. Comparison of results

In figure 28 I compare the tree topologies of Kergoat *et al*. (2012) and my phylogenetic reconstruction, converted to the same branch length scale. Mostly the relationships between and within clades are similar. There is one difference in the topology of a deeper node: In my phylogenetic reconstruction the clade EH is sister to MT+PEC, in Kergoat *et al*. S. *exigua* is sister to all other *Spodoptera* species. This difference might stem from several causes. Including *S. hipparis* in the taxon sampling (see table 22) might introduce conflict in the position of this clade, leading to a different branching order and the low support (pp=0.53) on this node. Another possibility is the different focus of the character sampling: a focus on nuclear markers in this thesis, contrasting with the focus on mitochondrial markers in Kergoat *et al*.

There is also a major difference in branch lengths. This is probably caused by the different rates of substitution between markers, reflected by differences in average number of parsimony informative sites. The 141-taxa data set of Kergoat *et al*. has only 16.4% parsimony informative sites. This is fewer than mine (22.2% parsimony informative sites in Concatenated Alignment 1 and 33.5% in Concatenated Alignment 2). However, one remark has to be made here: Whether a site is recovered as parsimony informative depends on the taxonomic sampling: a broader taxonomic sampling means a longer separate evolutionary history and thus more variable sites. Because my taxonomic sampling covers the entire superfamily (in CA2, not in CA1) it is not surprising that I've more parsimony informative sites. But note that CA1 also has a higher number of parsimony informative sites than 16.4%.

In my review of the paper by Kergoat *et al*. I describe several issues in the part on time-calibration of the phylogenetic reconstruction. These resulted in overestimations of the age of *Spodoptera*, explaining the difference between my age estimates and theirs. They reported the 95% HPD of the age for the crown of *Spodoptera* between 12.56 and 54.84 Mya. My 95% HPDs lie between 4.78 and 10.91 Mya for the different calibration combinations, thus there is no overlap at all between the estimates.

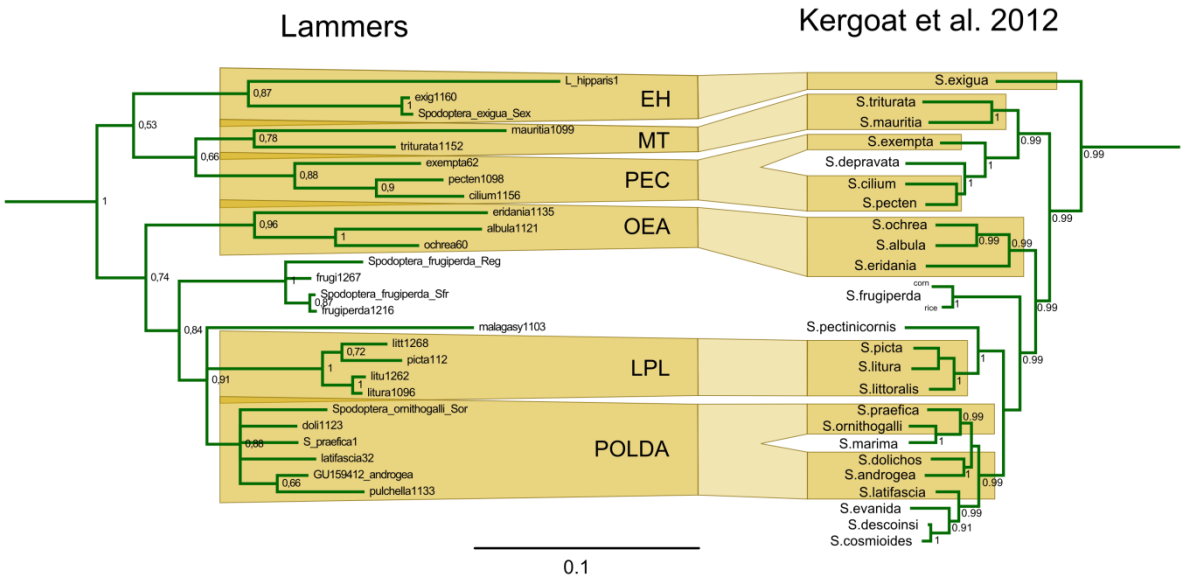


Figure 28. Comparison of the tree topology of the *Spodoptera*-part of this thesis with Kergoat *et al*. (modified from their figure 2B). Branch lengths are converted to the same branch length scale.

Table 22. The taxonomic sampling of this thesis and Kergoat *et al*. 2012.

Species	Sampled in this study?	Sampled in Kergoat *et al*. 2012?
Spodoptera albula	yes	yes
Spodoptera androgea	yes	yes
Spodoptera apertura	no	no
Spodoptera cilium	yes	yes
Spodoptera compta	no	no
Spodoptera cosmioides	no	yes
Spodoptera depravata	no	yes
Spodoptera descoinsi	no	yes
Spodoptera dolichos	yes	yes
Spodoptera eridania	yes	yes
Spodoptera evanida	no	yes
Spodoptera exempta	yes	yes
Spodoptera exigua	yes	yes
Spodoptera frugiperda	yes	yes
Spodoptera hipparis	yes	no
Spodoptera latifascia	yes	yes
Spodoptera littoralis	yes	yes
Spodoptera litura	yes	yes
Spodoptera malagasy	yes	no
Spodoptera marima	no	yes
Spodoptera mauritia	yes	yes
Spodoptera ochrea	yes	yes
Spodoptera ornithogalli	yes	yes
Spodoptera pecten	yes	yes
Spodoptera pectinicornis	no	yes
Spodoptera picta	yes	yes
Spodoptera praefica	yes	yes
Spodoptera pulchella	yes	no
Spodoptera roseae	no	no
Spodoptera triturata	yes	yes
Spodoptera umbraculata	no	no

6. Recommendations and future prospects

6.1. Improvements of data set and analyses

6.1.1. Merge Kergoat *et al*. and Lammers *et al*. data sets

The independently generated data sets of Kergoat *et al*. (2012) and the one used for this thesis should be merged into one large data set in order to gain more insight in the *Spodoptera* species level relationships. Our data set has three species which are missing in Kergoat *et al*. (2012), while they have six species missing in our data set. After merging of these data sets, only four of the 31 species will still be missing.

The *Spodoptera* species neither sampled in this study nor in Kergoat *et al*. (2012) should be collected from the field in order to make a complete molecular phylogenetic reconstruction of the genus. These missing species are *Spodoptera apertura* form the Afrotropical, Oriental and Australasian region; *S*. *compta* from the Neotropics (only known from Peru); *S*. *roseae*, an endemic from the Galapagos Islands; and *S*. *umbraculata* from Australia. Unfortunately there was not enough time available in my MSc thesis project to collect them. Once we have all species available in the collection I recommend the sequencing of all species for all 11 markers listed in Wahlberg and Wheat (2008) and all 19 markers used in Regier *et al*. (2013, see supplementary table S4 there). These sets of markers overlap: CAD, DDC and wingless are used in both. With such a dense character sampling and complete taxon sampling the monophyly and the position of *Spodoptera* in the Noctuidae can be recovered.
Even better would be to sequence the whole genomes of all species, which is possible nowadays and will be financially feasible in the near future.

6.1.2. Collect fresh *S. mauritia* from Fiji

In the discussion I propose several possible explanations for the COI-sequence divergence between the specimens of *S*. *mauritia* from Fiji and their conspecifics from other populations. The most likely explanation is the NUMT-hypothesis. Unfortunately, the specimens are 'used up' by now: no more legs, abdomen or head is available for sequencing. Moreover, the specimens are very old. Probably their DNA is highly degraded, explaining the failed attempts in doing PCR on nuclear markers. These issues clarify the need for fresh specimens. Someone should go to Fiji, collect new specimens and sequence their COI-barcode region as described in the discussion.

6.1.3. Investigate why amplification of IDH failed for *Spodoptera*

Despite several attempts I failed to amplify IDH for any specimen of *Spodoptera*. It should be investigated why this did not work. Perhaps *Spodoptera* has one or more crucial mutations in the primer binding sites, causing a mismatch between primer and template. A possible way of investigating this could be by sequencing of the transciptome of *Spodoptera* and thereby obtaining the full sequence of the IDH exon. Then the primer binding site can be compared to the primer sequence developed by Wahlberg and Wheat (2008).

6.1.4. Use Partitionfinder

Lanfear *et al*. (2012) published a program which calculates the optimal partitioning scheme from and for a multigene alignment. Optimal partitioning should lead to the most likely phylogenetic reconstruction. When I started my MSc thesis project this software package was not published yet. Due to time constraints I have not had the possibility of investigating its usefulness, but conceptually the program has great potential.

6.1.5. Do not use EF1α in Lepidoptera species-level systematics

Every genetic marker has a different phylogenetic resolution. EF1α has little or no sequence variation at the species level in *Spodoptera* (see this thesis, Kergoat *et al*. 2012) and Lepidoptera in general (Wahlberg and Wheat 2008). Sequencing this marker with the purpose of reconstructing a phylogeny with low taxonomic depth is a waste of time. Therefore no one should ever use this marker again in a Lepidoptera species-level molecular phylogenetics.

6.2. New analyses

6.2.1. Make phylogenetic networks

Several software packages for different types of networks are available, which provide more insight in the sequence data. SplitsTree (Huson and Bryant 2006) is a package for making tree networks out of data from one or more molecular markers (Huson *et al*. 2010). The benefit of phylogenetic networks is that they show what part of the phylogenetic signal comes from what part of the input data. This can especially be useful when there are indications of conflict between markers.
Another method for estimating a species tree when conflicting gene trees are present was published by Edwards *et al*. (2007). I have not looked in to the usefulness of this method for the *Spodoptera* data set, but it would be worth investigating in a follow-up study.

6.2.2. Character optimizations

6.2.2.1. Morphology

Pogue (2002) defined 24 morphological characters of adult *Spodoptera*, which can all be optimised over a phylogenetic reconstruction. These characters are all described for all species when combined with the additions and modifications proposed by Pogue (2011) and Van der Straten and Germain (in publ.). When these data are optimised over the tree in order to find the ancestral state and infer evolutionary trends we can investigate morphological synapomorphies for the clades I defined. The morphological characteristics of larvae can also be optimised when larvae from all *Spodoptera* species are available.

6.2.2.2. Host preference and specificity

Similar to morphological character optimizations, host preference or specificity can also be optimized over a species tree. An extra problem here, besides character definitions and delimitation, is the missing data. For many species rare species we simply don't know what their larvae eat, or can potentially eat, of prefer to eat. This will lead to many ambiguous character states at nodes. It has to be investigated whether it is possible with the current knowledge to do these optimizations at all.

6.2.3. Investigate the historical biogeography

Kergoat *et al*. (2012) published an historical biogeographic optimization using Dispersal-Extinction-Cladogenesis (DEC) analysis. This could work well, when the initial assumptions are made correctly (i.e. the distributions of all species, see my review of their paper above).
Another software package for analyses of dispersal and vicariance patterns is RASP (Yu *et al*. submitted). This is the new version of S-DIVA (Yu *et al*. 2010), a package used in many previous historical biogeographic investigations. Its utility for *Spodoptera* should be investigated in future research.

6.2.4. Time-calibrate an extended phylogeny over all ditrysian Lepidoptera

In this thesis I assumed that the emergence of bat echo-location and the emergence of non-homologous tympanal organs in nocturnal Lepidoptera is linked. This premise should be investigated for its validity. I recommend doing that by first extending the data set over all ditrysian ('higher') Lepidoptera; next including all available independent fossil calibrations and then check the estimated ages of all clades having a tympanal organ. If the emergence of all clades having a tympanal organ are of equal age and that age is consistent with the theory on the linkage with bat echo-location, it will be highly likely that the theory is correct.

6.3. Use the LPL-clade as a model system

Why a herbivorous insect species should specialise on a narrow range of host plants is a long lasting question in evolutionary biology (e.g. Schoonhoven *et al*. 2005). On the other hand, some herbivores seem to be able to eat just about anything and many of these emerge as crop pests. The *Spodoptera* LPL-clade is in interesting case for investigating all facets of generalism and specialism of host plants: Two of the species in this clade, *S*. *littoralis* and *S*. *litura*, are extremely polyphagous, both described from over 80 plant families. But the third species, *S*. *picta*, is narrowly oligophagous on Amaryllidaceae. Likely the ancestor of this clade was polyphagous as well and thus *S*. *picta* has switched its host preference to a very small angiosperm group for no obvious reason. Nonetheless this species has a wide distribution, almost as wide as *S*. *litura* (Pogue 2002), which could mean that it is equally 'successful'. Because the split between these species is at most 2 million years old, I expect not many substitutions to have accumulated in this time. Therefore, any non-silent substitution found in this species relative to the polyphagous sister species, is likely to be related to this change in host specificity. Using RNA-seq (the sequencing of all RNA isolated from a specimen, preferably only the mRNA by a poly-A-tail fishing step) it is possible to find which genes are being transcribed and what their specific sequence is (Wang *et al*. 2009, Grabherr *et al*. 2011). Comparing these transcriptomes between different species and between different stages within a species, or between larvae feeding on different host plants, can provide insight in the relative importance of the genes expressed by the larvae. I expect that such a marked change in ecology should be reflected by a large difference in transriptome composition. Ultimately this will lead to an understanding of why specialism arises and why other species are polyphagous.

7. Conclusion

In this thesis I have shown that (with this taxonomic sampling) *Spodoptera* is a monophyletic genus, having clearly discerned monophyletic clades which partly correspond to a biogeographic separation between Old World and New World clades. The crown age of *Spodoptera* I estimated to be between 5 and 11 million years old. Because many *Spodoptera* species are crop pests, while some others have a narrow host range or distribution, the genus would be perfect as a model system for studying pest species ecology and their evolutionary history. And for the same reason the genus is of interest for scientist aiming at answering why herbivorous species tend to specialise.

Acknowledgements

Many people helped me during my thesis. Here I wish to thank them all: Firstly, my supervisors Freek T. Bakker and Marja van der Straten for their time, help, valuable discussions and guidance; Willem Ellis for help in locating *Elaphria* specimens; Maurice Franssen for supplying a *Elaphria venustula* specimen and discussion; Ria Vrielink-van Ginkel for the introduction in the lab; Pulcherie Bissiengou for the help in the lab; Michael G. Pogue for sending me the COI-sequence data of his 2011-publication; Jakub Prokop for all extra information on and the photo of the Notodontid fossil; Julieta Brambila, John Nielsen and Felix Sperling for discussing the topic with me in Denver; Eric Schranz for all stimulating discussions; Lars Chatrou and Robin van Velzen for the helpful discussions and the help with the software; Martijn Staats for all discussions; the students of the Biosystematics Group (Rutger Wilschut, Jeike van der Poel, Rens Holmer, Erik-Jan Bijleveld, Sara van de Kerke, Daan Drukker, Marjolein Soethoudt) for the help and the distractions; Erik van Nieukerken for his help with the funding for the conference visit; Camiel Doorenweerd for discussing lab problems; Don Lafontaine for the discussion and guidance on echo-location; Marcus Koch for the brief but helpful discussion; and lastly my parents, my sisters and Karina Prins for their unlimited support.

This research was facilitated by: the Wageningen University Biosystematics Group, the nVWA and the CIRPES science gateway.

WAGENINGEN UR
For quality of life

Nederlandse Voedsel- en Warenautoriteit
Ministerie van Economische Zaken, Landbouw en Innovatie

CIPRES SCIENCE GATEWAY

I also thank my funders for the conference visit: the Wageningen University Fund and the Uyttenboogaart-Eliasen Stichting.

Wageningen University Fund.

