



Karlsruhe Institute of Technology

# The Anguillicolidae

## (Nematoda: Anguillicoloidea)

### Barcode, population structures and phylogenetics

Diploma thesis of

**Dominik R. Laetsch**

At the Faculty of Chemistry and Biosciences  
Institute of Zoology I  
Department of Ecology & Parasitology

Reviewer: Prof. Horst Taraschewski  
Second reviewer: Prof. Peter Nick  
Advisor: Prof. Mark L. Blaxter  
Second advisor: Prof. Horst Taraschewski

17<sup>th</sup> of December 2009 – 2<sup>nd</sup> of June 2010



*E quindi uscimmo a riveder le stelle.*

- Dante Aligheri, La Comedia (1307-1320), Canto XXXIV



## Declaration of Authorship

I certify that the work presented here is, to the best of my knowledge and belief, original and the result of my own investigations, except as acknowledged, and has not been submitted, either in part or whole, for a degree at this or any other University.

---

Location, Date

---

Dominik R. Laetsch



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Nematoda . . . . .	1
1.2	Parasitism and Nematodes . . . . .	3
1.3	The Anguillicolidae . . . . .	6
1.3.1	The <i>invasiveness</i> of <i>Anguillicoloides crassus</i> . . . . .	8
1.3.2	The host, its economic importance and the pathogenicity of <i>Anguillicoloides crassus</i> . . . . .	9
1.3.3	Taxonomic history of the Anguillicolidae and results from phylogenetic analyses of clade 8 nematodes <i>sensu</i> Holterman et al. (2006) . . . . .	12
1.4	Implications of the phylogenetic position of <i>A. crassus</i> on the origin of parasitism within the Nematoda . . . . .	13
1.5	Barcode genes and phylogenetic analysis . . . . .	14
<b>2</b>	<b>Aims</b>	<b>17</b>
<b>3</b>	<b>Materials and Methods</b>	<b>19</b>
3.1	Sampling . . . . .	19
3.2	Tissue extraction and lysis . . . . .	20
3.3	DNA amplification . . . . .	20
3.4	TOPO cloning . . . . .	21
3.5	Gel electrophoresis . . . . .	22
3.6	Gel extraction . . . . .	22
3.7	Sequencing . . . . .	22
3.8	Bioinformatic methods . . . . .	23
3.8.1	Processing of raw sequence data . . . . .	23
3.8.2	Screening for contaminants . . . . .	24
3.8.3	Assembly of consensus sequences . . . . .	24
3.8.4	Construction of consensus sequence alignments . . . . .	26
3.8.5	Inclusion of additional anguillicoloid taxa in the datasets . . . . .	27
3.9	Phylogenetic Analysis . . . . .	27
3.9.1	Search for outgroups . . . . .	27
3.9.2	Assessment of phylogenetic utility of the alignments . . . . .	29

3.9.3	Model selection for DNA evolution . . . . .	30
3.9.4	Indel Coding . . . . .	30
3.9.5	Phylogenetic analysis using Bayesian Inference, as implemented in MrBayes	31
3.9.6	Assessment of convergence of Markov chains . . . . .	32
3.9.7	Comparison of the models proposed by MrModeltest using MrBayes output	33
3.9.8	Tests of taxon monophyly using MrBayes . . . . .	33
3.9.9	Phylogenetic analysis of the different datasets using MrBayes . . . . .	34
3.9.10	Phylogenetic analysis using the maximum likelihood method as implemented in RAxML . . . . .	36
3.9.11	Interpretation of branch support . . . . .	36
3.9.12	Drawing and labelling of phylogenetic trees . . . . .	36
3.10	Analyses of population structure . . . . .	37
3.10.1	Statistical parsimony haplotype networks . . . . .	37
3.10.2	Calculation of simple population genetic parameters using Arlequin . . . . .	40
<b>4</b>	<b>Results and Discussion</b>	<b>41</b>
4.1	Laboratory results . . . . .	41
4.2	Analysis of the 18S ribosomal DNA (18S rDNA) dataset . . . . .	42
4.3	Analysis of the internal transcribed spacer I (ITS1) dataset . . . . .	44
4.4	Analysis of the 28S ribosomal DNA D2-D3 (28S rDNA D2-D3) dataset . . . . .	45
4.5	Analysis of the internal transcribed spacer 2 (ITS2) dataset . . . . .	48
4.6	Analyses of the Cytochrome oxidase I ( <i>Cox1</i> ) dataset . . . . .	48
4.6.1	Analysis of the <i>Cox1</i> sequences from <i>A. australiensis</i> . . . . .	49
4.6.2	Analysis of the <i>Cox1</i> sequences from <i>A. novaezelandiae</i> . . . . .	49
4.6.3	Analysis of the <i>Cox1</i> sequences from <i>A. papernai</i> . . . . .	51
4.6.4	Analysis of the <i>Cox1</i> sequences from <i>A. crassus</i> . . . . .	53
4.7	Assessment of the phylogenetic utility of the different barcoding genes . . . . .	56
4.8	Phylogenetic analyses . . . . .	57
4.8.1	Phylogenetic analyses of the 18S rDNA sequences . . . . .	57
4.8.2	Phylogenetic analyses of the 28S rDNA D2-D3 sequences . . . . .	62
4.8.3	Phylogenetic analyses of the <i>Cox1</i> sequences . . . . .	63
4.8.4	Phylogenetic analysis of the “supermatrix” . . . . .	67
4.8.5	Summary of the phylogenetic analyses . . . . .	68
4.9	A co-evolution theory for the host-parasite-system of <i>Anguilla</i> infected with anguillicolid nematodes . . . . .	68
<b>5</b>	<b>Conclusions</b>	<b>73</b>
<b>6</b>	<b>Suggestions for further work</b>	<b>79</b>

<b>A Appendix</b>	<b>81</b>
A.1 Perl programs . . . . .	81
A.2 DNA Sequences and alignments . . . . .	81
A.3 Statistical parsimony networks . . . . .	81
A.4 Arlequin analyses . . . . .	81
A.5 Phylogenetic analysis . . . . .	81
A.5.1 18S rDNA dataset . . . . .	82
A.5.2 28S rDNA D2-D3 dataset . . . . .	83
A.5.3 <i>Cox1</i> dataset . . . . .	85
<b>B List of Figures</b>	<b>87</b>
<b>C List of Tables</b>	<b>93</b>
<b>D Bibliography</b>	<b>95</b>
<b>E Colophon</b>	<b>111</b>



# 1. Introduction

## 1.1. Nematoda

The Nematoda is a phylum of abundant ecdysozoan animals (Dunn et al., 2008), comprising over 27,000 described species (Hugot et al., 2001) distributed among most of the available habitats on Earth. The representatives of this phylum are generally of microscopic size, although the female giant nematode *Dioctophyme renale* may reach a length of up to 1 metre.

Their numerical dominance, often exceeding more than 1 million individuals per square metre and accounting for about 80% of all individual animals on earth (Platt, 1994), their diversity in lifestyles, and their presence at various trophic levels, suggest an important role in many ecosystems (Holterman et al., 2006).

While their true diversity is estimated to range from 100,000 to 10 million species (Lambshead et al., 2000; Lambshead and Boucher, 2003), which remains to be verified, their ubiquity is unquestionable; reaching from the depths of the deep sea to the soil and sediments covering the crests of mountains.

The Nematoda are therefore one of the most successful animal groups (Dougherty, 1951). One attempt to illustrate the sheer species richness of this phylum was undertaken by a famous American nematologist:

“[...] if all the matter in the universe except the nematodes were swept away, our world would still be dimly recognisable, and if, as disembodied spirits, we could then investigate it, we should find its mountains, hills, vales, rivers, lakes, and oceans represented by a film of nematodes. The location of towns would be decipherable, since for every massing of human beings there would be a corresponding massing of certain nematodes. Trees would still stand in ghostly rows representing our streets and highways. The location of the various plants and animals would still be decipherable, and, had we sufficient knowledge, in many cases even their species could be determined by an examination of their erstwhile nematode parasites. ”

- N. A. Cobb (1914)

Chitwood (1933, 1937) and Chitwood and Chitwood (1950) divided the Nematoda into the Aphasmidia and Phasmidia, later renamed Adenophora (“gland bearers”) and the Secernentea (“secretors”) respectively (Chitwood, 1958). The division was based on the fact that the Secernentea share several characteristics including the presence of phasmids, a pair of

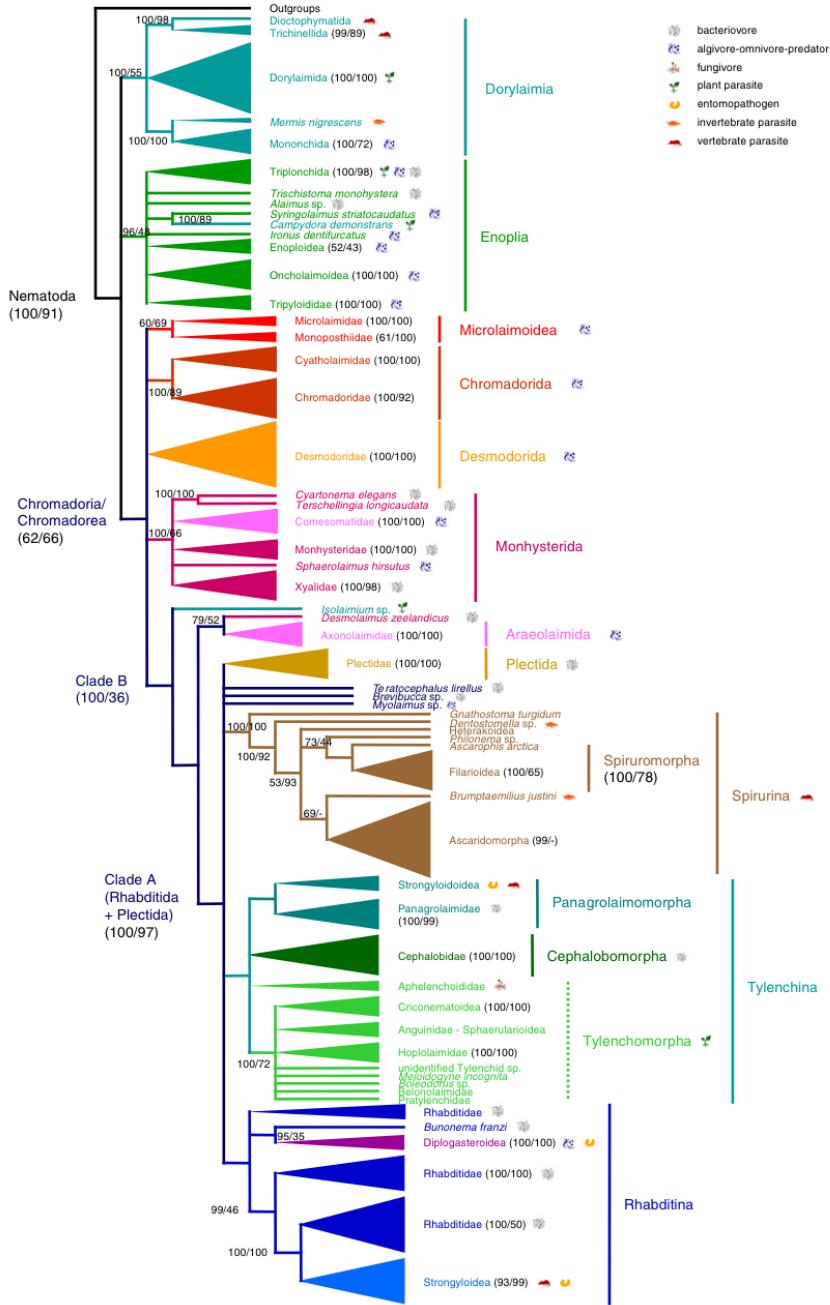
sensory organs located in the lateral posterior region. Adenophora include virtually all aquatic nematodes (Enoplida and Chromadorida) and selected terrestrial omnivores or plant-feeders (Dorylaimia), while almost all parasitic species (Strongylina, Tylenchina, Ascaridina, and Spirurida) belong to the Secernentea. Andrassy (1976) proposed a tripartite system, dividing the Adenophorean further into the Torquentia (roughly equivalent to the Chromadorida) and the Penetrantia (roughly equivalent to the Enoplida).

Lorenzen (1981) introduced the first taxonomic system based on cladistic principles, showing that there was no support for the Adenophorean as a monophyletic group. Furthermore, he showed that the number of morphological characters was too low to come up with a plausible alternative.

Advances in molecular biology techniques allowed an objective, empirical analysis of the evolutionary history of the Nematoda. Blaxter et al. (1998) confirmed the paraphyly of the Adenophora, by inferring the first phylogenetic framework of the phylum using 53 small subunit rDNA (18S rDNA) sequences from 53 taxa (covering all the major parasitic and free-living taxonomic groups). Additionally Blaxter et al. (1998) defined 5 major clades and showed that trophic ecologies such as animal and plant parasitism arose several times independently. Comparable results were presented by Aleshin et al. (1998) based on 19 nematode sequences.

Later, De Ley and Blaxter (2002, 2004) updated the classification of the phylum using molecular data available from additional species, with morphological data to assist the placement of taxa for which SSU sequences were not yet available. Holterman et al. (2006) analysed the nearly full-length small-subunit rDNA sequences from 339 taxa and suggested a further division into 12 clades. Interestingly, significantly accelerated nucleotide substitution rates were observed in virtually all parasitic taxa as well as taxa displaying short generation times. Furthermore fungal feeding nematodes were exclusively found basal to or as sister taxa next to plant parasitic nematodes, supporting the long standing hypothesis that plant parasitic nematodes may have arisen from fungivorous ancestors. Meldal et al. (2007) inferred a phylogeny using additional, previously unsequenced, species with special emphasis on marine taxa which is depicted in Figure 1.1.

The most extensive analysis up to date has been carried out by Van Megen et al. (2009) who analysed 1200 small subunit rDNA sequences, but found no major changes in the subdivision of this phylum compared to Holterman et al. (2006). However, extensive convergent evolution within the phylum Nematoda was observed, since most morphological, ecological and biological characteristics have arisen at least twice during nematode evolution.



**Figure 1.1.:** An overview of the Nematoda based on SSU phylogenetic analyses. Support values (in percentage) from Bayesian inference and LogDet transformation of distances are given for putative monophyletic groups. Trophic groups are shown as pictograms. Major taxa according to De Ley and Blaxter (2004) are highlighted in the same colour (from Meldal et al., 2007)

## 1.2. Parasitism and Nematodes

Since every biological system can be referred to as an open or continuous system, the number of interactions with among the systems are vast and vary in extent, duration and, of course, benefit for each of the systems involved.

In 1879 the German mycologist Heinrich Anton de Bary coined the term *symbiosis*, defining it as “the living together of unlike organisms”, and explicitly treating parasitism as a type of symbiosis (Douglas, 1994).

In order to classify the broad spectrum of interactions between organisms further, the term *mutualism* was introduced, characterising bidirectional essential interactions, distinguishing them from parasitic interactions, in which one organism (the parasite) benefits, while causing a detrimental effect to the other (the host).

As with most concepts in the biological sciences, the phenomenon of life provides us with examples for all the nuances within this spectrum.

It is difficult to draw a clear line between parasitism and classical predator-prey-relationships. Elton (1927) suggested that the resemblance between predators and parasites are more important than their differences: ectoparasitic plant-feeding nematodes, for example, can be viewed as herbivores, much smaller than an ungulate, but nevertheless predating on plants. Following this thought, what we define as parasitism may just be the small organism’s way of predating on life forms many fold bigger than themselves.

The most recent definition, or perhaps description, of parasitism is given by Crofton (1971) and states that a parasite-host-system must fulfil the following five criteria:

1. Ecological relationship between one organism (the parasite) and another (the host).
2. The parasite is physiologically or metabolically dependent upon its host.
3. Heavily infected hosts will be killed by their parasites.
4. The reproductive potential of the parasite exceeds that of their hosts.
5. There is an over-dispersed frequency distribution of parasites within the host population.

Especially the last criterion is of great importance, since an over-dispersed frequency distribution is generally not found in micro-predators, therefore excluding them from the definition.

Estimates of the omnipresence of the parasitic life-style suggest that ~40 – 50 % of all known species are parasitic (Price, 1980; Windsor, 1998; Dobson et al., 2008), making parasitism the most common life-style among eukaryotes. However, a prediction concerning the overall diversity of parasitic nematodes is complicated by the anthropocentric bias towards the study of medically and economically relevant nematodes. While the rate of discovery of new parasitic nematodes has grown linearly or even exponentially, the overall sampling depth of nematodes remains thin at best (Dobson et al., 2008). Nevertheless, parasitic nematodes, like parasites *per se*, seem to have a vast impact on ecosystems (Lafferty et al., 2006) and hence are of extreme agricultural, economic and medical importance.