UYTTENBOOGAART -ELIASEN STICHTING

References

- Alfaro ME, Zoller S, Lutzoni F (2003) Bayes or bootstrap? a simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Molecular Biology and Evolution 20: 255-266.
- APG III (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161: 105-121
- Bakker FT (2007) Hoofdstuk 7 Testen DNA kwaliteit en geschiktheid. In: Fransen JJ (2007) *Eindrapportage Plan van Aanpak+Vervolg FES-programma 'Versterking Infrasctructuur Plantgezondheid'*. Rapport Plantenziektekundige Dienst, Wageningen, The Netherlands.
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. Molecular Ecology 13: 729-744
- Barker FK, Lutzoni FM (2002) The utility of the Incongruence Length Difference Test. Systematic Biology 51: 625-637
- Bensasson D, Zhang D-X, Harti DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends in Ecology and Evolution 16: 314-321
- Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. Nucleic Acids Research 36: D25-30. Accessible on http://www.ncbi.nlm.nih.gov/genbank/
- Beutel RG, Friedrich F, Hörnschemeyer T, Pohl H, Hünefeld F, Beckmann F, Meier R, Misof B, Whiting MF, Vilhelmsen L (2010) Morphological and molecular evidence converge upon a robust phylogeny of the megadiverse Holometabola. Cladistics 27: 341-355
- Brady SG, Danforth BN (2004) Recent intron gain in Elongation Factor-1a of Colletid Bees (Hymenoptera: Colletidae). Molecular Biology and Evolution 21: 691-696
- Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proceedings of the National Academy of Sciences 91: 6491-6495
- Campbell V, Legendre P, Lapointe F-J (2011) The performance of the Congruence Among Distance Matrices (CADM) test in phylogenetic analysis. BMC Evolutionary Biology 11: 64-79
- Chinnery PF, Thorburn DR, Samuels DC, White SL, Dahl HM, Turnbull DM, Lightowlers RN, Howell N (2000) The inheritance of mitochondrial DNA heteroplasmy: random drift, selection or both? Trends in Genetics 16: 500-505
- Cho S, Mitchell A, Mitter C, Regier J, Matthews M, Robertson R (2008) Molecular phylogenetics of heliothine moths (Lepidoptera: Noctuidae: Heliothinae), with comments on the evolution of host range and pest status. Systematic Entomology 33: 581-594
- Cho S, Zwick A, Regier JC, Mitter C, Cummings MP, Yao J, Du Z, Zhao H, Kawahara AY, Weller S, Davis DR, Baixeras J, Brown JW, 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
- Cordaux R, Bouchon D, Grève P (2011) The impact of endosymbionts on the evolution of host sex-determination mechanisms. Trends in Genetics 27: 332-341
- Danforth BN, Ji S (1998) Elongation Factor-1a occurs as two copies in bees: implications for phylogenetic analysis of EF-1a sequences in insects. Molecular Biology and Evolution 15:225-235
- Darlu P, Lecointre G (2002) When Does the Incongruence Length Difference Test Fail? Molecular Biology and Evolution 19: 432-437
- DeSalle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transitions of the Royal Society B 360: 1905-1916
- Dorp K van (2004) Molecular systematics of *Lycaena* F., 1807 (Lepidoptera: Lycaenidae) – Some preliminary results. Proceedings of the Netherlands Entomological Society 15: 65-70
- Doublet V, Souty-Grosset C, Bouchon D, Cordaux R, Marcadé I (2008) A thirty million year-old inherited heteroplasmy. PLoS ONE 3(8): e2938. doi:10.1371/journal.pone.0002938
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) relaxed phylogenetics and dating with confidence. PLoS Biol 4(5): e88. doi:10.1371/journal.pbio.0040088
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology 7: 214. Available from http://beast.bio.ed.ac.uk

- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 Molecular Biology and Evolution 29: 1969-1973
- Dupuis JR, Roe AD, Sperling FAH (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. Molecular Ecology 21: 4422-4436
- Edwards SV, Liu L, Pearl DK (2007) High-resolution species trees without concatenation. Proceedings of the National Academy of Sciences 104: 5936-5941
- Edwards EJ, Osborne CP, Strömberg CAE, Smith SA, C4 Grasses Consortium (2010) The origins of C4 grasslands: integrating evolutionary and ecosystem science. Science 328: 587-591
- Eick GN, Jacobs DS, Matthee CA (2005) A nuclear DNA phylogenetic perspective on the evolution of echolocation and historical biogeography of extant bats (Chiroptera). Molecular Biology and Evolution 22: 1869-1886
- Eiting TP, Gunnell GF (2009) Global completeness of the bat fossil record. Journal of Mammalian Evolution 16: 151-173
- Ellis SE (2004) New Pest Response Guidelines: *Spodoptera*. USDA/APHIS/PPQ/PDMP. http://www.aphis.usda.gov/ppq/manuals/
- Farag M, Ahmed MHM, Yousef H, El-badawey SS, Abd El-Ghany MA, Abdel-Rahman AAH (2012) Repellent and insecticide activity of *Pelargonium* x *hortorum* against *Spodoptera littoralis* (Boisd.). Zeitschrift für Naturforschung C: 398-404
- Farris JS, Källersjö M, Kluge AG, Bult C (1994) Constructing a significance test for incongruence. Systematic Biology 44: 570-572
- Farris JS, Källersjö M, Kluge AG, Bult C (1995) Testing significance of incongruence. Cladistics 10: 315-319
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. Journal of Molecular Evolution 17: 368–376
- Fordyce JA (2010) Host shifts and evolutionary radiations of butterflies. Proceedings of the Royal Society B 277: 3735-3743
- Folmer O, Black W, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294-299
- Goto H, Dickins B, Afgan E, Paul IM, Taylor J, Makova KD, Nekrutenko A (2011) Dynamics of mitochondrial heteroplasmy in three families investigated via a repeatable re-sequencing study. Genome Biology 12: R59
- Grabherr MF, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology 29: 644-652
- Graham RI, Grzywacz D, Mushobozi WL, Wilson K (2012) *Wolbachia* in a major African crop pest increases susceptibility to viral disease rather than protects. Ecology Letters 15: 993-1000
- Graham RI, Wilson K (2013) Male-killing *Wolbachia* and mitochondrial selective sweep in a migratory African insect. BMC Evolutionary Biology 12: 204-214
- Graham RI, Wilson K (2013) Correction: Male-killing *Wolbachia* and mitochondrial selective sweep in a migratory African insect. BMC Evolutionary Biology 13: in press.
- Guenée (1852). In: Boisduval & Guenée (1852) *Histoire Naturelle des Insectes. Species Général des Lépidoptéres. Tome Cinquiéme. Noctuélites. Tome 1*: 1-407, *Atlas* (Noctuelites): pl. 1-24. Paris: Roret.
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006a) DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Sciences of the United States of America 103: 968-971
- Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitfield JB, Hebert PDN (2006b) A minimalist barcode can identify a specimen whose DNA is degraded. Molecular Ecology Notes 6: 959-964
- Hampson GF (1906) *Catalogue of the Lepidoptera Phalaenae in the British Museum*. Volume 6. London: Taylor and Francis. 532 pp.
- Harvey LF (1876) New Californian and Texan moths. The Canadian Entomologist 8: 52-56.
- Hasegawa M, Kishino H, Yano T (1985) Dating of human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22: 160-174
- Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B 270: 313-321

- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) 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 PDN, deWaard JR, Landry J-F (2010) DNA barcodes for 1/1000 of the animal kingdom. Biology Letters 6: 359-362
- Hedges SB, Kumar S (2004) Precision of molecular time estimates. TRENDS in Genetics 20: 242-247
- Hedin MC, Maddison WP (2001) Phylogenetic utility and evidence for multiple copies of Elongation Factor-1a in the spider genus *Habronattus* (Araneae: Salticidae). Molecular Biology and Evolution 18: 1512-1521
- Heikkilä M, Kaila L, Mutanen M, Peña C, Wahlberg N (2011) Cretaceous origin and repeated tertiary diversification of the redefined butterflies. Proceedings of the Royal Society B 279: 1093-1099
- Hennig W (1966) *Phylogenetic systematics*. Urbana, Illinois, USA: University of Illinois Press. 280 pp.
- Heppner JB (1998) *Spodoptera* armyworms in Florida (Lepidoptera: Noctuidae). Entomology Circular 390
- Hilgenboecker K, Hammerstein P, Schlattmann P, Telschow A, Werren JH (2008) How many species are infected with *Wolbachia*? – a statistical analysis of current data. FEMS Microbiology Letters 281: 215-220
- Ho SYW, Phillips MJ (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology 58: 367-380
- Hovemann B, Richter S, Walldorf U, Cziepluch C (1988) Two genes encode related cytoplasmic elongation factors 1α (EF-1α) in Drosophila melanogaster with continuous and stage specific expression. Nucleic Acids Research 16: 3175-3194
- Hugall AF, Lee MSY (2007) The likelihood node density effect and consequences for evolutionary studies of molecular rates. Evolution 61: 2293-2307
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23: 254-267
- Huson DH, Rupp R, Scornavacca C (2010) *Phylogenetic Networks*. Cambridge: Cambridge University Press. 374 pp.
- Janz N, Nyblom K, Nylin S (2001) Evolutionary dynamics of host-plant specialization: a case study of the tribe Nymphalini. Evolution 55: 783-796
- Jiggins FM, Hurst GDD (2011) Rapid insect evolution by symbiont transfer. Science 332: 185-186
- Johnson MTJ, Carpenter EJ, Tian Z, Bruskiewich R, Burris JN, Carrigan CT, Chase MW, Clarke ND, Covshoff S, dePamphilis CW, Edger PP, Goh F, Graham S, Greiner S, Hibberd JM, Jordon-Thaden I, Kutchan TM, Leebens-Mack J, Melkonian M, Miles N, Myburg H, Patterson J, Pires JC, Ralph P, Rolf M, Sage RF, Soltis D, Soltis P, Stevenson D, Stewart Jr. CN, Surek B, Thomsen CJM, Villarreal JC, Wu X, Zhang Y, Deyholos MK, Wong GK (2012) Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS ONE 7(11): e50226. doi:10.1371/journal.pone.0050226
- Jones G, Teeling EC (2006) The evolution of echolocation in bats. TRENDS in Ecology and Evolution 21: 149-156
- Jones G, Holderied GJ (2007) Bat echolocation calls: adaptation and convergent evolution. Proceedings of the Royal Society B 274: 905-912
- Jordal BH (2002) Elongation Factor 1α resolves the monophyly of the haplodiploid ambrosia beetles Xyleborini (Coleoptera: Curculionidae). Insect Molecular Biology 11 453-465
- Juárez ML, Murúa MG, García MG, Ontivero M, Vera MT, Vilardi JC, Groot AT, Castagnaro AP, Gastaminza G, Willink E (2012) Host association of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) Corn and Rice Strains in Argentina, Brazil, and Paraguay. Journal of Economic Entomology 105: 573-582
- Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (editor), *Mammalian Protein Metabolism*, pp. 21-132, New York: Academic Press.
- Kass RE, Raftery AE (1995) Bayes factors. Journal of the American Statistical Association 90: 773-795
- Kelchner SA, Thomas MA (2006) Model use in phylogenetics: nine key questions. TRENDS in Ecology and Evolution 22: 87-94
- Kergoat GJ, Prowell DP, Le Ru BP, Mitchell A, Dumas P, Clamens AL, Condamine FL, Silvain JF (2012) Disentangling dispersal, vicariance and adaptive radiation patterns: a case study using armyworms in

the pest genus *Spodoptera* (Lepidoptera: Noctuidae). Molecular Phylogenetics and Evolution 65: 855-70

- Kitching IJ, Rawlins JE (1998) *Noctuoidea. Handbook of Zoology. Lepidoptera*, Vol. 1: Systematics and Evolution (ed. NP Kristensen). p. 355-401. Berlin: W. de Gruyter.
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 6: 111-120
- Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proceedings of the National Academy of Sciences of the United States of America 78: 454-458,
- Kolaczkowski B, Thornton JW (2009) Long-branch attraction bias and inconsistency in Bayesian phylogenetics. PLoS ONE 4(12): e7891. doi:10.1371/journal.pone.0007891
- Kristensen NP, Scoble MJ, Karsholt O (2007) Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity. Zootaxa 1668: 699-747
- Kvaček Z, Böhme M, Dvořák Z, Konzalová M, Mach K, Prokop J, Rajchl M (2004) Early Miocene and swamp ecosystems of the Most Basin (Northern Bohemia) with particular reference to the Bílina Mine section. Journal of the Czech Geological Society 49: 1-40
- Lafontaine JD, Wood DM (1988) A zoogeographic analysis of the Noctuidae (Lepidoptera) of Beringia, and some inferences about past Beringian habitats. Memoirs of the Entomological Society of Canada 144: 109-123
- Lafontaine JD, Schmidt BC (2012) Annotated check list of the Noctuoidea (Insecta, Lepidoptera) of North America north of Mexico. ZooKeys 40: 1-239
- Lanfear R, Calcott B, Ho SYW, Guindon S (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29: 1695-1701
- Lemey P, Salemi M, Vandamme A-M (eds) (2009) *The Phylogenetic Handbook: A practical approach to phylogenetic analysis and hypothesis testing*. Cambridge: Canbridge University Press. 723 pp.
- Lemmon AR, Moriarty EC (2004) The importance of proper model assumption in Bayesian phylogenetics. Systematic Biology 53: 265-277.
- Letsch HO, Meusemann K, Wipfler B, Schütte K, Beutel R, Misof B (2012) Insect phylogenomics: results, problems and the impact of matrix composition. Proceedings of the Royal Society B 279: 3282-3290
- Levy HC, Garcia-Maruniak A, Maruniak J (2002) Strain identification of *Spodoptera frugiperda* (Lepidoptera: Noctuidae) insects and cell line: PCR-RFLP of cytochrome oxidase subunit 1 gene. Florida Entomologist 85: 186-190
- Lewis PO (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50: 913-925
- Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451-1452
- Lopez JV, Yuhki N, Masuda R, Modi W, O'Brien SJ (1994) Numt, a recent transfer and tandem replication of mitochondrial DNA to the nuclear genome of the domestic cat. Journal of Molecular Evolution 39: 174-190
- Mable BK (2004) 'Why polyploidy is rarer in animals than in plants': myths and mechanisms. Biological Journal of the Linnean Society 82: 453-466
- Maddison WP, Maddison DR (2007) Mesquite: a modular system for evolutionary analysis. Version 2.74. Available from http://mesquiteproject.org
- Mandrioli M, Borsatti F, Mola L (2006) Factors affecting DNA preservation from museum-collected lepidopteran specimens. Entomologia Experimentalis et Applicata 120: 239-244
- May RM (1988) How Many Species are There on Earth? Science 241: 1441-1449
- May RM (2010) Tropical Arthropod Species, More or Less? Science 329: 41-42
- Mayfield-Jones D, Washburn JD, Arias T, Edger PP, Pires JC, Conant GC (2013) Watching the grin fade: Tracing the effects of polyploidy on different evolutionary time scales. Seminars in Cell and Developmental Biology, in press.
- Meagher RL, Brambila J, Hung E (2008) Monitoring for exotic *Spodoptera* species (Lepidoptera: Noctuidae) in Florida. Florida Entomologist 91: 517-522
- Mehrkhou F, Talebi AA, Moharramipour S, Naveh VH (2012) Demographic parameters of *Spodoptera exigua* (Lepidoptera: Noctuidae) on different soybean cultivars. Environmental Entomology, 41: 326-332

- Meredith RW, Janečka JE, Gatesy J, Ryder OA, Fisher CA, Teeling EC, Goodbla A, Eizirik E, Simão TLL, Stadler T, Rabosky DL, Honeycutt RL, Flynn JJ, Ingram CM, Steiner C, Williams TL, Robinson TJ, Burk-Herrick A, Westerman M, Ayoub NA, Springer MS, Murphy WJ (2011) Impacts of the Cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334, 521-334
- Meusemann K, von Reumont BM, Simon S, Roeding F, Strauss S, Kück P, Ebersberger I, Walzl M, Pass G, Breuers S, Achter V, von Haeseler A, Burmester T, Hadrys H, Wägele JW, Misof B (2010) A phylogenomic approach to resolve the Arthropod tree of life. Molecular Biology and Evolution 27: 2451-2464
- Mikkola K (2008) The lock-and-key mechanisms of the internal genitalia of the Noctuidae (Lepidoptera): How are they selected for? European Journal of Entomology 105: 13-25
- Misof B (2012) Phylogenomics and the evolution of insects - The 1KITE project. Proceedings of the Biodiversity Genomics Conference 2012, Canberra, Australia.
- Mitchell A, Mitter C, Regier JC (2006) Systematics and evolution of the cutworm moths (Lepidoptera: Noctuidae): evidence from two protein-coding nuclear genes. Systematic Entomology 31: 21-46
- Mora C, Tittensor DP, Adl S, Simpson AGB, Worm B (2011) How Many Species Are There on Earth and in the Ocean? PLoS Biol 9(8): e1001127. doi:10.1371/journal.pbio.1001127
- 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 277: 2839-2848
- Otto SP, Whitton J (2000) Polyploid incidence and evolution. Annual Review of Genetics 34: 401-437
- Nagoshi RN, Brambila J, Meagher RL (2011) Use of DNA barcodes to identify invasive armyworm *Spodoptera* species in Florida. Journal of Insect Science 11: 154
- Nakano R, Takanashi T, Skals N, Surlykke A, Ishikawa Y (2010) To females of a noctuid moth, male courtship songs are nothing more than bat echolocation calls. Biology Letters 6: 582-584
- Nieukerken EJ van, Kaila L, Kitching IJ, Kristensen NP, Lees DC, Minet J, Mitter C, Mutanen M, Regier JC, Simonsen TJ, Wahlberg N, Yen S-H, Zahiri R, Adamski D, Baixeras J, Bartsch J, Bengtsson BÅ, Brown JW, Bucheli SR, Davis DR, De Prins J, De Prins W, Epstein ME, Gentili-Poole P, Gielis C, Hättenschwiler P, Hausmann A, Holloway JD, Kallies A, Karsholt O, Kawahara AY, Koster JC, Kozlov MV, Lafontaine JD, Lamas G, Landry J-F, Lee S, Nuss M, Park K-T, Penz C, Rota J, Schintlmeister A, Schmidt BC, Sohn J-C, Solis MA, Tarmann GM, Warren AD, Weller S, Yakovlev RV, Zolotuhin VV, Zwick A (2011) Order Lepidoptera Linnaeus, 1758. In: Zhang, Z.-Q. (Ed.) *Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness*. Zootaxa 3148: 212-221. Auckland, New Zealand: Magnolia Press.
- Nosil P (2002) Transition rates between specialization and generalization in phytophagous insects. Evolution 56: 1701-1706
- Nylin S, Janz N (2009) Butterfly host plant range: an example of plasticity as a promoter of speciation? Evolutionary Ecology 23:137-146
- Otto SP, Whitton J (2000) Polyploid incidence and evolution. Annual Review of Genetics 34: 401-437
- Page RDM, Holmes EC (1998) *Molecular evolution: a phylogenetic approach*. Oxford: Blackwell Science Ltd. 352 pp.
- Pagel M (1999) The Maximum Likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Systematic Biology 48: 612-622
- Papadopoulou A, Anastasiou I, Vogler AP (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. Molecular Biology and Evolution 27: 1659-1672
- Passoa S (1991) Color identification of economically important *Spodoptera* larvae in Honduras (Lepidoptera: Noctuidae). Insecta Mundi: 414. http://digitalcommons.unl.edu/insectamundi/414
- Pirie MD, Chatrou LW, Erkens RHJ, Maas JW, Niet T van der, Mols JB, Richardson JE (2005) Phylogeny reconstruction and molecular dating in four Neotropical genera of Annonaceae: the effect of taxon sampling in age estimations. In: Bakker FT, Chatrou LW, Gravendeel B, Pelser PB (eds) *Plant Species-level Systematics: new perspectives on pattern & process*. Regnum Vegetabile 143. Liechenstein: A.R.G. Gantner Verlag, Ruggell
- Plotnick RE, Smith DM (2012) Exceptionally preserved fossil insect ears from the Eocene Green River Formation of Colorado. Journal of Paleontology, 86: 19-24
- Pogue MG (2002) A world revision of the genus *Spodoptera* Guenée (Lepidoptera: Noctuidae). Memoires of the American Entomological Society 43: 1-202