Whatever definition one considers best, parasitism by nematodes can be divided into two

groups: endo- and ectoparasitism. Ectoparasitism, the exploitation of a host's metabolic resources from *outside* the host, may involve several different host individuals during the life-cycle of the nematode, while an endoparasitic nematode, feeding *inside* the host, is generally associated with only one individual in the host population per life-stage. This, of course, leads to distinct implications for each kind of parasite regarding, for instance, the evasion of host-immune-responses, mating strategies and the general depth of host-parasite-interactions ranging from generalist to specialist.

The impact of parasitic nematodes on human society is best described by the amount of effort that is undertaken to prevent it. Worldwide around US\$ 20 billion are spent by farmers on crop protection against cosmopolitan plant-parasitic nematodes (Abawi and Widmer, 2000). Nevertheless as much as US\$ 100 billion in annual losses of crops are caused across the planet (Koenning et al., 1999), contributing significantly to malnutrition and associated diseases. The greatest amount of crop damage is caused by the tylenchid nematodes, comprising around 60 root-knot nematodes species of the genus *Meloidogyne*. Found in areas with warm and hot climates, the reason for their success may reside in their ability to reproduce in a vast range of host plant species: in the case of the most prominent representative, *Meloidogyne incognita*, the possible number of different host plants exceeds 2,000 species. On the other hand, the cyst nematodes *Heterodera glycines* and *Globodera* spp. show a narrower host range, restricted to one or a few related plant species. They are significant pests on soybeans and potatoes, respectively. Certain nematodes, such as *Xiphinema americanum* are vectors for nepoviruses (Griesbach and Maggenti, 1989; McGuire, 1973).

Among the group of animal-parasitic nematodes, the most important are, from an anthropocentric point of view, the medically relevant species. Over 1.2 billion people worldwide are infected with the large gut worm *Ascaris lumbricoides*, causing malnutrition and obstructive bowel disease. The strongylid human hookworms (*Ancylostoma duodenale* and *Necator americanus*), as well as the pork trichina worm *Trichinella spiralis* account for ~700 – 800 million infections each (De Silva et al., 2003). It is estimated that the annual deaths caused by infections due to soil-transmitted nematodes (*A. duodenale*, *N. americanus*, *A. lumbricoides*, *Enterobius vermicularis*, *Strongyloides stercoralis* and *Trichuris trichiura*) alone vary between 12,000 (WHO World Health Report, 2002) to 135,000 (WHO World Health Report, 2004).

In Table 1.1, a list of the most important parasitic nematodes threatening human health is depicted.

Zooparasitic nematodes are also known to parasitise livestock, causing reduced production of meat and milk in cattle in the case of *Ostertagia ostertagi*, *Dictyocaulus viviparus*, *Cooperia* spp. and *Trichostrongylus* spp. (Vercruyse and Claerebout, 2001). Other nematodes parasitise fish as larvae or adults, attacking most body organs of the host. As agents of serious fish

Nematode species	Infect. ( $10^6$ )	Distribution
<i>Ancylostoma duodenale</i> and <i>Necator americanus</i>	1298	Worldwide (developing countries)
<i>Ascaris lumbricoides</i>	1472	Worldwide (developing countries)
<i>Brugia malayi</i>	13	East Indonesian islands, Philippines, South-East Asia, Southern China and India
<i>Dracunculus medinensis</i>	0.08	Sub-Saharan Africa, Yemen
<i>Enterobius vermicularis</i>	200	Worldwide
<i>Loa loa</i>	13	West and Central Sub-Saharan Africa
<i>Onchocerca volvulus</i>	18	Central and South America, Sub-Saharan Africa
<i>Strongyloides stercoralis</i>	70	Worldwide (warm countries)
<i>Trichinella spiralis</i>	800	Worldwide
<i>Trichuris trichiura</i>	1049	Worldwide (developing countries)
<i>Wuchereria bancrofti</i>	107	Asia, Central and South America, Sub-Saharan Africa, West Pacific countries

**Table 1.1.:** Geographical distribution of the major medically relevant nematode species and the estimated number of infections (in millions) caused in Humans. (after Crompton, 1999; De Silva et al., 2003; Lukeš et al., 2005)

diseases they produce an impact on the aquaculture industry; e.g. *Hysterotylaciun aduncum* in cultured salmon (*Salmo salar*) (Carvajal et al., 1995) and *Camallanus cotti* in guppies (*Poecilia reticulata*) (Rigby et al., 1997). Due to the consumption of undercooked host-tissue, humans are prone to infections of *Anisakis* spp., *Pseudoterranova* spp., *Gnathostoma* spp. and *Capillaria philippinensis* causing important public health problems in some regions (Suzuki et al., 2010; Torres et al., 2007; Ando et al., 1992; Lu et al., 2006).

### 1.3. The Anguillicolidae

Among the ichthyoparasitic nematodes, the family Anguillicolidae, parasites of fresh-water eels of the genus *Anguilla*, has attracted the attention of many researchers in the past decades (e.g. Nagasawa et al., 1994; Taraschewski, 2006; Kennedy, 2007). The Anguillicolidae *sensu* Moravec (2006) comprise five species, grouped into two genera as depicted in Figure 1.2.

Like nematodes in general, members of the Anguillicolidae moult four times during their life and are suspected to have similar life cycles (Moravec and Taraschewski, 1988; Kirk, 2003).

**Superfamily** Anguillicoloidea (Yamaguti, 1935)**Family** Anguillicolidae (Yamaguti, 1935)**Genus** *Anguillicola* (Yamaguti, 1935)*A. globiceps* (Yamaguti, 1935)**Genus** *Anguillicoloides* (Moravec et Taraschewski, 1988)*A. australiensis* (Johnston et Mawson, 1940; Moravec et Taraschewski, 1988)*A. crassus* (Kuwahara, Niimi et Itagaki, 1974; Moravec et Taraschewski, 1988)*A. novaezelandiae* (Moravec et Taraschewski, 1988)*A. papernai* (Moravec et Taraschewski, 1988)**Figure 1.2.: Systematic classification of the Anguillicolidae *sensu* Moravec (2006).**

For the sake of simplicity, only the life cycle of *A. crassus*, the best studied species of the Anguillicolidae, is illustrated here.

De Charleroy et al. (1990) provided the first complete characterisation of the life cycle of *A. crassus*. The dioecious adults parasitise in the swim bladder of the eel host feeding off blood-vessels. They do not exit the host, but decay *post mortem*. After copulation, females constantly release eggs, which hatch to L2 larvae in the swim bladder lumen. These L2 larvae exit the host via the *ductus pneumaticus*, which connects swim bladder and intestinal tract. Upon release, undulating body movements are performed to attract zooplanktonic predators, which serve as intermediate hosts. Several copepod and ostracod species have been identified as suitable intermediate hosts (Thomas, 1993). Inside the intermediate host, the L2 moults into infectious L3 larvae. At this stage, many freshwater fish species, tadpoles and larvae of aquatic insects can serve as paratenic hosts, in which moulting into the L4 larvae occurs (Reimer et al., 1994; Moravec and Škoríková, 1988). After being ingested by the final host, the larvae penetrate the mucosa and enter the connective tissue of the swim bladder wall where the last moult takes place. Finally, the adults move into the swim bladder lumen where sexual dimorphism develops.

The only species within the genus *Anguillicola*, *A. globiceps*, has been found to be endemic only in certain prefectures of Japan and provinces in China (Yamaguti, 1933; Moravec and Taraschewski, 1988; Nagasawa et al., 1994), where it infects the Japanese eel *Anguilla japonica* Temminck et Schlegel 1847.

Members of the genus *Anguillicoloides* are widely distributed. *A. australiensis* is found in South and East Australia (Johnston and Mawson, 1940; Moravec and Taraschewski, 1988; Kennedy, 1994) and infects the long-finned eel *Anguilla reinhardtii* Steindachner 1867.

*A. crassus* has today the widest geographical range (see Section 1.3.1), but was originally only known from East Asian countries, such as Taiwan, Japan, China and Korea, where it parasitises native *A. japonica* as well as the introduced European eel *Anguilla anguilla* Linnaeus

1758. The widest host range is also displayed by this species (see section 1.3.2), including, in addition to two eel species mentioned before, the American eel *Anguilla rostrata* LeSueur, 1821, the Indonesian short-fin eel *Anguilla bicolor* McClelland, 1844, the giant mottled eel *Anguilla marmorata* Quoy et Gaimard, 1824 and the African long-fin eel *Anguilla mossambica* Peters, 1852 (Moravec, 2006).

*A. novaezelandiae* was first described in *Anguilla australis* Richardson, 1841 from New Zealand (Moravec and Taraschewski, 1988), although, as early as 1982, this species, reported erroneously as *A. australiensis*, was found in *A. anguilla* from Lake Bracchiano in Italy (Paggi et al., 1982). This finding can be explained by the repeated introduction of *A. australis* from New Zealand into Lake Bracchiano. However, recent investigations show no evidence for the occurrence of *A. novaezelandiae* in this locality and it is assumed that it gradually has been replaced by *A. crassus* (Moravec et al., 1994). Moravec and Rohde (1992) also reported *A. novaezelandiae* from South-East Australia.

The species which least is known about is *A. papernai*, endemic in South Africa and Madagascar, where it infects *A. mossambica*. However, Taraschewski et al. (2005) demonstrated through laboratory infections that *A. papernai* is capable of completing its life cycle with European species of intermediate and final host, and suggested that, if introduced to Europe or North America, it could spread through the eel populations on these continents.

Three main reasons exists for the interest in this family of parasitic nematodes.

- The growing geographical range, or the invasiveness, of one representative of this family.  
(Section 1.3.1)
- The economic importance of the host and the pathogenicity associated with infection.  
(Section 1.3.2)
- The taxonomic and phylogenetic position(s) of the Anguillicolidae and their implications.  
(Section 1.3.3)

### **1.3.1. The invasiveness of *Anguillicoloides crassus***

Until recently no member of this family occurred either in European eels (*Anguilla anguilla*) or in North American eels (*Anguilla rostrata*).

Paggi et al. (1982) reported the first sighting in Europe of an anguillicolid nematode, later identified as *Anguillicoloides novaezelandiae* (Moravec and Taraschewski, 1988), in *Anguilla anguilla* in Lake Bracchiano in Italy.

Neumann (1985) reported swim bladder nematodes in *Anguilla anguilla* in the Weser-Ems River region in Northern Germany. These were identified as *Anguillicoloides crassus* and were probably introduced from Taiwan (Køie, 1991; Wielgoss et al., 2008) through the live eel trade. Koops and Hartmann (1989) identified one transfer of 35 tonnes of live Taiwanese eels

in 1980 as the most likely source of the spread of *A. crassus*. Subsequently this nematode was reported in many European countries (Kirk, 2003; Rolbiecki and Rokicki, 2005), in Turkey (Genç et al., 2008), North Africa in Egypt (Koops and Hartmann, 1989), Morocco (El Hilali et al., 1996) and Tunisia (Maamouri et al., 1999) reaching prevalence rates of more than 70% in most European countries (Kirk, 2003).

In 1994, *A. crassus* was also introduced into North America, infecting the American eel *A. rostrata* both in aquaculture and in the wild (Fries et al., 1996; Barse and Secor, 1999). Analysis of microsatellite loci and mitochondrial markers suggest that this colonisation process originated from a Japanese population of *A. crassus* Wielgoss et al. (2008).

Sasal et al. (2008) reported the occurrence of *A. crassus* in native and introduced eels on the Island of Reunion in the Indian Ocean. It is assumed that the flexibility of the life cycle with respect to intermediate- and paratenic hosts is one of the main factors responsible for the increase in prevalence and distribution of *A. crassus* throughout the world (Kirk, 2003).

This species is thus considered a global invader *sensu* Colautti and MacIsaac (2004) and Taraschewski (2006).

### **1.3.2. The host, its economic importance and the pathogenicity of *Anguillicoloides crassus***

Fresh-water eels of the genus *Anguilla* comprise 15 species and three subspecies (Ege, 1981; Lecomte-Finiger, 2003). These are widely distributed among most tropical and subtropical areas of the world. The distribution ranges for the individual species are listed in Table 1.2 together with their respective anguillicolid parasites.

As indicated in Table 1.2, anguillicolid nematodes have only been observed in certain eel species, namely *A. anguilla*, *A. australis*, *A. bicolor*, *A. japonica*, *A. marmorata*, *A. mossambica*, *A. reinhardtii* and *A. rostrata*. Unfortunately, the European eel *Anguilla anguilla*, the Japanese eel *Anguilla japonica* and the American eel *Anguilla rostrata* are considered to be the eel species with the greatest commercial importance for the food market.

All species of the genus *Anguilla* display catadromous life-cycles, comprising two long-range, marine migrations, a continental growth phase and a marine pelagic spawning stage in localised, tropical areas, subject to highly saline, warm ( $> 20^{\circ}\text{C}$ ) subtropical water currents (Tsukamoto et al., 2002). Both, spawning ground circumference and location vary between species. For instance spawning Japanese eels only occupy a narrow area close to the Mariana's Trench in the Pacific Ocean (Tsukamoto, 2006), while the North Atlantic eel species *A. anguilla* and *A. rostrata* apparently spawn over a large, broadly overlapping area in the Sargasso Sea (Schmidt, 1923; Schoth and Tesch, 1982; Kleckner and McCleave, 1988).

In the case of the European eel, the transparent and willow-leaf-shaped larva, referred to as

## 1. Introduction

---

Eel species	Distribution	Anguillicolid parasites
<i>A. anguilla</i>	Europe, Mediterranean Sea, Northern Africa	C (+), N (-), P (-) (Moravec, 2006)
<i>A. australis</i>	East Australia, New Zealand	N (-) (Moravec, 2006)
<i>A. bicolor</i>	Madagascar, Indochina, Philippines, South East Asia	C (-) (Sasal et al., 2008)
<i>A. borneensis</i>	Indonesia	None recorded
<i>A. celebesensis</i>	Indonesia	None recorded
<i>A. dieffenbachii</i>	New Zealand	None recorded
<i>A. interioris</i>	New Guinea	None recorded
<i>A. japonica</i>	Japan, South East Asia	C (-), G (-) (Moravec, 2006)
<i>A. marmorata</i>	South East Africa, Madagascar, Indo-Pacific	C (-) (Sasal et al., 2008)
<i>A. megastoma</i>	Pacific Ocean	None recorded
<i>A. mossambica</i>	South East Africa, Madagascar	P (-) (Moravec, 2006), C (-) (Sasal et al., 2008)
<i>A. nebulosa</i>	South East Africa, Indochina, South East Asia	None recorded
<i>A. obscura</i>	Pacific Ocean, Fiji Islands	None recorded
<i>A. reinhardtii</i>	East Australia, Tasmania, New Zealand	A (-) (Moravec, 2006)
<i>A. rostrata</i>	East Coast of the United States	C (+) (Moravec, 2006)

**Table 1.2.:** Geographic distribution and qualitative anguillicolid parasite burden of the 15 species of the genus *Anguilla* (After Minegishi et al., 2004). Anguillicolid parasites: A = *A. australiensis*, C = *A. crassus*, G = *A. globiceps*, N = *A. novaezelandiae*, P = *A. papernai*; (+) = high abundance, (-) = low abundance.

a leptocephalus, displays a planktovorous diet and is believed to drift along the North Atlantic currents towards continental water systems over a period of two years (Kettle and Haines, 2006). However, the analysis of otoliths suggests an even shorter migration time (Lecomte-Finiger, 1994; Arai et al., 2000), implying that active swimming occurs. Upon reaching the continental shelf, the first metamorphosis takes place and the leptocephali transform into transparent anguilliform stages, referred to as glass eels. Although eels may stay their whole life in the open ocean (Tsukamoto and Arai, 2001), most glass eels gather in the tidal zone of estuaries and prepare for their ascent into freshwater systems by alternating positive and negative rheotaxis with the turn of tides (Bolliet et al., 2007). After further growth and the onset of pigmentation they are referred to as elvers and start a predominantly benthic life-style in freshwater, preying on crustaceans and insects (Tesch, 1977). When reaching lengths around 20 centimetres, they are referred to as yellow eels, due to their dorso-lateral yellow-green or olive coloration (Tesch, 2002). After several years of continental residency, where they show a markedly piscivorous diet, the second metamorphosis takes place, marking the onset of sexual maturation. The gonads become fully developed, the eye diameters and visual capabilities increase, and dorso-lateral blackening starts contrasting with a ventro-lateral silverying. This stage is named the silver eel (Tesch, 2002). In order to complete the migration into the spawning area in marine waters, the epidermal mucosa develops permitting a readaption

to raising salinities. Although there is evidence that maturing migrants can hold a compass course during movement from the estuaries toward the open sea (Tesch, 2002), no single eel spawner has ever been caught in the oceans and the exact migration routes, as well as the reproductive processes, remain mysteries.