- Pogue MG, Sullivan JB (2003) Re-evaluation of the *Elaphria festivoides* (Guenée) species complex (Lepidoptera: Noctuidae). Proceedings of the Entomological Society of Washington 105: 331-347 Available online: http://biodiversitylibrary.org/page/16195612
- Pogue MG (2011) Using genitalia characters and mitochondrial COI sequences to place "*Leucochlaena*" *hipparis* (Druce) in *Spodoptera* Guenée (Lepidoptera: Noctuidae). Proceedings of the Entomological Society of Washington 113: 497-507
- Poole RW (1989) Fascicle 118, Noctuidae. In: Arnett Jr RH, Coalson AL (eds.) *Lepidopterorum Catalogus.* Leiden: EJ Brill and Flora and Fauna Publications. 1313 pp.
- Posada D (2003) Selecting models of evolution. In: *The Phylogenetic Handbook*. Vandemme AM, Salemi M (eds). Cambridge: Cambridge University press. pp. 256-282
- Posada D (2008) jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution 25: 1253-1256
- Prokop J (2003) Remarks on palaeoenviromental changes based on reviewed Tertiary insect associations from the Krušné hory piedmont basins and the České středohoří Mts in northwestern Bohemia (Czech Republic). Acta zoological cracoviensa 46: 329-344
- Prokop J, Wappler T, Knor S, Kvaček Z (2010) Plant-arthropod associationsfrom the lower Miocene of the Most Basin in Northern Bohemia (Czech Republic): A preliminary report. Acta Geologica Sinica 84: 903-914
- Prowell DP, McMichael M, Silvain J-F (2004) Multilocus genetic analysis of host use, introgression, and speciation in host strains of Fall Armyworm (Lepidoptera: Noctuidae). Annals of the Entomological Society of America 97: 1034-1044
- Puslednik L, Yeates DK, Faith DP, Ballard JWO (2012) Protein-protein interactions of the cytochrome c oxidase DNA barcoding region. Systematic Entomology 37: 229-236
- Quek S-P, Davies SJ, Itino T, Pierce NE (2004) Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). Evolution 58: 554-570
- Rambaut A (2006) FigTree: Tree figure drawing tool version 1.3.1. Available from http://tree.bio.ed.ac.uk/software/figtree/
- Rambaut A, Drummond AJ (2009) Tracer: MCMC trace analysis tool. Available from http://beast.bio.ed.ac.uk/
- Ratnasingham S & Hebert PDN (2007) BOLD: The Barcode of Life data system. Molecular Ecology Notes 7: 355-364. Accessible on www.barcodinglife.org
- Ree RH, Moore BR, Webb CO, Donoghue MJ (2005) A likelihood framework for inferring the evolution of geographic range on phylogenetic trees. Systematic Biology 59: 2299-2311
- Ree RH, Smith SA (2008) Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57: 4-14
- Regier JC, Shi D (2005) Increased yield of PCR product from degenerate primers with nondegenerate, nonhomologous 5' tails. BioTechniques 38: 34-38
- Regier JC, Zwick A, Cummings MP, Kawahara AY, Cho S, Weller S, Roe A, Baixeras J, Brown JW, Parr C, Davis DR, Epstein M, Hallwachs W, Hausmann A, Janzen DH, Kitching IJ, Alma Solis M, Yen S-H, Bazinet AL, Mitter C (2009) Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. BMC Evolutionary Biology 9: 280
- Regier JC, Shultz JW, Zwick A, Hussey A, BallB, Wetzer R, Martin JW, Cunningham CW (2010) Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463: 1079-1083
- Regier JC, Brown JW, Mitter C, Baixeras J, Cho S, Cummings MP, Zwick A (2012) A molecular phylogeny for the leaf-roller moths (Lepidoptera: Tortricidae) and its implications for classification and life history evolution. PLoS ONE 7(4): e35574. doi:10.1371/journal.pone.0035574
- Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, Sohn J-C, Zwickl DJ, Cho S, Davis DR, Baixeras J, Brown J, Parr C, Weller S, Lees DC, Mitter KT (2013) A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). PLoS ONE 8(3): e58568. doi:10.1371/journal.pone.0058568
- Ren F, Tanaka H, Yang Z (2005) An empirical examination of the utility of codon-substitution models in phylogeny Reconstruction. Systematic Biology 54: 808-818
- Richly E, Leister D (2004) NUMTs in sequenced Eukaryotic genomes. Molecular Biology and Evolution 21: 1081-1084

- Ronquist F, Huelsenback JP (2001) MrBayes version 3.1.2: Bayesian analysis of phylogeny. Available from http://mrbayes.sourceforge.net/
- Sage RF (2004) The evolution of C4 photosynthesis. New Phytologist 161: 341-370
- Saito O (2000) Flight activity of three *Spodoptera* spp, *Spodoptera litura*, *S. exigua* and *S. depravata*, measured by flight actograph. Physiological Entomology 25: 112-119
- Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: A Penalized Likelihood approach. Molecular Biology and Evolution 19: 101-109
- Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19: 301-302
- Schoonhoven LM, van Loon JJA, Dicke M (2005) *Insect-Plant Biology.* Second Edition. Oxford and New York: Oxford University Press. 421 pp.
- Shokralla S, Zhou X, Janzen DH, Hallwachs W, Landry J-F, Jacobus LM, Hajibabaei M (2011) Pyrosequencing for Mini-Barcoding of Fresh and Old Museum Specimens. PLoS ONE 6 (7): e21252.
- Siemers BM, Schauermann G, Turni H, Merten S von (2009) Why do shrews twitter? Communication or simple echo-based orientation. Biology Letters 5: 593-596
- Simmons ND (2005) An Eocene Big Bang for bats. Science 307: 527-528
- Simmons NB, Seymour KL, Habersetzer J, Gunnell GF (2008) Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. Nature 451: 818-821
- Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87: 651-701
- Simonsen TJ, Wahlberg N, Warren AD, Sperling FAH (2010) The evolutionary history of *Boloria* (Lepidoptera: Nymphalidae): phylogeny, zoogeography and larval foodplant relationships. Systematics and Biodiversity 8: 513-529
- Simonsen TJ, Zakharov EV, Djernaes M, Cotton AM, Vane-Wright RI, Sperling FAH (2011) Phylogenetics and divergence times of Papilioninae (Lepidoptera) with special reference to the enigmatic genera *Teinopalpus* and *Meandrusa*. Cladistics 27: 113-137
- Smith MA, Bertrand C, Crosby K, Eveleigh ES, Fernandez-Triana J, Fisher BL, Gibbs J, Hajibabaei M, Hallwachs W, Hind K, Hrcek J, Huang D-W, Janda M, Janzen DH, Li Y, Miller SE, Packer L, Quicke D, Ratnasingham S, Rodriguez J, Rougerie R, Shaw MR, Sheffield C, Stahlhut JK, Steinke D, Whitfield J, Wood M, Zhou X (2012) *Wolbachia* and DNA Barcoding Insects: Patterns, Potential and Problems. PLoS ONE 7(5): e36514. doi:10.1371/journal.pone.0036514
- Speakman J (2008) Evolutionary biology: A first for bats. Nature 451: 774-775
- Speidel W, Fänger H, Naumann CM (1996) The phylogeny of the Noctuidae (Lepidoptera). Systematic Entomology 21: 219-251
- Sperling FAH, Harrison RG (1994) Mitochondrial DNA variation within and between species of the *Papilio machaon* group of swallowtail butterflies. Evolution 48: 408-422
- Staats M, Cuenca A, Richardson J, Vrielink-van Ginkel R, Petersen G, Seberg O, Bakker FT (2011) DNA damage in plant herbarium tissue. PLoS ONE 6(12): e28448
- Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690
- Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the raxml web servers. Systematic Biology 57: 758-771
- Straten MJ van der, Bakker FT (2011) Morphological and moleculair identification of species of Spodoptera (FES part 2.2). Scientific report Plant Protection Service.
- Strömberg CAE (2011) Evolution of grasses and grassland ecosystems. Annual Review of Earth and Planetary Sciences 39: 517-544
- Sullivan J, Swofford DL (2001) Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? Systematic Biology 50: 723-729
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10: 512-526
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731-2739

- Tavaré S (1986). Some Probabilistic and Statistical Problems in the Analysis of DNA Sequences. Lectures on Mathematics in the Life Sciences (American Mathematical Society) 17: 57-86.
- Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ (2005) A molecular phylogeny for bats illuminates biogeography and the fossil record. Science 307: 580-584
- Todd EL, Poole RW (1980) Keys and illustrations for the armyworm moths of the noctuid genus *Spodoptera* Guenée from the Western hemisphere. Annals of the Entomological Society of America 73: 722-738
- Todisco V, Gratton P, Zakharov EV, Wheat CW, Sbordoni V, Sperling FAH (2012) Mitochondrial phylogeography of the Holarctic *Parnassius phoebus* complex supports a recent refugial model for alpine butterflies. Journal of Biogeography 39: 1058-1072
- Tojo S, Hayakawa Y, Phaophan P (2008) Strains in the common cutworm, *Spodoptera litura* (Lepidoptera: Noctuidae) with differing host ranges. Applied Entomology and Zoology 43: 491-496
- Velzen R van, Bakker FT, van Loon JJA (2007) DNA barcoding reveals hidden species diversity in *Cymothoe* (Nymphalidae). Proc. Neth. Entom. Soc. Meet. 18: 95–103.
- Vila R, Bell CD, Macniven R, Goldman-Huertas B, Ree RH, Marshall CR, Bálint Z, Johnson K, Benyamini D, Pierce NE (2011) Phylogeny and palaeoecology of *Polyommatus* show Beringia was a climate-regulated gateway to the New World. Proceedings of the Royal Society B 278: 2737-2744
- Wadall PJ, Penny D, Moore T (1997) Hadamard conjugations and modeling sequence evolution with unequal rates across sites. Molecular Phylogenetics and Evolution 8: 33-50
- Wahlberg N and Wheat CW (2008) Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. Systematic Biology 57: 231-242
- Wahlberg N, Weingartner E, Warren AD, Nylin S (2009) Timing major conflict between mitochondrial and nuclear genes in species relationships of *Polygonia* butterflies (Nymphalidae: Nymphalini). BMC Evolutionary Biology 9: 92
- Wahlberg N, Rubinoff D (2011) Vagility across *Vanessa* (Lepidoptera: Nymphalidae): mobility in butterfly species does not inhibit the formation and persistence of isolated sister taxa. Systematic Entomology 36: 362-370
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10: 57-63
- Whitfield JB, Kjer KM (2008) Ancient rapid radiations of insects: Challenges for phylogenetic analysis. Annual Review of Entomology 53: 449-72
- Wiegmann BM, Regier JC, Mitter C (2002) Combined molecular and morphological evidence on the phylogeny of the earliest lepidopteran lineages. Zoologica Scripta 31: 67-81
- Wiens JJ, Morrill MC (2011) Missing data in phylogenetic analysis: reconciling results from simulations and empirical data. Systematic Biology 60: 719-731
- Wilgenbusch JC, Warren DL, Swofford DL (2004) AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. Available from http://ceb.csit.fsu.edu/awty and http://king2.scs.fsu.edu/CEBProjects/awty/awty_start.php
- Wilson JW (1932) Notes on the biology of *Laphygma exigua* Huebner. The Florida Entomologist 16: 33-39
- Wilson JJ (2010) Assessing the value of DNA barcodes and other priority gene regions for molecular phylogenetics of Lepidoptera. PLoS ONE 5 (5): e10525
- Yack Je, Scudder GGE, Fullard JH (1999) Evolution of the metathoracic tympanal ear and its mesothoracic homologue in the Macrolepidoptera (Insecta). Zoomorphology 119: 93-103
- Yang Z (1996) Among-site rate heterogeneity and its impact on phylogenetic analysis. Trends in Ecology & Evolution 11: 367-372
- Yang Z, Rannala B (2012) Molecular phylogenetics: principles and practice. Nature Reviews 13: 303-314
- Yeates DK, Cameron SL, Trautwein M (2012) A view from the edge of the forest: recent progress in understanding the relationships of the insect orders. Australian Journal of Entomology 51: 79-87
- Yu Y, Harris AJ, He X-J (2013) RASP (Reconstruct Ancestral State in Phylogenies) 2.1 beta. Available at http://mnh.scu.edu.cn/soft/blog/RASP
- Yu Y, Harris AJ, He XJ (2010) S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. Molecular Phylogenetics and Evolution. 56: 848-850
- Zachos J, Pagani M, Sloan L, Thomas E, Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292: 686-693

- Zahiri R, Kitching IJ, Lafontaine JD, Mutanen M, Kaila L, Holloway JD, 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 JD, Kitching IJ, Lafontaine JD, Mutanen M, Wahlberg N (2012) Molecular phylogenetics of Erebidae (Lepidoptera, Noctuoidea). Systematic Entomology 37: 102-124
- Zharkikh A (1994) Estimation of evolutionary distances between nucleotide sequences. Journal of Molecular Evolution 39: 315-329
- Zhou X, Xu S, Xu J, Chen B, Zhou K, Yang G (2011) Phylogenomic analysis resolves the interordinal relationships and rapid diversification of the Laurasiatherian Mammals. Systematic Biology 61: 150-164

Appendices

Appendix 1: Valid *Spodoptera* species
Appendix 2: Wish list
Appendix 3: Original planning
Appendix 4: Budget
Appendix 5: Letters of recommendation, personal invitation and evaluation report of my presentation at the International Lepidopterists' Conference 2012 in Denver, Colorado, USA
Appendix 6: Master thesis agreement
Appendix 7: Results of one-gene and concatenated alignment analysis
Appendix 8: Calibration combination results per clade

Appendix 1: Valid *Spodoptera* species

Species	Author, year	Distribution	Host
Spodoptera albula	Walker, 1857	Southern Nearctic, Neotropical	Strongly polyphagous
Spodoptera androgea	Stoll in Cramer, 1782	Southern Nearctic, Neotropical	Strongly polyphagous
Spodoptera apertura	Walker, 1865	Afrotropical, Oriental, Australasian	Unknown
Spodoptera cilium	Guenée, 1852	Southern Palearctic, Afrotropical, Oriental, Northern Australasian	Narrowly polyphagous on 'grasses'
Spodoptera compta	Walker, 1869	Neotropical (Peru only)	Unknown
Spodoptera cosmioides	Walker, 1858	Neotropical	Unknown
Spodoptera depravata	Butler, 1879	Eastern Palearctic	Unknown
Spodoptera descoinsi	Lalanne-Cassou & Silvain, 1994	Equatorial Neotropic	Probably strongly polyphagous
Spodoptera dolichos	Fabricius, 1794	Nearctic, Neotropical	Strongly polyphagous
Spodoptera eridania	Stoll in Cramer, 1782	Southern Nearctic, Neotropical	Strongly polyphagous
Spodoptera evanida	Schaus, 1914	Equatorial Neotropics	Unknown
Spodoptera exempta	Walker, 1856	Southern Palearctic, Afrotropical, Oriental, Australasian	Narrowly polyphagous on 'grasses'
Spodoptera exigua	Hübner, 1803/08	Cosmopolitan	Strongly polyphagous
Spodoptera frugiperda	Smith & Abbot, 1797	Nearctic, Neotropical	Strongly polyphagous
Spodoptera hipparis	Druce, 1887	Southern Nearctic, Northern Neotropic	
Spodoptera latifascia	Walker, 1856	Southern Nearctic, Northern Neotropic	Strongly polyphagous
Spodoptera littoralis	Boisduval, 1833	Southern Palearctic, Afrotropical	Strongly polyphagous
Spodoptera litura	Farbicius, 1775	Oriental, Eastern Palearctic, Australasian	Strongly polyphagous
Spodoptera malagasy	Viette, 1967	Madagascar	Unknown
Spodoptera marima	Schaus, 1904	Neotropical	Probably strongly polyphagous
Spodoptera mauritia	Boisduval, 1833	Eastern Afrotropical, Oriental, Australasian	Narrowly polyphagous on 'grasses'
Spodoptera ochrea	Hampson, 1909	Neotropical (Ecuador and Peru)	Unknown
Spodoptera ornithogalli	Guenée, 1852	Nearctic, Northern Neotropic	Strongly polyphagous
Spodoptera pecten	Guenée, 1852	Oriental, Northern Australasian	Strongly polyphagous
Spodoptera pectinicornis	Hampson, 1895	Oriental, Northern Australasian	Monophagous on *Pistia stratioites*
Spodoptera picta	Guérin-Meneville, 1830	Oriental, Eastern Palearctic, Australasian	Oligophagous on Amaryllidaceae
Spodoptera praefica	Grote, 1875	Western Nearctic	Strongly polyphagous
Spodoptera pulchella	Herrich-Schäffer, 1868	Caribbean	Unknown, only found on cotton
Spodoptera roseae	Schaus, 1923	Galapagos Islands	Unknown
Spodoptera triturata	Walker, 1856	Afrotropical	Narrowly polyphagous on 'grasses'
Spodoptera umbraculata	Walker, 1858	Eastern Australia	Unknown

Appendix 2: Original wish list

These were the species I hoped to be able to sequence in my thesis project, in order to add them to the pre-existing *Spodoptera* data set.

Category	Species
Look-a-like	*Neogalea sunia*
Look-a-like	*Trichordesta prodeniformis*
Look-a-like	*Copitarsia* spp.
Look-a-like	*Antachara diminuta*
Look-a-like	*Elaphria nucicolora*
Allied genera	*Elaphria* spp
Allied genera	*Galgula* spp
***Spodoptera*, missing**	*S. hipparis*
***Spodoptera*, missing**	*S. apertura*
***Spodoptera*, missing**	*S. compta*
***Spodoptera*, missing**	*S. cosmioides*
***Spodoptera*, missing**	*S. depravata*
***Spodoptera*, missing**	*S. descoinsi*
***Spodoptera*, missing**	*S. evanida*
***Spodoptera*, missing**	*S. marima*
***Spodoptera*, missing**	*S. pectinicornis*
***Spodoptera*, missing**	*S. praefica*
***Spodoptera*, missing**	*S. roseae*
***Spodoptera*, missing**	*S. umbraculata*
***Spodoptera*, low sampling**	*S. cilium*
***Spodoptera*, low sampling**	*S. ochrea*
***Spodoptera*, low sampling**	*S. picta*
***Spodoptera*, low sampling**	*S. triturata*
***Spodoptera*, low sampling**	*S. malagasy*
***Spodoptera*, low sampling**	*S. pecten*
***Spodoptera*, resample**	*S. mauritia* FIJI

Appendix 3: Original planning

This was the original planning of the project. Due to many unexpected events I (in consultation with my supervisors) greatly deviated from this planning

Week	Monday date	Stage	Activity
1	12-3-2012	Phylogeny	Create a species wish list, gather specimens
2	19-3-2012		DNA isolation, PCR optimization
3	26-3-2012		Generation amplicons
4	2-4-2012		Sanger sequencing
5	9-4-2012		Sequence trace analysis & Sequence alignment
6	16-4-2012		Tree building & PCA on morphological characters
7	23-4-2012		Tree building
8	30-4-2012		Lab
9	7-5-2012		Tree building
10	14-5-2012		Making CAD, GAPDH & IDH alignments
11	21-5-2012		Lab
12	28-5-2012	Morphology	Data collection
13	4-6-2012		Phylogenetic analysis
14	11-6-2012		Analysis
15	18-6-2012	Report	Start writhing
16	25-6-2012		Writhing
17	2-7-2012		Writhing
18	9-7-2012		Hand in thesis report
19	16-7-2012		Travel to Denver, Colorado
20	23-7-2012		Congress Lepidopterists' Society

Appendix 4: Budget

Expenses		**Budget**	
a) Flight tickets	€1258.00	Requested:	
b) Conference fee	€58.50	I. Wageningen University Fund	
c) Printing poster	€10.00	*50% of expenses (following WUF guidelines) = 0.5*(a+b+c+d+e+g)*	€856.50
d) Visa	€96.00		
Hotel accommodation:			
e) *Night before and after presentation*	€95.00	II. Uyttenboogaart-Eliasenstichting	
f) *5 other nights during congress*[1]	€237.00	*Standard amount EUS*	€1000.00
g) Unforeseen costs (10% of budget)	€195.50	Own contribution[2] = *Total expenses – I – II*	€93.50
Total	**€1950.00**	**Total**	**€1950.00**

[1] In order to make the most out of this visit, I want to take part in all conference-related activities, such as plenary sessions and symposia, for networking purposes.

[2] Own contribution will be expanded with other expenses:
- Food and drink
- Excursions
- Local travel expenses

Authorisation

Date: 30/3/2012

Student: Mark Lammers

Supervisor: Frederik T. Bakker

Appendix 5: Letter of recommendation, personal invitation and evaluation report of my presentation at the International Lepidopterists' Conference 2012 in Denver, Colorado, USA

WAGENINGEN UNIVERSITY
WAGENINGEN UR

For quality of life

P.O. Box Postbus 647 | 6700 AP Wageningen | The Netherlands

Wageningen University Fund

Biosystematics Group

DATE
March 30, 2012

SUBJECT
Application Lammers

POSTAL ADDRESS
P.O. Box Postbus 647
6700 AP Wageningen
The Netherlands

VISITORS' ADDRESS
Wageningen Campus
Building 107
Droevendaalsesteeg 1
6708 PB Wageningen

INTERNET
http://www.bis.wur.nl/UK
www.wageningenuniversity.nl

CoC NUMBER
09215846

HANDLED BY
Freek T. Bakker

TELEPHONE
+31 (0)317 483175

EMAIL
freek.bakker@wur.nl

Dear Sir/Madam,

With this letter I'd like to fully endorse the enclosed application by Mark Lammers, whom I am supervising at the moment during his master thesis.

For a keen & enthousiastic young scientist such as mr. Lammers participating in an international professional conference such as the annual meeting of the Lepidopterist's Society, held this year in Denver, Colorado USA, will provide an excellent opportunity for getting to know a specific scientific community (and its *modus operandi*), broadening his scientific horizon, and improving his skills.

In my opinion, the work covered in mr. Lammers' thesis (i.e. *Spodoptera* systematics) would suit very well in the meeting concerned and would significantly contribute to both the Noctuid systematics community as well as to a further (international) exposure/public awareness of WUR.

With kind regards,

Freek T. Bakker
Assistant Professor
Biosystematics group

DENVER MUSEUM OF NATURE & SCIENCE

Dr Frank T. Krell
Curator of Entomology
Department of Zoology

24 April 2012

Mark Lammers
c/o Erik J. van Nieukerken
Netherlands Centre for Biodiversity Naturalis (NCB Naturalis)
Dep. of Terrestrial Zoology
PO Box 9517
NL-2300 RA Leiden
The Netherlands

Invitation to present a paper at the international Lepidopterist's Conference in Denver, 23-29 July 2012

Dear Mr. Lammers,

it is a pleasure to invite you to present your research on the molecular phylogenetics of the pest genus *Spodoptera* Guenée, 1842 (Lepidoptera: Noctuidae) and related taxa at the combined Annual Meeting of the Lepidopterists' Society and the Societas Europaea Lepidopterologica that we will be hosting from 23 to 29 July 2012 at the Denver Museum of Nature & Science. You find relevant information and a link to the registration form on the conference on the conference webpage (www.dmns.org/krell-lab). Please note that the Denver Museum of Nature & Science can neither pay for nor subsidize travel or registration costs.

We would be delighted to welcome you here in Denver.

Please let me know if I can be of any further help.

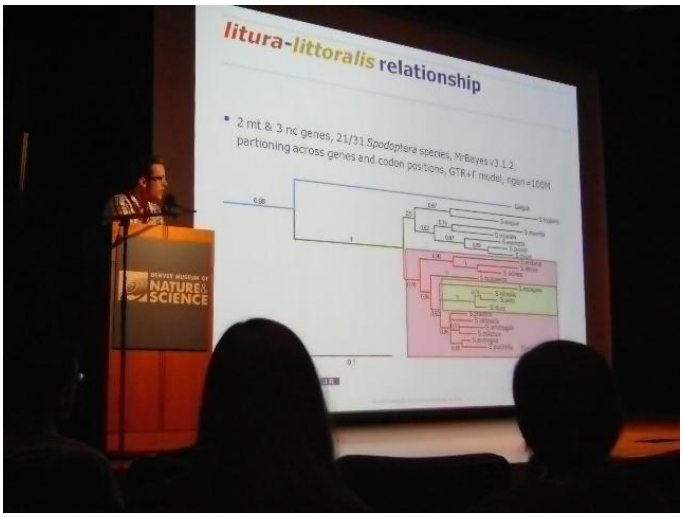
Sincerely

Dr Frank-T. Krell

Conference Host, frank.krell@dmns.org
Conference email: lepidopterist@dmns.org

2001 Colorado Blvd.
Denver, CO 80205-5798
P 303.322.7009
F 303.331.6492
www.dmns.org

Evaluatieverslag 'Combined Annual Meeting of the Lepidopterists' Society and the Societas Europaea Lepidopterologica', 23-29 juli 2012, te Denver, Colorado.
Presentatie gehouden met de titel "Dating *Spodoptera* Guenée, 1852 (Lepidoptera: Noctuidae) clades using fossil and non-fossil evidence", op 25 juli, 10:00 uur.



Na enkele dagen vakantie in de Rocky Mountains was ik ruimschoots hersteld van de lange reis naar Denver, Colorado. De ochtend van mijn grote moment, mijn entree in de internationale gemeenschap, mijn visitekaartje aan de wetenschap, was eindelijk aangebroken. Goed voorbereid en ietwat gespannen hield ik daar mijn presentatie voor tussen de 100 en 150 aanwezigen, onder wie zowel fanatieke amateurs als vooraanstaande wetenschappers. Naderhand kreeg ik twee vragen: Robert Pyle, schrijver van vele boeken over dagvlinders, vroeg of ik andere voorbeelden van verwante soorten ken die een vergelijkbaar verspreidingspatroon laten zien, waar ik het antwoord op schuldig moest blijven. Felix Sperling, een vooraanstaande Canadese wetenschapper, vroeg hoe het zit met de substitutiesnelheden die ik gevonden had en welke impliciete aannames daarin zaten, waarop ik kon antwoorden dat mijn substitutiesnelheden goed in de bestaande literatuur vallen. Over de aannames heb ik naderhand in een pauze met hem verder gediscussieerd.

Gedurende de dagen na de presentatie kreeg ik vele complimentjes over de kwaliteit en inhoud van mijn presentatie. Het leukst om te horen vond ik dat meerdere mensen vonden dat ik de beste presentatie van alle studenten gehouden had en dat ik de student paper competition zou moeten winnen. Het voornaamste argument dat gegeven werd, was dat ik duidelijk mijn probleemstelling uiteenzette, een heldere aanpak van het probleem had en een concrete conclusie kon geven aangaande het probleem. Helaas was de jury het niet eens met deze mensen en heeft een Amerikaanse student gewonnen.
Ook vroegen veel mensen mij of ik hun mijn thesis wilde opsturen. Sommigen uit interesse voor het onderwerp en de relevantie voor hun werkveld, anderen uit interesse in de methodes omdat die ook toepasbaar zijn op hun onderzoeksveld.
Daarnaast kreeg ik ook enkele kritische opmerkingen over enkele onderdelen, waar ik nog niet over nagedacht had. Dat was een zeer welkome aanvulling! Hoewel ik geen tijd meer heb om al deze punten uit te zoeken, gaat het me wel lukken alles in te passen in de discussie van mijn thesis.

Het congres zelf had een fijne open sfeer. Alle aanwezigen benaderden elkaar als gelijkwaardig, autoriteit of leeftijd speelde geen merkbare rol. Het was heel leuk allemaal mensen te ontmoeten van wie ik vele artikelen gelezen had. Net zoals ik veel kon leren van de tips die ik van anderen kreeg, had ik ook tips aan met name andere studenten die worstelen met dezelfde software als ik gebruikt heb. Ook was het congres een uitstekende manier om mijn Engels te oefenen.
Verder werden er presentaties gehouden uit onderzoeksvelden waarvan ik niet eens wist dat ze bestonden. Het merendeel daarvan was nog interessant ook!

Al dit tezamen heeft mij een grote motivatie gegeven om door te gaan als wetenschapper. Het was een enorm inspirerende meeting. Ik heb vele ideeën en aanvullingen gekregen voor dit onderzoek en ook voor vervolgstudies. Daarnaast was het een uitgelezen plek om te netwerken en daarmee heb ik enkele zeer interessante potentiële stageplaatsen gevonden.

Ik wil het Wageningen Universiteitsfonds en de Uyttenboogaart-Eliasenstichting bedanken voor de verstrekte subsidies. Zonder hen was het bijwonen van dit internationale congres en het houden van een presentatie daar onmogelijk en onbetaalbaar geweest voor een masterstudent als ik.
Ook bedank ik alle mensen die mij geholpen hebben zo ver te komen, in het bijzonder mijn twee begeleiders: Freek Bakker en Marja van der Straten.

Financiële verantwoording
De posten komen overeen met de ingediende begroting. Ter vergelijking zijn de begrote bedragen rechts ingevoegd. Daarnaast staan nog nummers voor opmerkingen, die refereren naar de onderste tabel.
Een aantal bedragen zijn hoger uitgevallen. Dit kwam doordat ik pas laat kon boeken, omdat het WUF pas laat (3 juli) groen licht gegeven had.

Inkomsten		**Begroot**	**Opmerking**
Subsidies			
I. Bijdrage WUF	€ 933,07	€ 856,50	1
II. Bijdrage UES	€ 1.000,00	€ 1.000,00	
Totaal subsidies	**€ 1.933,07**	€ 1.856,50	
Eigen bijdrage	**€ 195,57**	€ 93,50	2
= Totale kosten – I – II			
Groot totaal	**€ 2.128,64**	**€ 1.950,00**	

Uitgaven	**In andere valuta**	**In euro's**	**Bon-nummer**	**Begroot**	**Opmerking**
a) Vliegtickets					
Ticket Amsterdam - Denver		€ 710,00	L2		
Ticket Denver - Lulea		€ 906,13	L3		3
Totaal a) Vliegtickets		**€ 1.616,13**		**€ 1.258,00**	4
b) Conference fee	**$ 135,00**	**€ 107,52**	L4	**€ 58,50**	5
c) Drukkosten poster		**€ -**		**€ 10,00**	6
d) Visum	**$ 14,00**	**€ 11,24**	L5	**€ 96,00**	7
e) Nacht voor en na presentatie	**$ 158,36**	**€ 131,25**	L1	**€ 95,00**	
f) 4 andere nachten gedurende het congres	**$ 316,72**	**€ 262,50**	L1	**€ 237,00**	8
Totaal Hotel		€ 393,75		€ 332,00	9
g) Onvoorziene kosten		**€ -**		**€ 195,50**	10
Groot totaal		**€ 2.128,64**		**€ 1.950,00**	

Opmerkingen	
1	Hoger dan begroot, aangezien de kosten ook hoger uitvielen. Dit bedrag is exact volgens de regels van het WUF: 50% van totale kosten = 0,5*(a+b+c+d+e). Daarnaast is ook de eigen bijdrage hoger dan begroot, dit zit ook in de vliegticketprijzen.
2	Bovenop dit bedrag kwam ook nog al het eten, excursies, het banket en dergelijke (= 352,62 euro).
3	Op het moment van boeken was de kale prijs voor het goedkoopste ticket naar Amsterdam 832 euro, 7 euro goedkoper dan naar Luleå. Om die reden lijkt het mij verantwoordbaar om de financiering van deze alternatieve bestemming te rechtvaardigen.
4	Hoger door late booking.
5	Hoger (30 dollar) door late inschrijving.
6	Het houden van een presentatie sloot het opzetten van een poster uit.
7	Het begrote bedrag was gebaseerd op de oude regelgeving.
8	Op de begroting stonden in totaal 7 nachten gepland, terwijl er maar 6 gedurende het congres waren.
9	Belastingen (=10,18 euro per nacht) zaten nog niet bij de prijs in.
10	Compenseert voor onvoorziene belastingen en hogere vliegticketprijzen.