Despite isolated success in raising Japanese eels artificially to the glass eel stage (Tanaka et al., 2003), eels cannot be stably bred in captivity. Due to the complex and opaque biology of the genus *Anguilla*, the global eel market relies uniquely on the fishing of eels at all life stages with associated international trade in livestock (Nielsen and Prouzet, 2008). While from the mid-1990s on the eel market volume has increased rapidly (FAO Report, 2001) a serious decline in populations of *A. anguilla*, *A. rostrata* and *A. japonica* has been observed (Moriarty and Dekker, 1997; Tzeng, 1997; Tatsukawa, 2001). The global volume of live eel exports (all species and all stages) was estimated to be 25,794 tonnes in 1997 (compared to about 5,000 tonnes before 1983) and was valued at 385 US\$ (FAO Report, 2002). Apart from the fishing pressure, additional threats to the worldwide eel populations, such as pollution, human impact on water systems (dams, habitat destruction), climate change and parasitism, have been discussed (Muchiut et al., 2002).

Within the Anguillicolidae, pathogenicity to the host has been studied in detail only in *A. crassus*. A potential parasitological threat was pointed out early by Egusa (1978), who observed that *A. crassus* causes pathological changes in European eels, while it generally does not in Japanese eels. In the swim-bladder of heavily infected *A. anguilla* pathological changes occur and a reduced food intake paired with increased mortality has been observed (Molnár et al., 1993). Effects on the swim bladder wall include inflammation, proliferative changes and alteration of the gas composition of the swim bladder. Especially the latter, the reduction of oxygen content in the swim bladder, together with the implications of the decomposition of dead adults within this organ, are suspected to be the principle factors responsible for the death of eels during high water temperatures (Molnár et al., 1993; Würtz et al., 1996).

Results from exhaustion swim tunnel experiments demonstrate that sexually mature silver eels, highly infected with *A. crassus*, reach lower cruising speeds and have higher costs of transport compared to healthy individuals (Palstra et al., 2007). High infection rates of *A. crassus* could therefore, if this experiment is comparable to the eel's reproductive migration, be lethal in migrating, non-feeding silver eel stages (Lefebvre et al., 2007).

*A. crassus* is now considered a significant pathogen of *A. anguilla* (Taraschewski et al., 1987) and *A. rostrata* (Kennedy, 2007), being only absent in cold regions with an average freshwater temperature below 4 °C (Kirk, 2003), such as Iceland (Kristmundsson and Helgason, 2007).

### 1.3.3. Taxonomic history of the Anguillicolidae and results from phylogenetic analyses of clade 8 nematodes *sensu* Holterman et al. (2006)

Upon the discovery of *A. globiceps*, Yamaguti (1933) created for it the new genus *Anguillicola* and the new family Anguillicolidae, and placed it within the Filaroidea *sensu* Weinland, 1858.

Later, Johnston and Mawson (1940) described *A. australiensis* from the swim bladder of *A. reinhardtii* and grouped it in the same genus.

Skryabin (1954) dealt with the Dracunculoidea as one of the two superfamilies of the sub-order Camallanata *sensu* Chitwood, 1936 and recognised three families: the Anguillicolidae, Dracunculidae and Tetanonematidae. Ivashkin et al. (1971) removed the Anguillicolidae from the Dracunculoidea and created the independent superfamily Anguillicoidea in which he also place the Skrjabillanidae. In contrast, Anderson et al. (1975) did not remove the Anguillicolidae from the Dracunculoidea.

Triggered by the introduction of the highly-pathogenic East-Asian swim bladder nematode *A. crassus* into Europe, Moravec and Taraschewski (1988) carried out a taxonomic revision of members of this genus, describing two new species of eels from New Zealand (*A. novaezealandiae*) and South African eels (*A. papernai*), briefly redescribing already known species, erecting the new subgenus *Anguillicoloides* (type species *A. crassus* Kuwahara, Niimi et Itagaki, 1974) and providing a key for species identification.

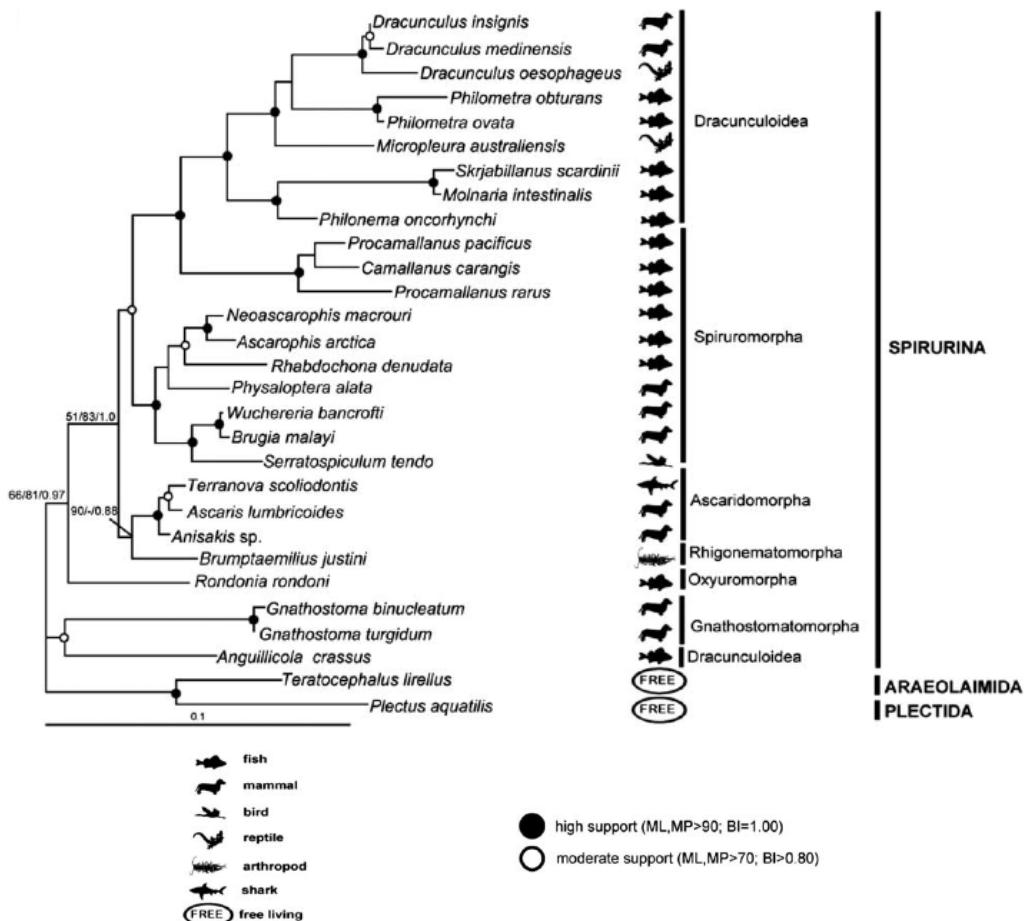
Later, Hirose et al. (1998) sequenced the 18S rDNA gene of *A. crassus* and *A. globiceps* and found high sequence identity between both species, suggesting the inability of this marker to distinguish between them.

Wijová et al. (2006) conducted a phylogenetic analysis of the 18S rDNA sequences of spirurine nematodes with special emphasis on dracunculoids and found *A. crassus* together with *Gnathostoma binucleatum* and *Gnathostoma turgidum* to lie basal to all Spirurina (see Figure 1.3). This result was confirmed by a more comprehensive analysis undertaken by Nadler et al. (2007) which included 113 18S rDNA sequences of clade III nematodes *sensu* Blaxter et al. (1998) and contradicted the evidence for placing the *A. crassus* within the Dracunculoidea further.