Appendix 6: Master thesis agreement

Master Thesis Proposal Biology

Intake Form

Please note:

This form should be fully completed and signed before the start of the work. One photocopy of the completed form should be returned instantly by the student to the secretary: Marianne v.d. Peppel, bode 69.

Statement study advisor:

(Name:) Peter de Jong, study advisor for MSc programme Biology,

herewith certifies that student (named:) Mark Lammers

is granted official permission for starting his/her Master Thesis Proposal as part of the MSc programme.

Date: 9-2-2012 Signature: p.o.

You must have finished your two majors needed for your Thesis before you can take part in this course. If not, an appointment with Peter the Jong is necessary!

Personal data student:

Name: Mark Lammers Reg. nr.: 890206498090

Address: Haarweg 309d 6709 RX Wageningen

Tel.: 06-18040529

Personal data (1st) supervisor:

Name: Freek T. Bakker

Chair group: Biosystematiek

Function: UD Tel.: 83175

Information on Master Thesis Proposal:

Subject / title: Systematics of Spodoptera Guenée 1852 (Lepidoptera: Noctuidae)

Planning:

1. Start of the project: 6/2/'12
2. Delivering concept research plan: 22/2/'12
3. Oral presentation 6/3/'12
4. Delivering final version research plan: 12/3/'12

Further arrangements: Co-supervision by Mrs Marijke van der Straten (n/VWA, Wageningen PD)

Signatures:

(1st) Supervisor: Student:

date: 9 feb. date: 9/2/12

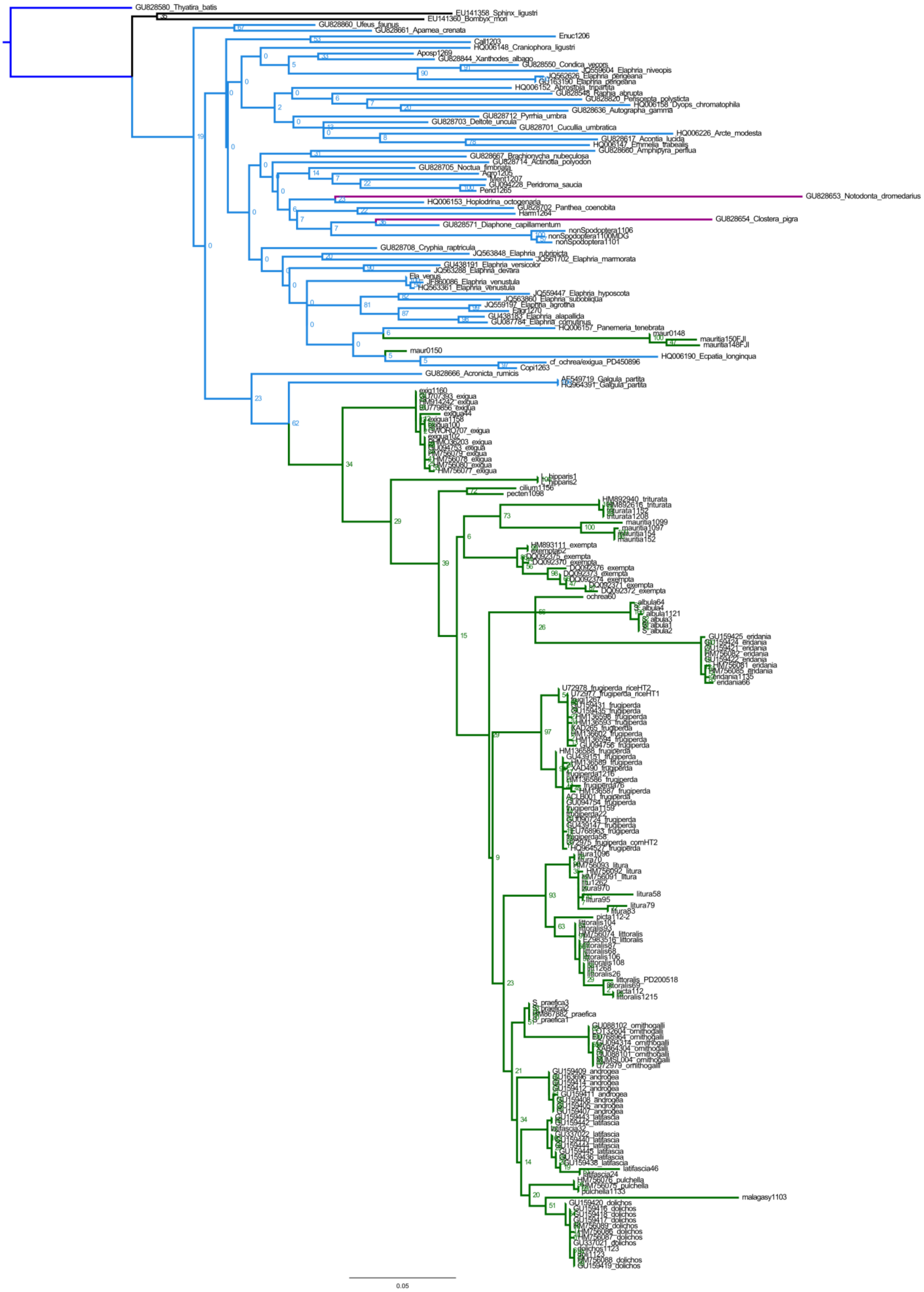
Agreed on behalf of the MSc programme coordinator:

(Signature) , date 13-02-2012

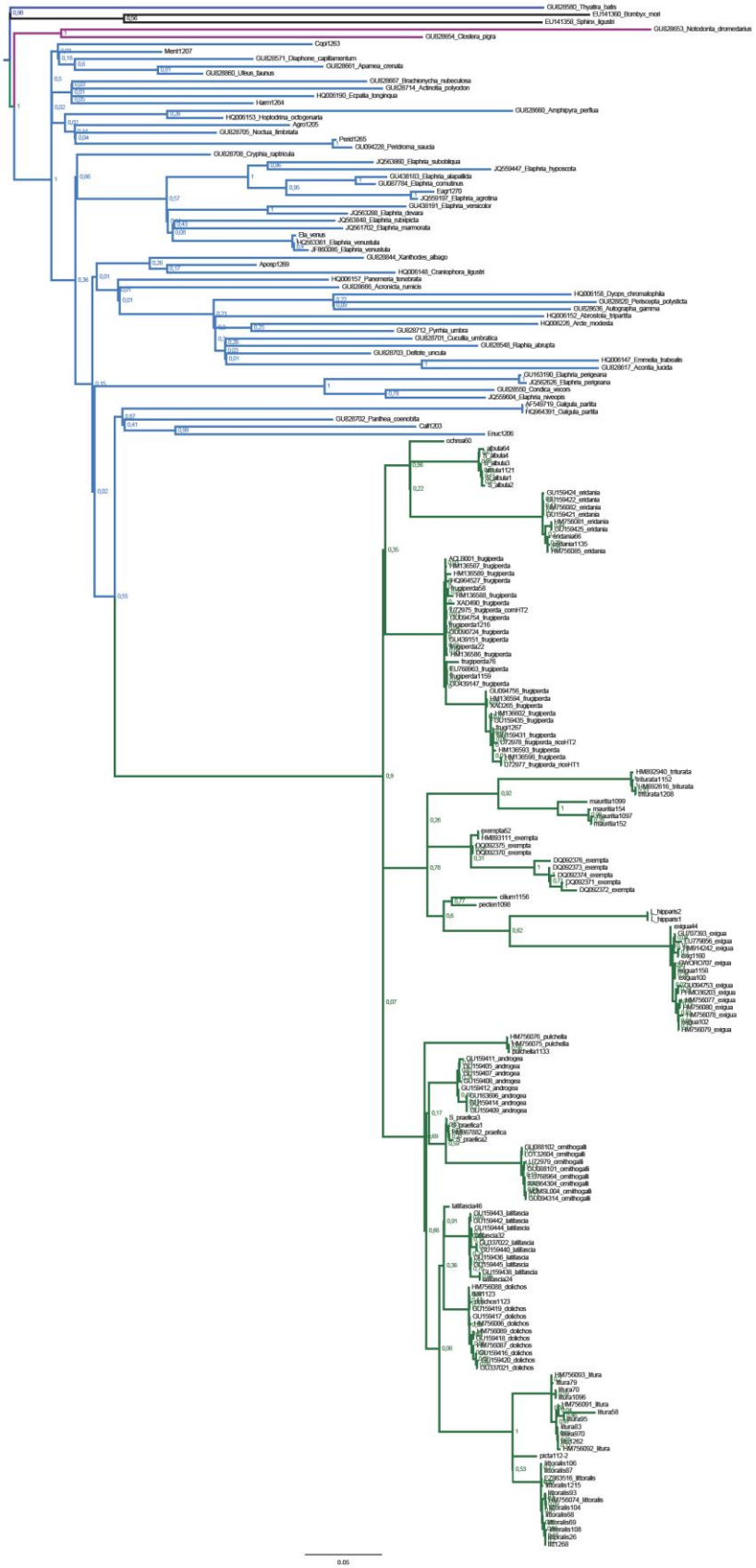
Appendix 7: Results of one-gene and concatenated alignment analyses

COI

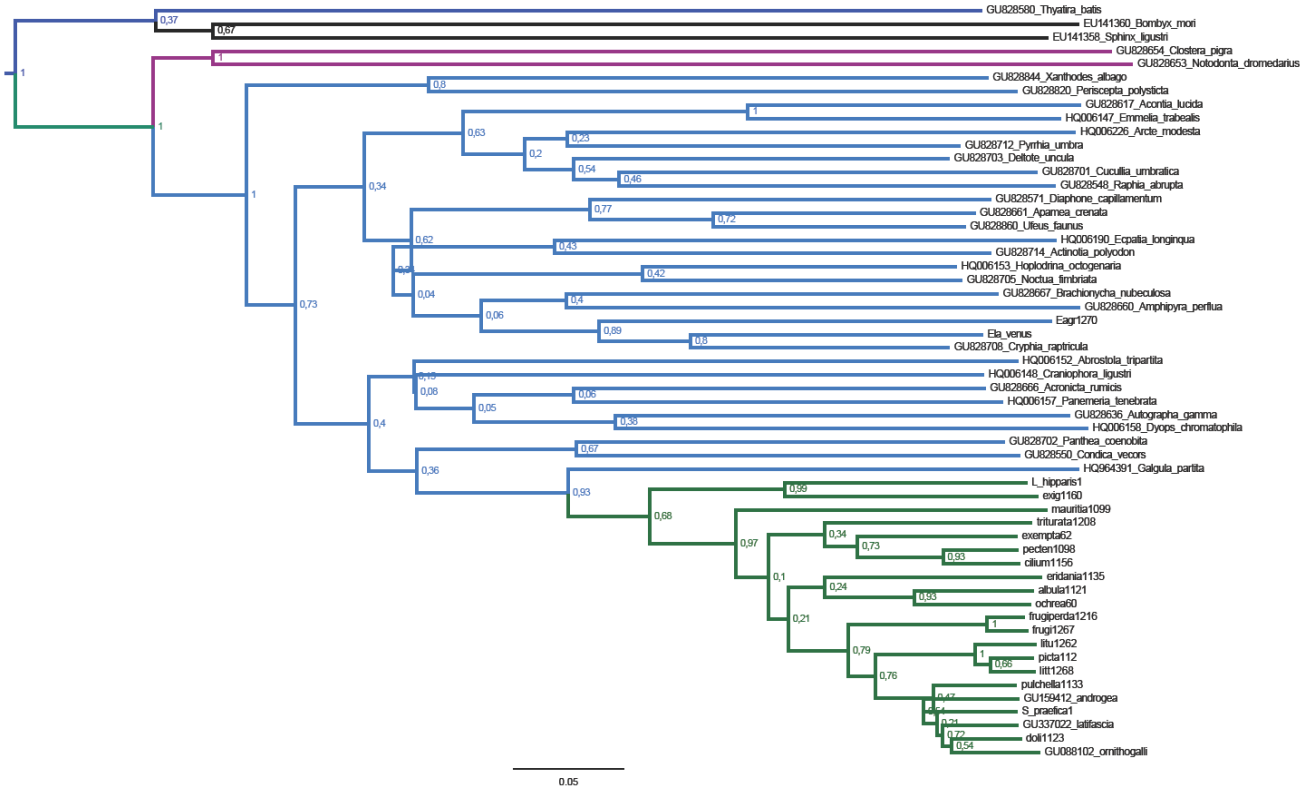
COI/221SpodCOIRAxML/longrun/221SpodCOIRAxML.tre.pdf



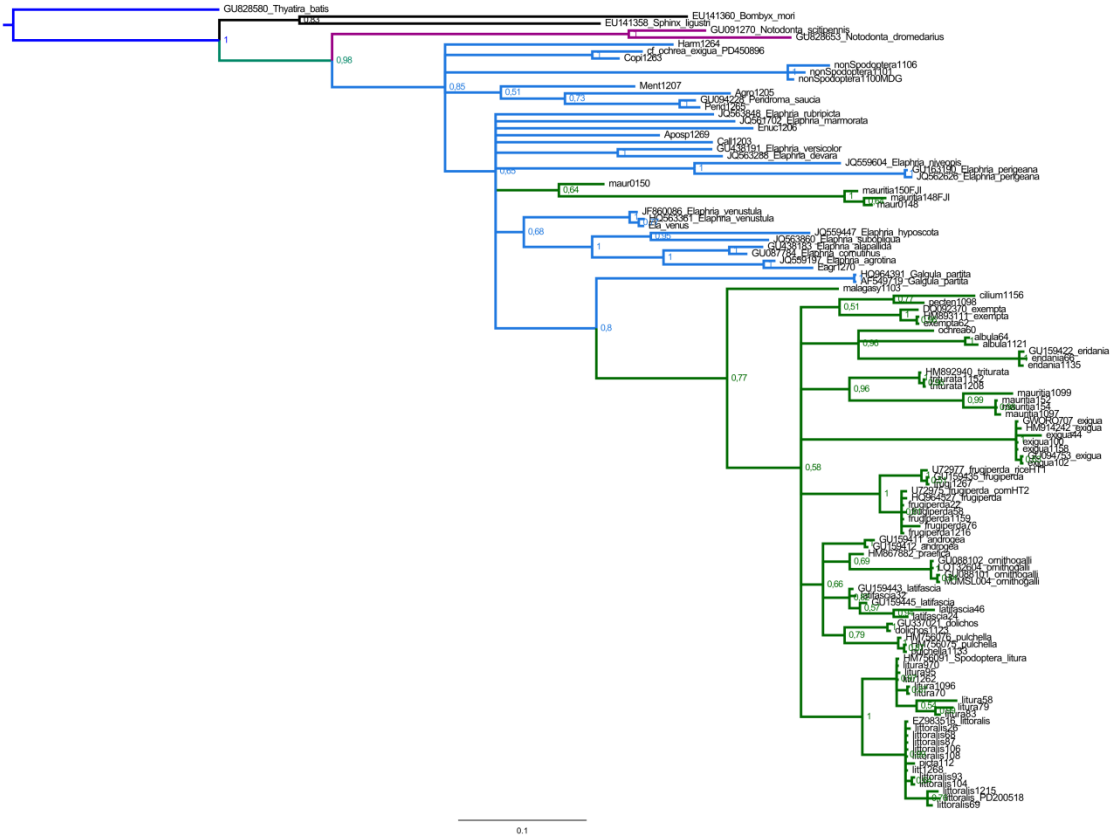
COI/210strictSpodCOIBEAST/210strictSpodCOIBEASTlognormalbat200mld.(subst).tree.pdf