In a review article, Moravec (2004) pointed out the inaptness of the present classification system of dracunculoids to reflect phylogenetic relationships and the necessity of a taxonomic revision of this group of nematodes based on detailed morphological, life history and molecular phylogenetic studies.

Holterman et al. (2006), in their analysis of 339 taxa, suggested a division of the phylum Nematoda into 12 clades, placing *G. turgidum* basal to clade 8. In the most extensive phylogenetic analysis of the phylum Nematoda up to date, Van Meegen et al. (2009) analysed 1200 nematode taxa and found, *Truttaedactinis truttae* together with *A. crassus* and three gnathost-

matid nematodes to occupy the most basal position within clade 8.



**Figure 1.3.: Phylogeny of Spirurina *sensu* De Ley and Blaxter (2002) based on phylogenetic analyses of the 18S rDNA sequences (see Wijová et al., 2006).**

#### 1.4. Implications of the phylogenetic position of *A. crassus* on the origin of parasitism within the Nematoda

It is safe to assume that the common ancestor of the Nematoda was free-living, since the number of free-living nematodes exceeds the number of parasitic ones. The alternative hypothesis, a secondary loss of the parasitic trait in most lineages, can thus be rejected based on the violation of the general assumption of evolutionary parsimony.

By inferring a phylogeny for the Nematoda, Blaxter et al. (1998) noted that parasitism arose at least seven times within this phylum. Four groups of animal parasites and three groups of plant parasites are interspersed with free-living, aquatic and terrestrial species in the phylogenetic tree, suggesting that parasitism evolved independently from one another.

The underlying processes necessary for a free-living nematode to adopt a parasitic life-style, however, remain a mystery. It is clear that some kind of initial association between a proto-parasitic nematode and its future host had to exist, in order to set the stepping stone for a host-parasite-system. For instance, many free-living nematodes rely on other organisms as transport hosts, as means of getting from one food patch to another. This non-parasitic symbiotic association is either opportunistic or essential for the nematode to colonise new resources (Bovien, 1937).

The concept of initial association has been termed pre-adaptation (Osche, 1956) and comes from the idea that certain adaptations to a current environment might be co-opted to a new function and facilitate the transition to a new environment. Osche (1956) states that phoresy (the usually unspecific use of a different species as a transport host) as well as necromeny (the non-parasitic association with a different species in order to feed on the posthumous development of microbes on the host's tissue) could be considered as pre-adaptations towards parasitism. *Caenorhabditis elegans*, for example, is considered to display phoresy since the Dauer larvae, a developmental arrested and non-feeding stage specialised on survival, uses insects or other invertebrates for transportation between habitats (Kiontke and Sudhaus, 2006).

However, the concept of pre-adaptation is highly hypothetical, since no organism can foresee evolution and therefore cannot pre-adapt to future environments. Furthermore, phylogenetic data (Holterman et al., 2006; Nadler et al., 2007; Van Megen et al., 2009), at least for the entirely parasitic clade 8 *sensu* Holterman et al. (2006), seems to contradict this theory, since vertebrate-parasitic taxa, namely *Truttaedactinis truttae*, *A. crassus* and three gnathostomatid nematodes, are found to lie basal to numerous arthropod-parasitic nematodes. All five species display rather complex, heteroxenous life-cycles, while, if the pre-adaptationist theory of the origin of parasitism was true, they would be expected to show either facultative parasitism or monoxenic life cycles.

## 1.5. Barcoding genes and phylogenetic analysis

*"We wish to suggest a structure for the salt of deoxyribose nucleic acids (D.N.A.).*

*This structure has novel features which are of considerable biological interest."*

- J. D. Watson and F. H. Crick (1953)

The quotation above not only served as the introduction for one of the most important publications in the biological sciences (see Watson and Crick, 1953), but initiated a new era in science. Among the gargantuan number of discoveries that followed the publication of this paper, the most important for this study is the variation in the information that is found coded in this

molecule, between organisms. As an example for the possible diversity of information, Hebert et al. (2003) states that by focusing on any sequence of 45 nucleotides of a protein-coding gene one gains access to 15 sites weakly affected by selection, since most mutations occur at the third nucleotide position due to the four-fold degeneracy of the genetic code. This leads to 1 billion possible combinations only weakly affected by selection.

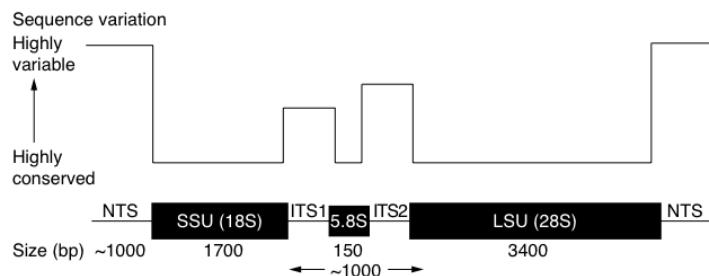
In the case of the Nematoda, species identification through morphological traits requires great amount of expertise and time, making it impossible to describe biodiversity solely based on this approach (Blaxter, 2003). This made the use of characteristic species-specific DNA sequences, termed “DNA barcodes”, for species identification attractive, and today DNA barcoding is the *modus operandi* of choice for many nematologists (e.g. Floyd et al., 2002; Powers, 2004; De Ley et al., 2005; Valentini et al., 2009).

Basically, a barcoding gene must meet the same criteria as any gene suitable for inferring a phylogeny among species: presence in all taxa of interest, availability of highly conserved, peripheral regions to archive primer annealing for amplification and sequencing and, ultimately, highly diverse regions in between from which information can be abstracted.

The primary sequence used for barcoding Metazoa is the 648-bp region of the cytochrome c oxidase subunit I (*Cox1* or *COI*) (Hebert et al., 2003; Savolainen et al., 2005). The gene cytochrome c oxidase is located within the mitochondrial genome (mtDNA) and its protein plays a crucial role in the respiratory chain of mitochondria, conducting the transfer of electrons to oxygen molecules, which as a result increases the electrochemical proton gradient across the inner membrane necessary for the synthesis of ATP. The *Cox1* fragment has two main advantages: first, due to the average copy number of mtDNA of 45 copies per cell (Lemire, 2005) *Cox1* can easily be amplified and sequenced; and second, the high mutational rate in comparison to nuclear genes allows discrimination between taxa at a population level (Blouin et al., 1998). However, the use of *Cox1* alone for barcoding and phylogenetic inference has lately been subject to criticism. Factors such as maternal inheritance of the mitochondrial genome, nuclear integration of mtDNA (“pseudogenes”), direct selection (on mtDNA itself) and indirect selection (arising from the disequilibrium with other maternally transmitted genetic material) are suggested to make inferences from mtDNA data alone unreliable (see Hurst and Jiggins, 2005). In order to take advantage of the strengths but also to account for its shortcomings, mtDNA is used together with other independent data sources, such as nuclear DNA (Rubinoff and Holland, 2005).

The nuclear ribosomal RNA genes (rDNA or rRNA) have been used widely in nematode phylogenetics and barcoding (e.g. Blaxter et al., 1998; Dorris et al., 1999; Floyd et al., 2002), as they are present in multiple copies (tandem repeats) per genome. *C. elegans*, for instance, has about 55 identical sets of rRNA cistrons (Ellis et al., 1986), which are maintained through concerted evolution of the rRNA loci accomplished by gene conversion (Charlesworth et al.,

1994). The graphic representation of the structure of a rDNA repeat is illustrated in Figure 1.4.



**Figure 1.4.:** The nuclear ribosomal DNA (rDNA) cistron. Sizes are approximate and not to scale. Each cistron comprises the small subunit rDNA gene (18S rDNA), the internal transcribed space I (ITS1), the 5.8S gene, ITS2 and the large subunit rDNA gene (28S rDNA). An external non-transcribed spacer (NTS) separates each transcribed cistron. The relative rate of sequence variation observed between taxa is shown schematically above the cistron (from Dorris et al., 1999).