COI/56SpodCOIBEAST/56SpodCOIBEASTlognormalnotod100mln.(subst).tree.pdf

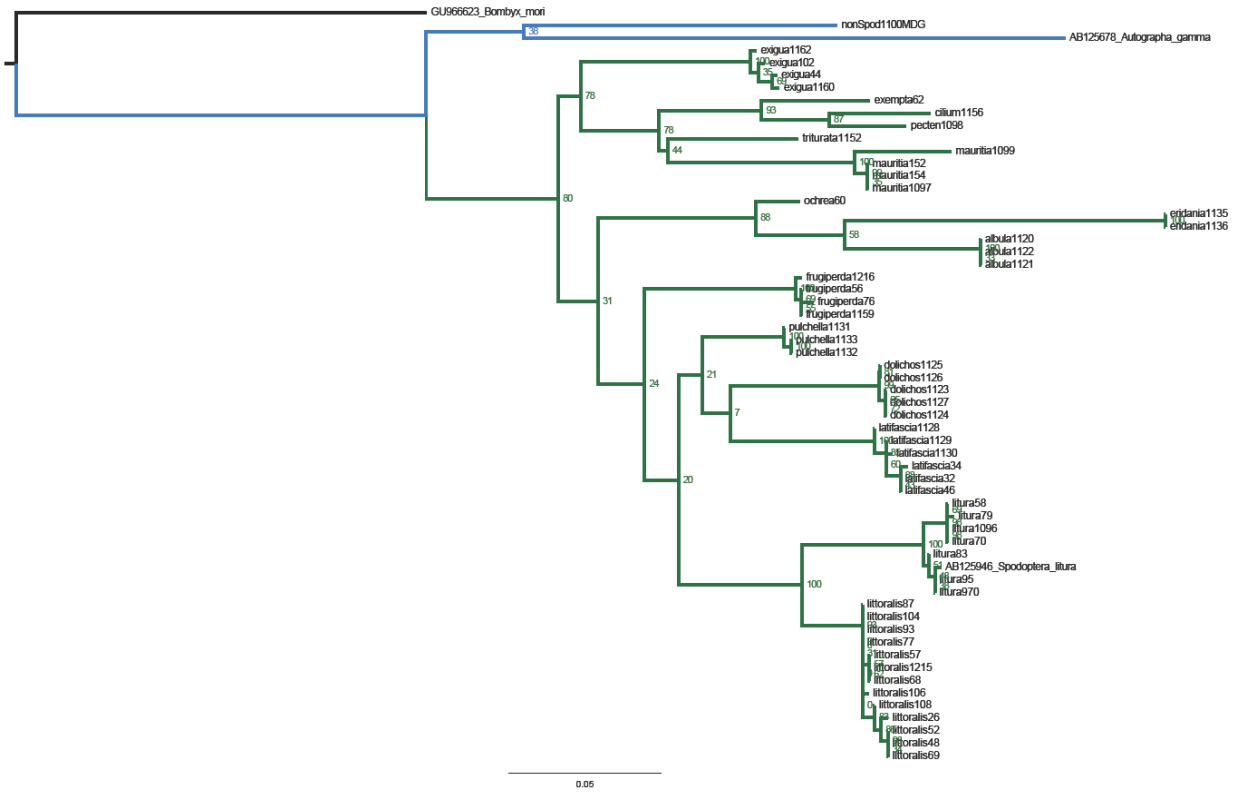


COI/116SpodCOIBayesCodonHotnoArch/116SpodCOIBayesHotnoArch.nex.con.tre.pdf

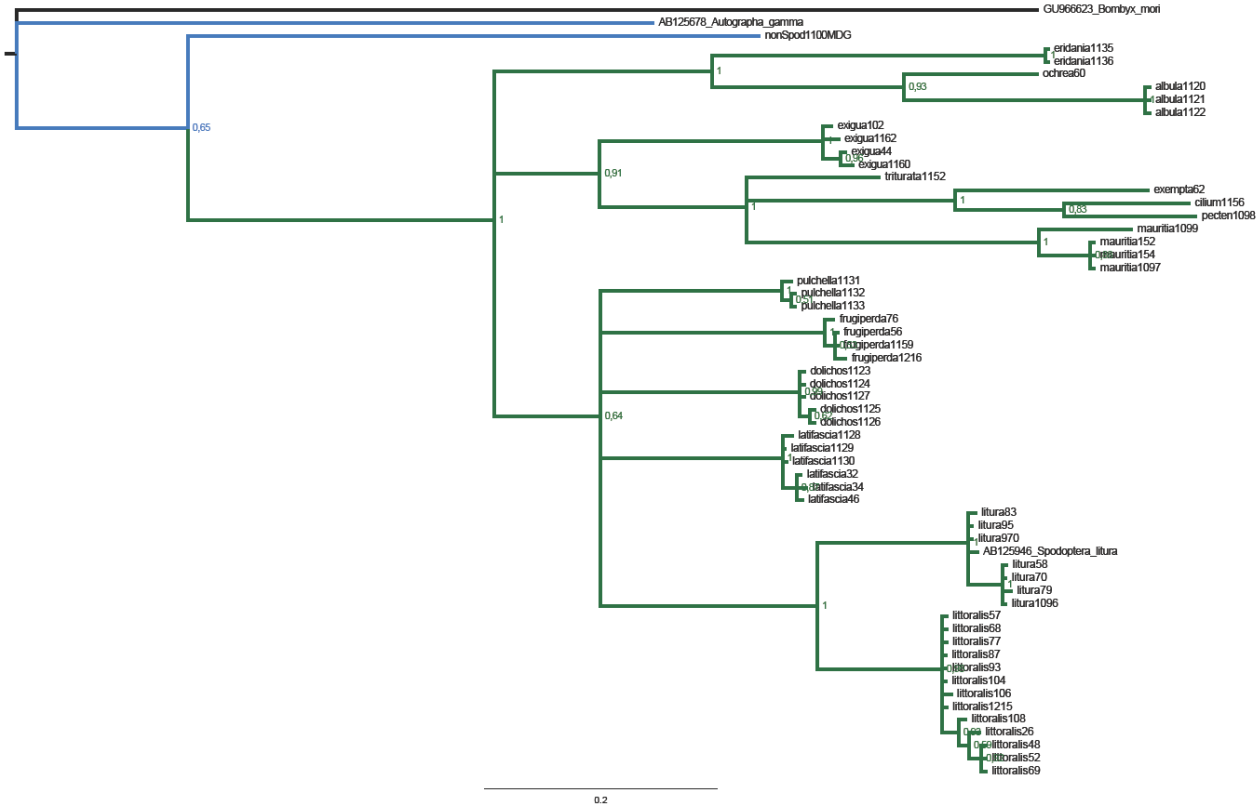


CytB

CytB/60SpodCytBRAxML/60SpodCytBRAxML.tre.pdf

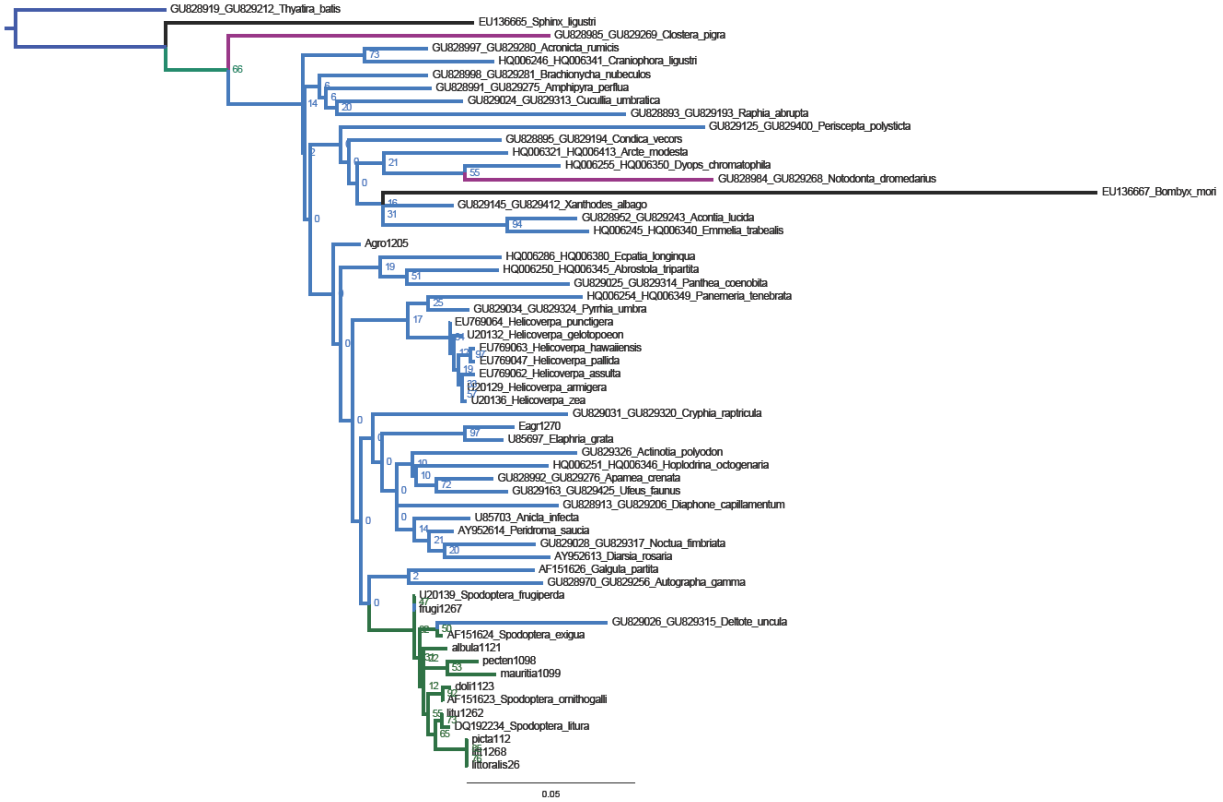


CytB/60SpodCytBBayes/60SpodCytBBayes100mln.nex.con.tre.pdf



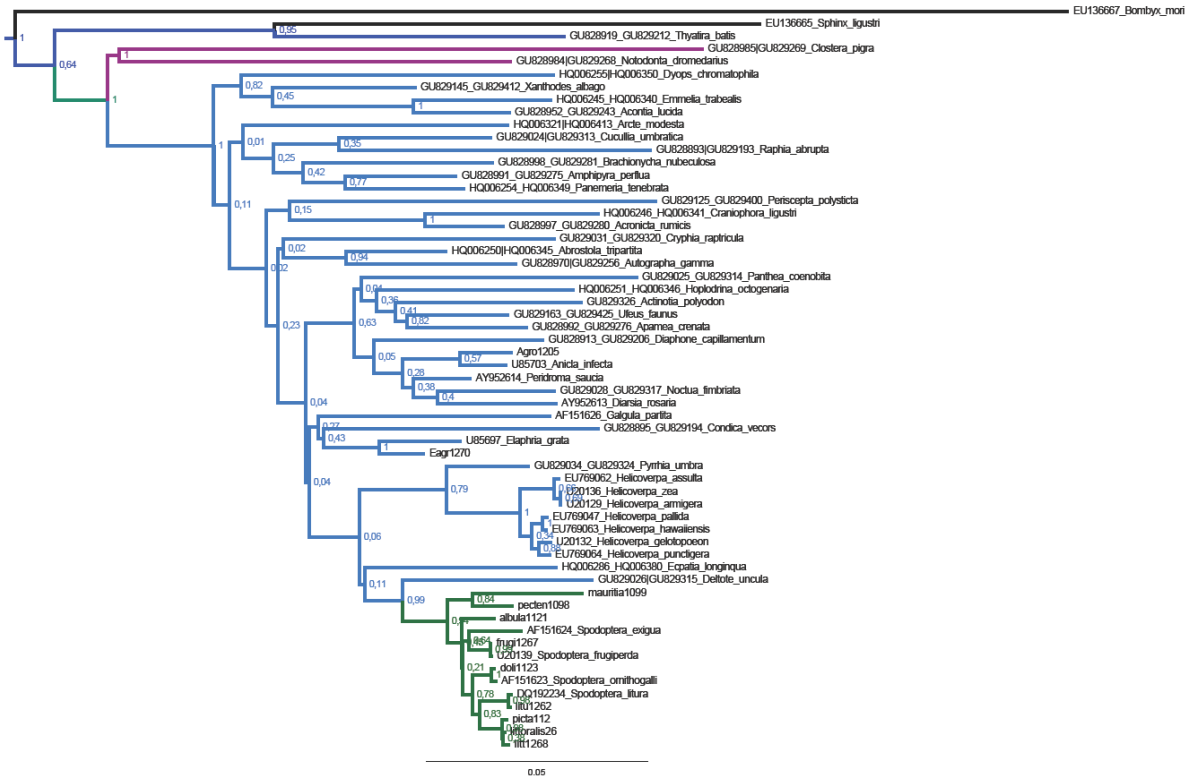
EF1α

EF1a/59SpodEF1aRAxML/59SpodEF1aRAxML.tre.pdf



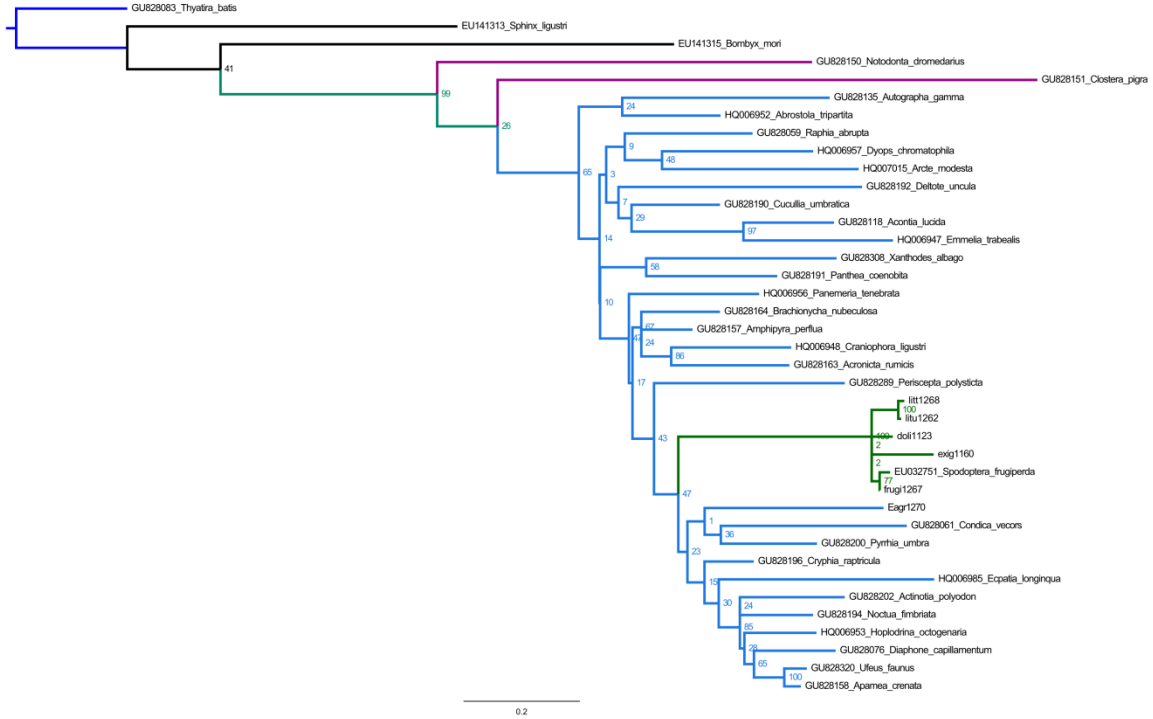
RAxML is weak!

EF1a/59SpodEF1aBEAST/59SpodEF1aBEASTnotod100mln.(subst).tree.pdf

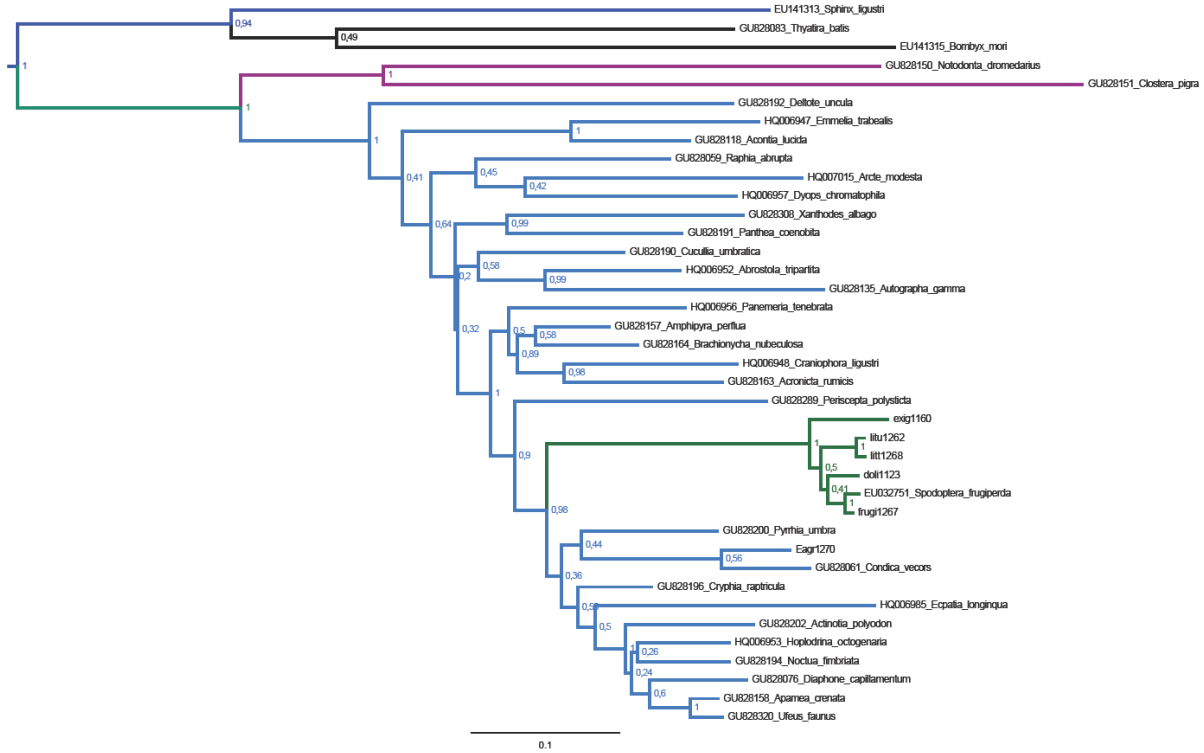


CAD

CAD/39SpodCADRAxML/39SpodCADRAxML.tre.pdf

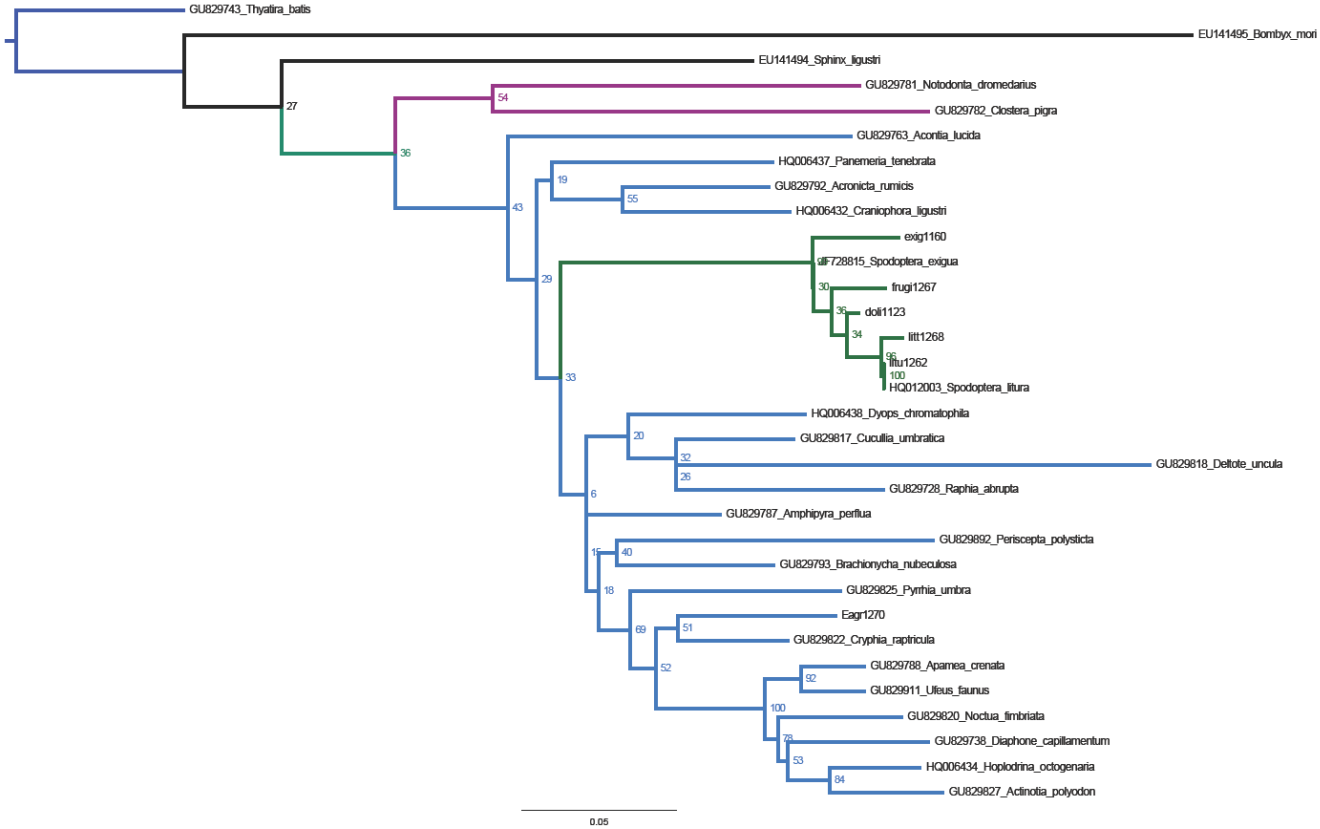


CAD/39SpodCADBEAST/39SpodCADBEASTlognormalnotodbat100mln.(subst).tree.pdf

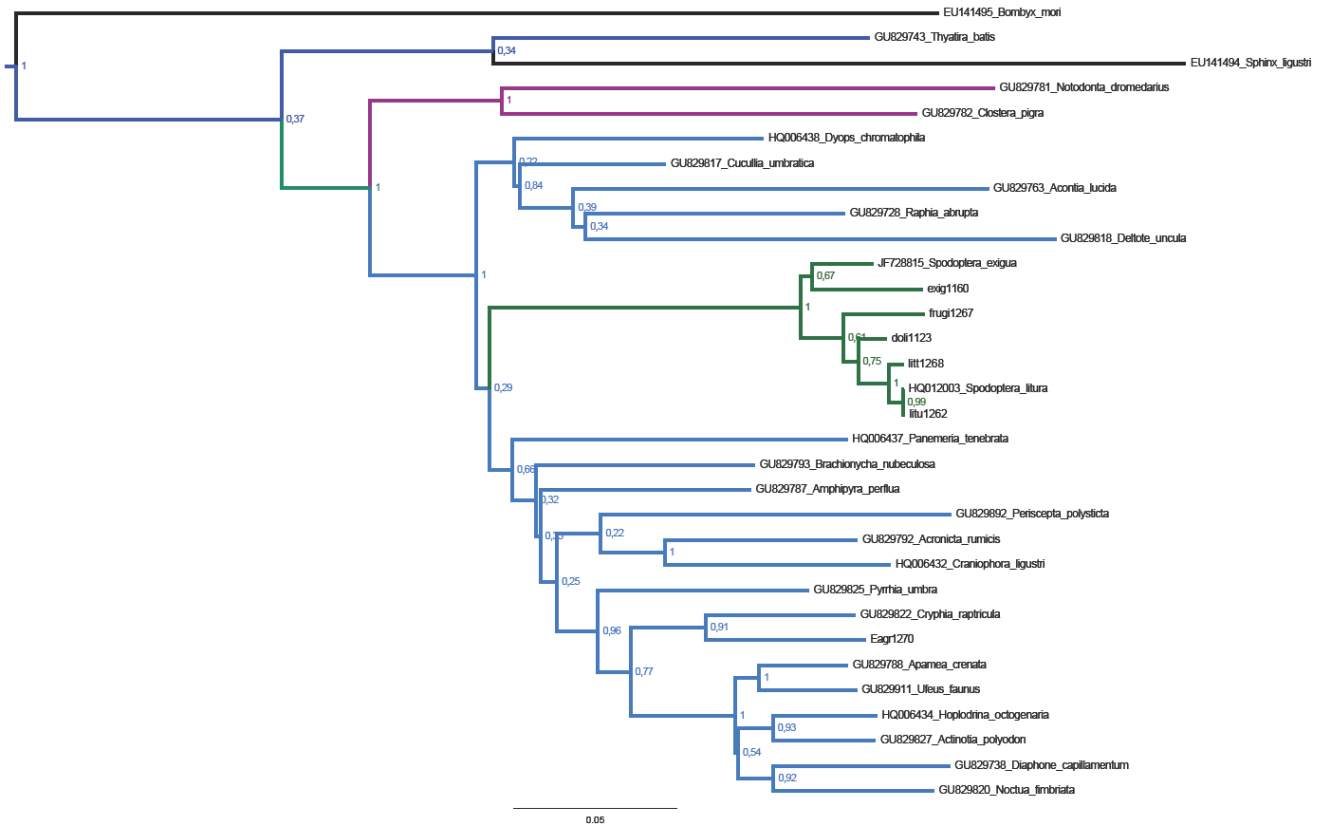


GAPDH

GAPDH/32SpodGAPDHRAxML/32SpodGAPDHRAxML.tre.pdf

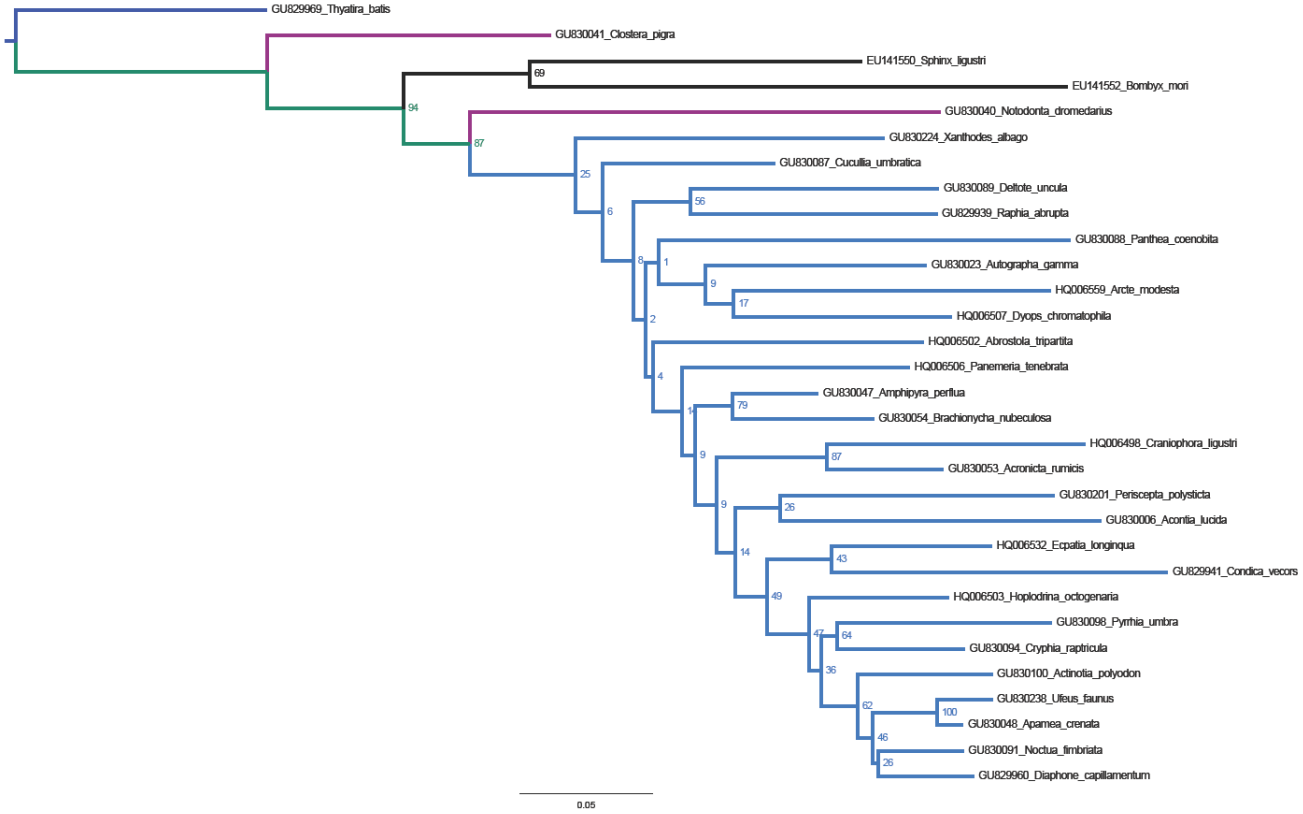


GAPDH/32SpodGAPDHBEAST/32SpodGAPDHBEASTnotod100mln.(subst).tree.pdf

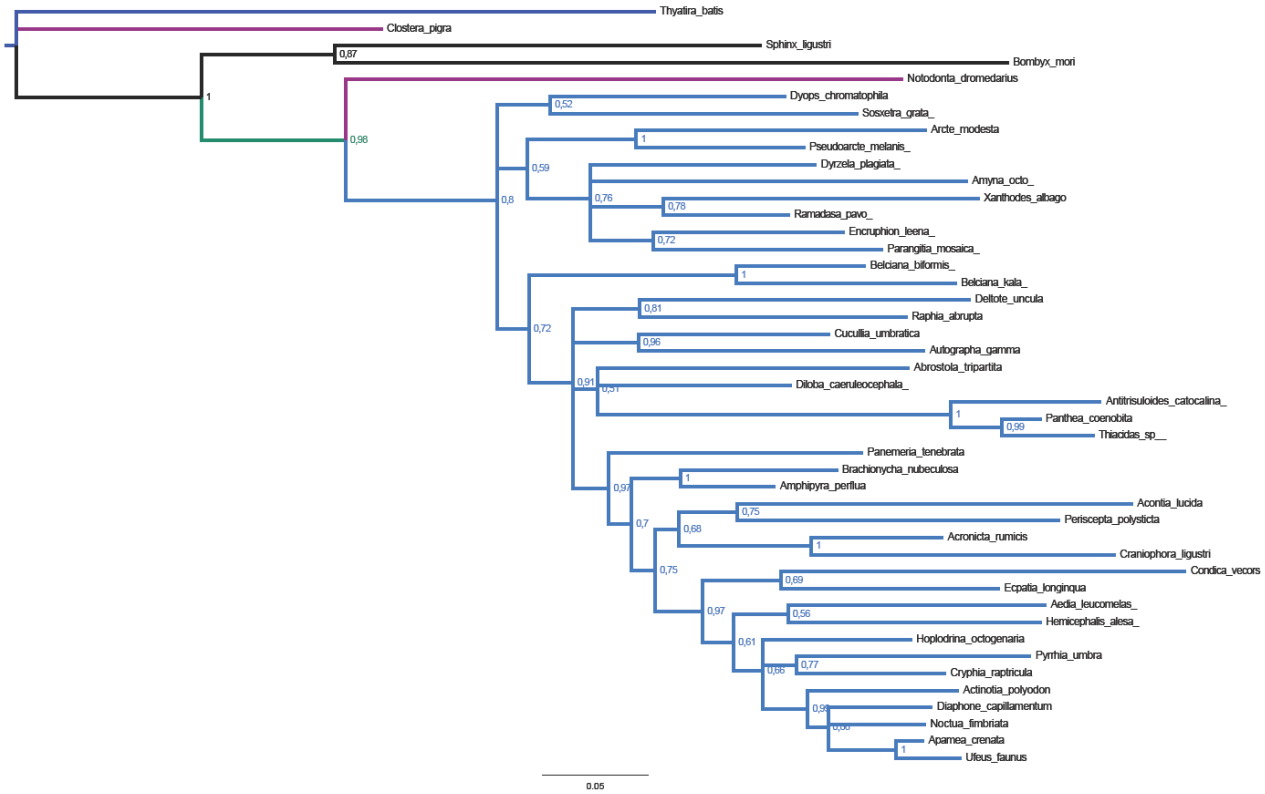


IDH

IDH/31SpodIDHRAxML/31SpodIDHRAxML.tre.pdf

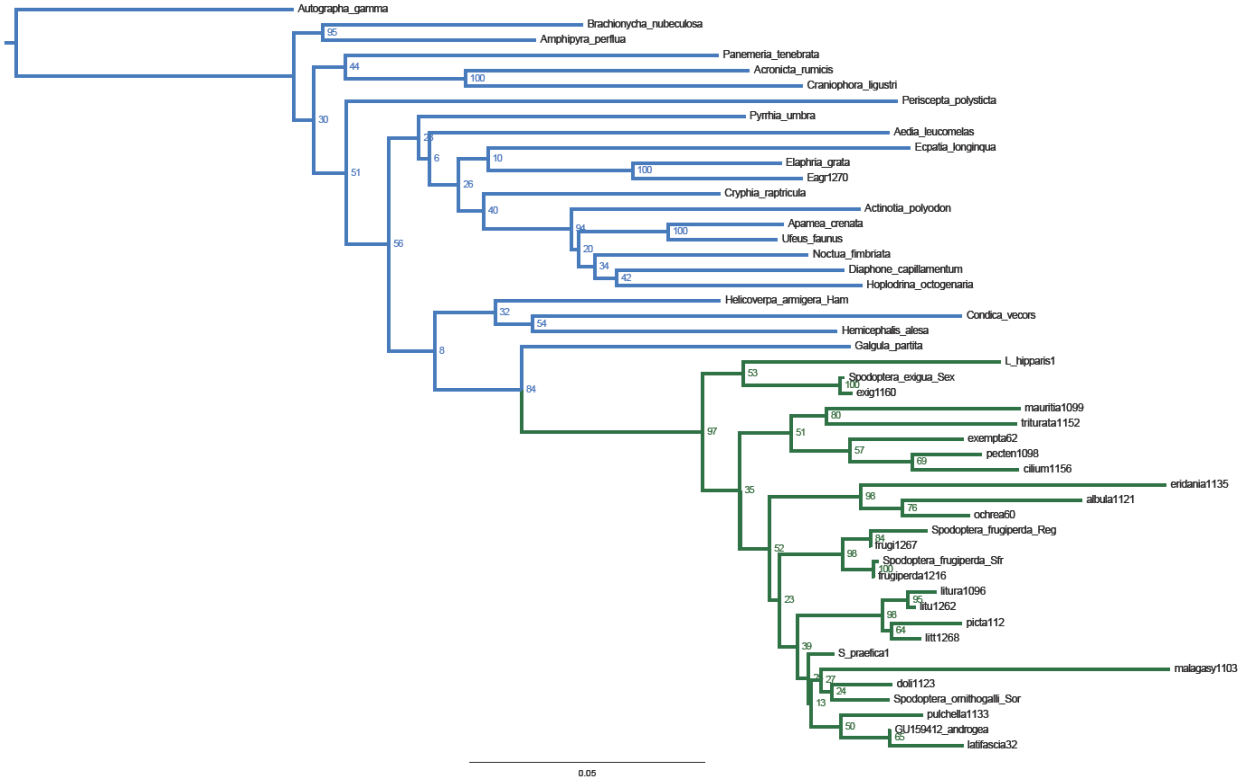


IDH/45SpodIDHBayes100mln/45SpodIDHBayes100mln.nex.con.tre.pdf

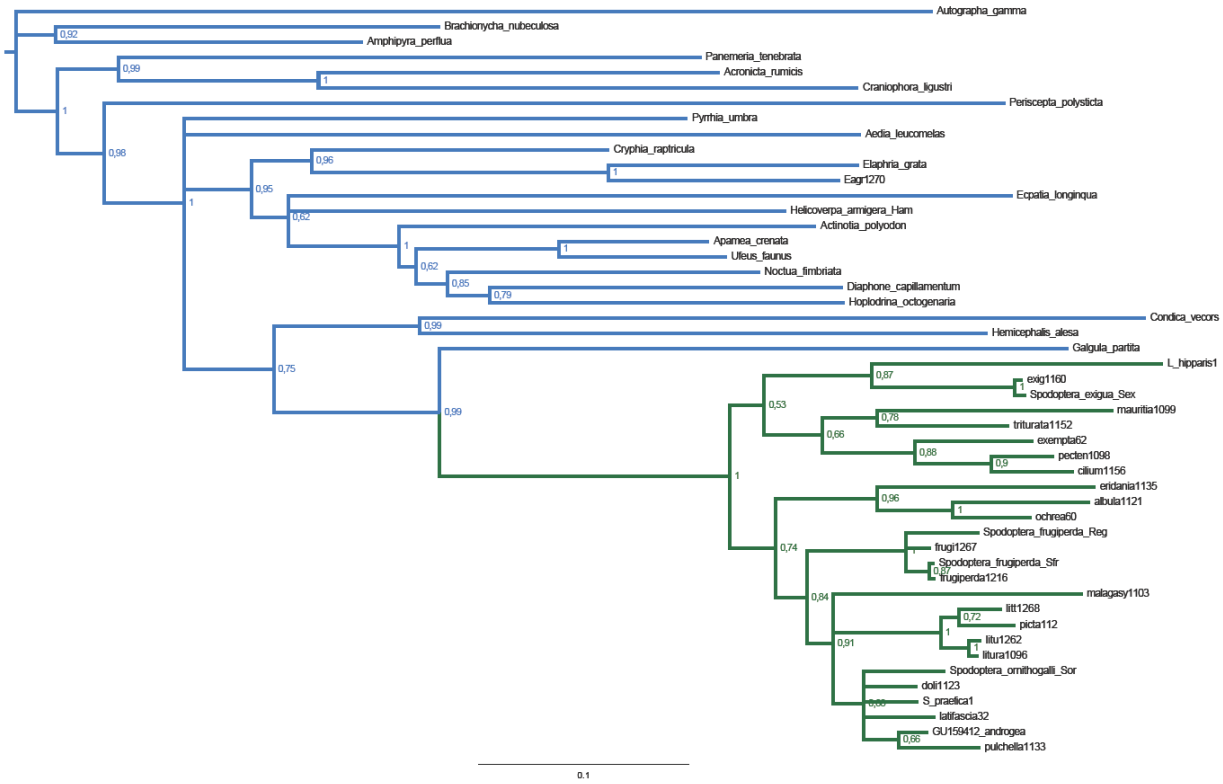


Concatenated alignment 1

Fused/49SpodFusedRAxML/49SpodFusedRAxML.tree.pdf

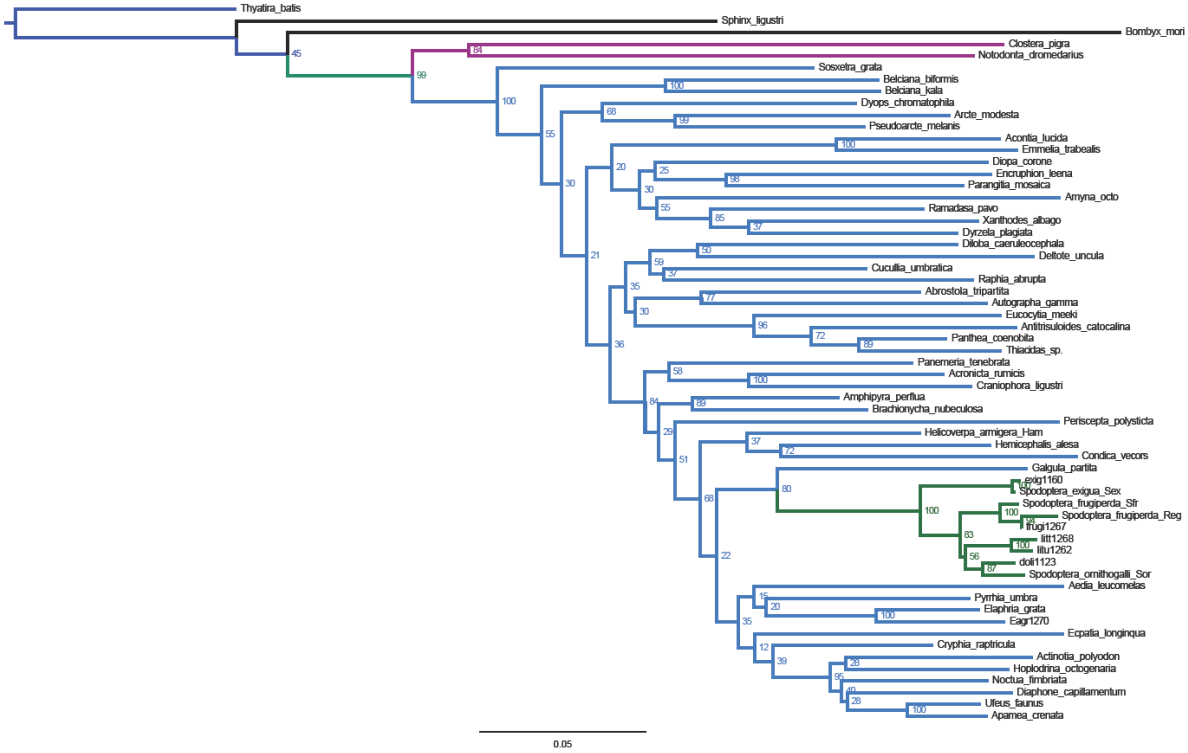


Fused/49SpodFusedBayes197mln/ 49spodfusedbayes197mln.nex.con.tre.pdf

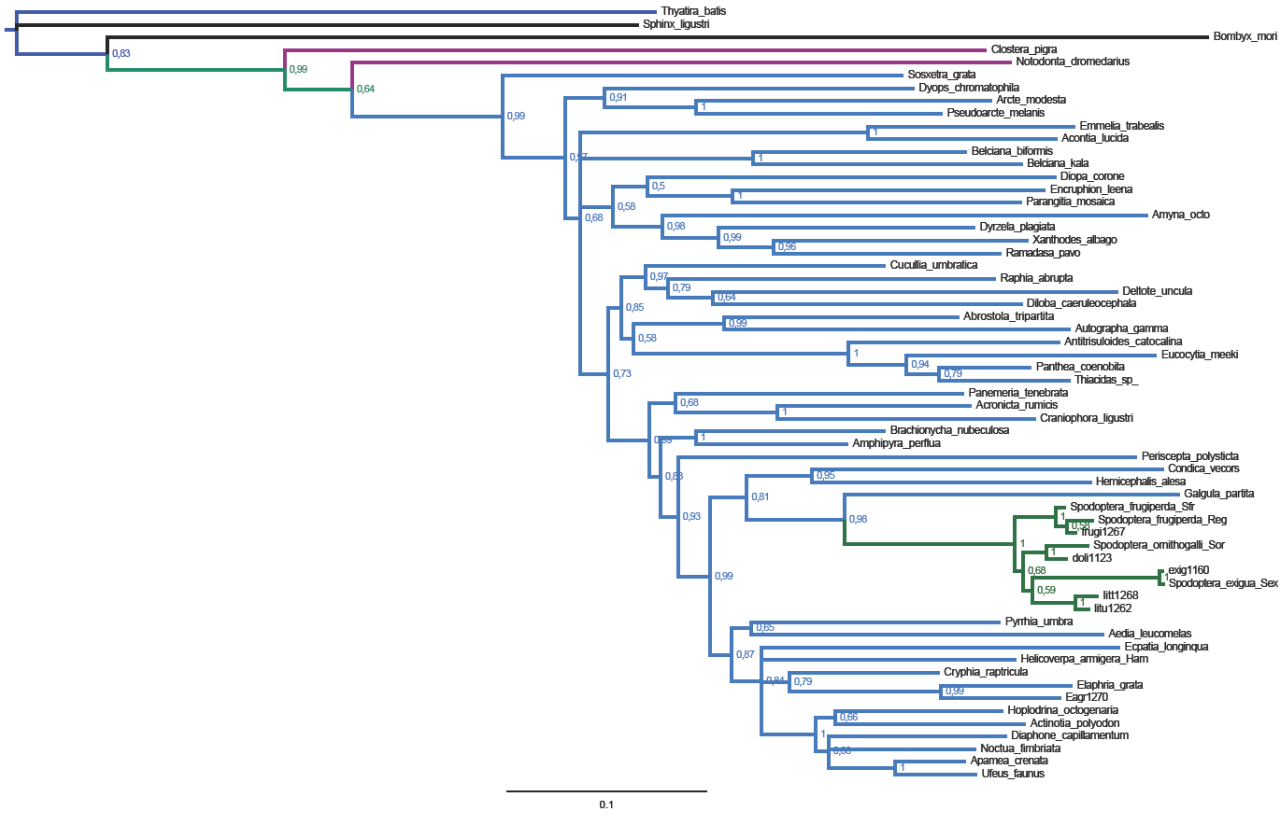


Concatenated Alignment 2

Fused/61SpodFusedRAxML/61SpodFusedRAxML.tre.pdf



Fused/61SpodFusedBayes/61SpodFusedBayes100mlnmetcodonparti.nex.con.tre.pdf



Fused/61SpodFusedBEAST/61SpodFusedBEASTlognormalbottom50mln.(subst).tree.pdf

Appendix 8: Calibration combination results per clade

in 95%-HPD	Calibration:	*Notod*		*Bat*		*Bottom*		*NotodBat*		*NotodBottom*		*BatBottom*		*NotodBatBottom*	
Group	Stem/cro	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Noctuoidea	Stem	24.38	40.76	52.34	68.51	58.98	97.82	50.72	66.67	49.53	85.42	56.87	71.59	55.15	69.99
	Crown	20.96	33.41	41.41	57.35	47.84	82.52	39.46	54.70	39.76	68.97	45.72	60.37	43.80	58.27
Notodontidae	Stem	20.96	33.41	41.41	57.35	47.84	82.52	39.46	54.70	39.76	68.97	45.72	60.37	43.80	58.27
	Crown	20.06	30.32	36.68	53.60	42.37	75.77	34.83	50.71	35.58	63.03	40.76	56.80	37.95	53.73
Noctuidae	Stem	20.96	33.41	41.41	57.35	47.84	82.52	39.46	54.70	39.76	68.97	45.72	60.37	43.80	58.27
	Crown	15.24	26.31	32.10	45.91	37.16	64.53	30.44	43.90	39.76	55.25	35.48	48.48	34.34	47.11
Spodoptera +*Galgula*	Stem	9.36	16.39	18.82	27.24	low pp	low pp	18.05	26.42	low pp	low pp	21.34	29.64	20.61	28.76