Variation in 18S rDNA is a common metric for phylogenetic analysis across phyla up to genus-levels (e.g. Blaxter et al., 1998; Wijová et al., 2006; Holterman et al., 2006; Nadler et al., 2007; Meldal et al., 2007). The 28S rDNA is longer in sequence and composed of a greater mixture of conserved and divergent regions. These regions have been termed “divergence (D) regions” and are numbered in the 5’ to 3’ direction of the mature rRNA (Hassouna et al., 1984). De Ley et al. (2005) considered the D2-D3 region of the 28S rDNA as promising loci for DNA barcoding of nematodes, and its use for analysing relationships among orders of the Nematoda has been demonstrated extensively (e.g. Nadler et al., 2006; Subbotin et al., 2007). The regions with the highest amount of sequence variation observed between taxa are ITS1 and ITS2. Both have been used widely for inferring species-relationships (e.g. Dorris et al., 1999; Nadler and Hudspeth, 2000) and are variable enough to separate even cryptic species (Chilton et al., 1995).

## 2. Aims

The family Anguillicolidae (Nematoda: Anguillicoloidea) comprise five species of ichthyoparasitic nematodes, grouped into the two genera *Anguillicola* and *Anguillicoloides*, which, as adults, feed on blood within the swim bladder of fresh-water eels of the genus *Anguilla*. The species *Anguillicoloides crassus*, previously restricted to populations of the Japanese eel *A. japonica*, is now considered a global invader *sensu* Colautti and MacIsaac (2004) and Taraschewski (2006) after colonising Europe, North Africa and North America, where the two *naive* recipient host species *A. anguilla* and *A. rostrata* resulted highly susceptible and vulnerable. Furthermore, phylogenetic analyses of clade 8 nematodes *sensu* Holterman et al. (2006) found this species to lay within the basal taxa of this entirely parasitic clade of nematodes, which raised questions regarding the processes governing the evolution of parasitism within the phylum Nematoda.

The aims of this diploma thesis are:

- The sequencing of nuclear and mitochondrial barcoding genes for all five anguillicolid nematodes to deliver an evaluation of their usefulness for studying the Anguillicolidae.
- The use and development of bioinformatic methods in order to extract the information contained in the sequences.
- The evaluation of the information extracted through bioinformatic analyses at the species and population level for this family of nematodes.
- The phylogenetic analyses of the different barcoding genes in order to re-evaluate the phylogenetic position of *A. crassus* within clade 8 through the incorporation of additional taxa.
- The comparison of the phylogeny of the hosts to the phylogeny of their parasites in order to understand the effects that led to the *invasiveness* of *A. crassus*
- The offering of suggestions for further studies in order to address question that remained unanswered.

## 2. Aims

---

### 3. Materials and Methods

#### 3.1. Sampling

A total of 150 anguillicolid nematodes were extracted by colleagues from the swim bladders of their respective hosts from 23 different locations in Africa, Asia, Europe and Australia.

In addition, three other nematodes parasitising eels from South Africa, Madagascar and Germany, respectively, have been collected. The nematode, collected in Germany was the only one for which a morphological identification to the species level (*Camallanus lacustris*) was possible. The South African nematode (SNR118), a thin and minute larval stage, was isolated from the serosa of the swim bladder of *A. mossambica* and the whole specimen was used for DNA barcoding. The nematode from Madagascar was extracted from a slightly deteriorated *A. mossambica*, so that no exact information about the location inside the host could be given. The locations of the sampling sites are summarised in Table 3.1. The specimens were preserved in 75% ethanol and stored at room temperature in 1.5 ml Eppendorf cups.

Prefix	Site (Country)	Geodetic coordinates			Host species	Collector
		$\phi$	$\lambda$			
AQT	Willow's Golf Course, Townsville, Queensland (Australia)	19°18'S	146°44'E		<i>A. reinhardtii</i>	BS 2007
AQB	Lake's Golf Course, Brisbane, Queensland (Australia)	27°38'S	153°12'E		<i>A. reinhardtii</i>	BS 2007
ATD	Deloraine, Tasmania (Australia)	41°31'S	146°39'E		<i>A. australis</i>	LP 2008
CGG	Guangzhou, Guangdong (China)	23°07'N	113°15'E		<i>A. anguilla</i>	HT 03/07
CGZ	Zhuhai, Guangdong (China)	22°16'N	113°34'E		<i>A. anguilla</i>	HT 03/07
EAV	Albufera de Valencia (Spain)	39°21'N	0°20'W		<i>A. anguilla</i>	PMR 01/09
GST	Steinfeld (Germany)	49°02'N	8°02'E		<i>A. anguilla</i>	AK 2009
GRA	RuBheimer Altrhein (Germany)	49°12'N	8°25'E		<i>A. anguilla</i>	EH, UW 2009
JPN	Natural water system, Wakayama (Japan)	34°13'N	135°10'E		<i>A. japonica</i>	HS 2006 2007
MAD	Marovoay (Madagascar)	16°06'S	46°36'E		<i>A. mossambica</i>	OW 05/2008
POL	Sniardwy Lake, Mikolajki (Poland)	53°45'N	21°43'E		<i>A. anguilla</i>	UW 2009
POR	Ribeira das Lampreias (Portugal)	38°47'N	9°01'W		<i>A. anguilla</i>	JLC 03/09
SFH	Farm Dam, Fort Hare (South Africa)	32°47'S	26°50'E		<i>A. mossambica</i>	HT 03/08
SKR	Koonap River (South Africa)	32°1'S	26°08'E		<i>A. mossambica</i>	HT 03/08
SSD	Sunday's River, Slagboom Dam (South Africa)	33°22'S	25°40'E		<i>A. mossambica</i>	HT 03/08
SDD	Sunday's River, Darlington Dam (South Africa)	33°12'S	25°8'E		<i>A. mossambica</i>	HT 03/08
SGF	Great Fish River, Double Drift (South Africa)	33°05'S	26°46'E		<i>A. mossambica</i>	HT 03/08
SNR	Nahoon River, Nahoon Dam (South Africa)	32°54'S	27°48'E		<i>A. mossambica</i>	HT 03/08
TUR	Asi River, Hatay (Turkey)	36°24'N	36°21'E		<i>A. anguilla</i>	EG 12/08
TKR	Sinyuan, Kaoping River (Taiwan)	22°30'N	120°25'E		<i>A. japonica</i>	HT 9/06
TCU	Eel culturing pond, Budai, (Taiwan)	22°38'N	120°26'E		<i>A. japonica</i>	HT 9/06

**Table 3.1.:** Sampling location for the nematodes including label prefix, information on geographic position including latitude ( $\phi$ ) and longitude ( $\lambda$ ) of the sampling site, host species and information about sampling. AK : Albert Keim, BS : Björn Schäffner, EH : Emanuel Heitlinger, EG : Ercüment Genç, HT : Horst Taraschewski, JLC : José Lino Costa, LP : Lea Perseke, OW : Olaf Weyl, PMR : Pilar Muñoz Ruiz, HS : Hiroshi Sato,

### 3.2. Tissue extraction and lysis

In order to extract tissue samples for sequencing, each specimen was transferred individually to a polypropylene petri dish ( $\varnothing 50$  mm) filled with 75% ethanol. Using sterile dissection instruments a small piece of tissue was then abscised from the dorsal cuticula. Caution was taken to avoid mechanical damage to the digestive tract to prevent contamination of the nematode tissue sample through host blood cells.

Subsequently, the tissue sample was washed in a second petri dish containing de-ionised water, placed in 100  $\mu\text{l}$  PCR reaction tube filled with 20  $\mu\text{l}$  microLysis-Plus lysis-buffer (Microzone Ltd., Haywards Health, UK) and incubated on a DNA Engine Dyad Thermal Cycler (MJ Research, now Bio-Rad) following the manufacturer's protocol.

Encapsulated and partly deteriorated specimens of *A. crassus* were extracted from within the intestinal wall of *A. japonica* using sterile dissection pins and transferred individually into petri dishes with de-ionised water. In 25 reactions the collagen sheath forming the capsules was mechanically destroyed and the debris was transferred into 20  $\mu\text{l}$  microLYSIS-PLUS lysis buffer and incubated following the manufacturer's protocol. The same protocol was applied to 5 capsules not mechanically opened and, as a negative control, to 10 pieces of *A. japonica* intestinal tissue showing no traces of infection.

### 3.3. DNA amplification

The lysates of all samples were used directly as templates in the PCR reactions. Since a fairly constant amount of nematode tissue was used in the lysation processes, there was no need to adjust the template volume employed in the PCR reactions beyond the range of 2 – 5  $\mu\text{l}$ .

While lower volumes were used for lysates in which the tissue was not fully degraded, higher volumes were applied when no sound PCR product could be detected *ex post*. All PCR reactions were performed using the chemicals at the concentrations indicated as follows.