	Crown	6.14	13.26	12.45	23.10	14.61	31.36	11.84	22.21	13.23	27.24	13.96	25.10	13.11	23.94
Spodoptera	Stem	6.14	13.26	12.45	23.10	14.61	31.36	11.84	22.21	13.23	27.24	13.96	25.10	13.11	23.94
	Crown	2.43	5.26	4.96	9.34	6.02	12.84	4.78	8.97	5.18	10.91	5.60	10.07	5.43	9.77
frugi+LPL +POLDA	Stem	2.43	5.26	4.96	9.34	6.02	12.84	4.78	8.97	5.18	10.91	5.60	10.07	5.43	9.77
	Crown	1.64	3.44	3.23	6.08	4.06	8.52	3.17	5.85	3.48	7.24	3.72	6.61	3.55	6.35
frugi	Stem	2.43	5.26	4.96	9.34	6.02	12.84	4.78	8.97	5.18	10.91	5.60	10.07	5.43	9.77
	Crown	0.24	1.36	0.43	2.43	0.55	3.28	0.40	2.41	0.43	2.88	0.46	2.78	0.49	2.59
LPL+POLDA	Stem	1.64	3.44	3.23	6.08	4.06	8.52	3.17	5.85	3.48	7.24	3.72	6.61	3.55	6.35
	Crown	1.36	3.01	2.78	5.37	3.36	7.37	2.69	5.18	2.91	6.35	3.04	5.73	3.07	5.67
LPL	Stem	1.36	3.01	2.78	5.37	3.36	7.37	2.69	5.18	2.91	6.35	3.04	5.73	3.07	5.67
	Crown	0.37	1.11	0.70	1.96	0.86	2.69	0.68	1.97	0.72	2.28	0.79	2.16	0.77	2.10
POLDA	Stem	1.36	3.01	2.78	5.37	3.36	7.37	2.69	5.18	2.91	6.35	3.04	5.73	3.07	5.67
	Crown	0.53	1.88	1.03	3.35	1.35	4.59	1.02	3.27	1.16	3.94	1.16	3.67	1.14	3.59

in medians		Calibration:						
Group	Stem/crown	*Notod*	*Bat*	*Bottom*	*NotodBat*	*NotodBotto*	*BatBottom*	*NotodBatBottom*
Noctuoidea	Stem	30.90	60.18	77.60	58.34	65.65	64.22	62.64
	Crown	25.50	49.37	64.30	47.03	53.91	53.22	51.21
Notodontidae	Stem	25.50	49.37	64.30	47.03	53.91	53.22	51.21
	Crown	23.50	45.18	58.83	42.35	48.21	48.63	45.93
Noctuidae	Stem	25.50	49.37	64.30	47.03	53.91	53.22	51.21
	Crown	19.89	38.74	50.50	36.77	42.84	42.15	40.78
Spodoptera +*Galgula*	Stem	11.98	23.22	low pp	22.36	low pp	25.40	24.62
	Crown	9.20	17.64	23.12	17.10	19.81	19.46	18.71
Spodoptera	Stem	9.20	17.64	23.12	17.10	19.81	19.46	18.71
	Crown	3.69	7.11	9.23	6.83	7.88	7.71	7.46
frugi+LPL+POLDA	Stem	3.69	7.11	9.23	6.83	7.88	7.71	7.46
	Crown	2.40	4.59	6.05	4.47	5.16	5.07	4.89
frugi	Stem	2.40	4.59	6.05	4.47	5.16	5.07	4.89
	Crown	0.63	1.20	1.59	1.18	1.36	1.34	1.26
LPL+POLDA	Stem	2.40	4.59	6.05	4.47	5.16	5.07	4.89
	Crown	2.09	3.97	5.21	3.86	4.46	4.37	4.23
LPL	Stem	2.09	3.97	5.21	3.86	4.46	4.37	4.23
	Crown	0.68	1.28	1.67	1.24	1.43	1.39	1.36
POLDA	Stem	2.09	3.97	5.21	3.86	4.46	4.37	4.23
	Crown	0.63	2.09	2.76	2.07	2.36	2.30	2.24