- 16.7  $\mu\text{l}$  *ddH<sub>2</sub>O* (Milli-Q)
- 2.7  $\mu\text{l}$  10x PCR Buffer (containing 15mM *MgCl<sub>2</sub>*) (Qiagen, Hilden, Germany)
- 2.7  $\mu\text{l}$  2mM dNTP (2mM dATP, dTTP, dGTP, dCTP)

- 0.4 µl of each PCR primer (10 µM (see Table 3.2)
- 0.1 µl Taq DNA Polymerase ( 5 units/µl ) (Quiagen, Hilden, Germany)

The primers used for the PCR reactions are summarised in Table 3.2. As a topological reference, the primers SSU\_F\_04 and SSU\_R\_26 bind at position 30 - 49 and 927 - 907 in the 18S rDNA of *Caenorhabditis elegans* (X03680 Region: 934..2693), respectively. The primers therefore amplify the first half of the 18S rDNA. This being noted, this will be referred as to the fragment amplified by those primers and to the sequences obtained after sequencing, as 18S rDNA.

The binding sites for the remaining primers are specified in the respective publications, they were taken from.

Gene	Primer name	Primer Sequence	Reference
18S rDNA	SSU_F_04	5'-GCTTGTCTCAAAGATTAAGCC-3'	Blaxter et al. (1998)
	SSU_R26	5'-CATTCTGGCAAATGCTTCG-3'	Blaxter et al. (1998)
ITS1	ITS1	5'-CCCTTTGTACACACCGCC-3'	Miscampbell et al. (2004)
	NC13r	5'-GCTGCGTTCTTCATCGAT-3'	Chilton et al. (1997a)
28S rDNA D2-D3	D2A	5'-ACAAGTACCGTGAGGGAAAGT-3'	Nunn (1992)
	D3B	5'-TGCAGAGGAACCAGCTACTA-3'	Nunn (1992)
ITS2	NC13	5'-GCTTGTCTCAAAGATTAAGCC-3'	Chilton et al. (1997b)
	NC2	5'-TTAGTTCTTTCCCTCGCT-3'	Gasser et al. (1993)
<i>Cox1</i>	LCO1490	5'-GGTCAACAAATCATAAAGATATTGG-3'	Folmer et al. (1994)
	HC02198	5'-TAAACTTCAGGGTGACCAAAAAATCA-3'	Folmer et al. (1994)

**Table 3.2.:** PCR primers used for amplification of the barcoding genes.

For all specimens the 18S rDNA, 28S rDNA D2-D3 and *Cox1* regions were amplified. In the initial phase of the project each sample was amplified in duplicate. This was later reduced to one analysis per sample.

The protocol for amplification of the ITS1 and ITS2 region was applied to several specimens of *A. crassus* samples to provide insight into the relationships between the different populations of this species and to identify possible cryptic species. In Table 3.3 the cycling profile for each of the PCR reactions is shown.

### 3.4. TOPO cloning

DNA amplification of *Cox1* of individual samples was archived by cloning the fragment, after initial PCR amplification, into the pCR® 2.1-TOPO® vector (Invitrogen, Carlsbad, CA, USA)

### 3. Materials and Methods

---

Cycling profile for	18S rDNA	ITS1	28S rDNA	ITS2	Cox1	
<b>Initial denaturation</b>	94°C for 5 min	95°C for 2 min	95°C for 5 min	95°C for 2 min	94°C for 1 min	
<b>Denaturation</b>	94°C for 1 min	95°C for 50 s	94°C for 30 s	95°C for 50 s	94°C for 1 min	94°C for 1 min
<b>Annealing</b>	55°C for 90 s	55°C for 25 s	55°C for 45 s	55°C for 25 s	45°C for 90 s	50°C for 90 s
<b>Elongation</b>	72°C for 2 min	72°C for 1 min	72°C for 2 min	72°C for 1 min	72°C for 90 s	72°C for 1 min
<b>Number of cycles</b>	34	40	34	40	4	34
<b>Final elongation</b>	10 min at 72°C	72°C for 7 min	72°C for 10 min	72°C for 5 min	72°C for 5 min	

**Table 3.3.:** Cycling profiles of the different PCR reactions.

and subsequently transforming TOP10 One Shot® *Escherichia coli* cells (Invitrogen, Carlsbad, CA, USA) using the TOPO TA Cloning® Kit (Invitrogen, Carlsbad, CA, USA).

Extraction of the cloned fragments from the cells was achieved by using the PureLink™ Quick Plasmid Miniprep Kit (Invitrogen, Carlsbad, CA, USA).

### 3.5. Gel electrophoresis

The PCR products were analysed prior to sequencing using standard submarine gel electrophoresis. From each sample 3 µl were loaded onto a 1.5 % w/v agarose/TBE - gel.

After staining the gel with SYBR® Safe DNA gel stain (Invitrogen, Carlsbad, CA, USA), amplification of the target sequences was confirmed through fragment length comparison with a 1 Kb DNA ladder (Invitrogen, Carlsbad, CA, USA) under UV light in a standard gel documentation system.

A second gel electrophoresis using 1.5 % w/v agarose/TAE gel and ethidium bromide, at a concentration of 0.5 µg/ml, for staining, was performed in order to maintain cloned DNA fragment suitable for extraction and further sequencing.

### 3.6. Gel extraction

In order to extract the DNA fragments from a gel, the band in question was excised using a sterile scalpel. The DNA was purified using the QiaQuick PCR purification kit (Qiagen, Hilden, Germany) following manufacturer's protocol.

### 3.7. Sequencing

To guarantee the best performance of the Sanger-sequencing reactions, amplified products were processed removing excess primers and dNTPs.

The method of choice was digestion of single stranded DNA (primers) into dNTPs using *Escherichia coli* Exonuclease I (ExoI) (USB Corporation, Cleveland, USA), followed by removal of phosphate groups from dNTPs employing Shrimp alkaline phosphatase (SAP) (Hanke and Wink, 1994). The concentrations of the reactants involved are listed below.

- 17 µl PCR product
- 1 µl Shrimp alkaline phosphatase (SAP) (1 U/µl)
- 1.5 µl 1/20 diluted Exonuclease I (ExoI) (diluted using SAP dilution buffer)

The purified PCR products were subsequently prepared for sequencing using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) following manufacturer's protocol.

All fragments were sequenced in both directions to minimise PCR artefacts, ambiguities and base-calling errors. Diluted PCR primers were used as sequencing primers. The reaction volumes used for each preparation were as follows:

- 3.68 µl ddH<sub>2</sub>O (Milli-Q)
- 2 µl 5X BigDye® Terminator 3.1 Sequencing Buffer
- 2 µl BigDye® Terminator 3.1
- 0.32 µl Sequencing primer (3.2 pmol/µl)
- 2 µl Purified PCR product

Ultimately, sequencing reactions took place in an automated ABI Prism 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA) in "the Gene- Pool" sequencing facilities based at the School of Biological Sciences at The University of Edinburgh (Scotland, UK).

## 3.8. Bioinformatic methods

### 3.8.1. Processing of raw sequence data

Raw sequences yielded by the dye terminator sequencing reaction provide a transcript of the sequence and raw chromatograph traces (Applied Biosystems (ABI) chromatogram files) that lead to the transcript.

The Perl program trace2seq.pl<sup>1</sup> was used to base-call the raw chromatograph traces generating quality sequences (FASTA format) as well as quality files, such as SCF files (with base

---

<sup>1</sup>A Perl program that uses phred (©1993 - 2002 by Phil Green and Brent Ewing. All rights reserved.) to identify high-quality base calls and crossmatch (©1993 - 1996 by Phil Green and Brent Ewing. All rights reserved.) to identify vector sequences; A. Anthony and M. Blaxter, unpublished

calls and phred-quality-scores) and PHD files (text files which contain base call and quality information).

### 3.8.2. Screening for contaminants

Although during the dissection of the nematodes mechanical damage to the digestive tract was avoided by all means, all sequences were screened for contaminants using the `remoteblast.pl` Perl program (see appendix A.1).

Upon running the program the user is asked to specify the type of BLAST search (`nblast`, `blastp`, `blastx`, `tblastn` or `tblastx`) he wants to run, the database (e.g. `nr`) against which the sequenced should be BLASTed, the hit-report cut-off and the E-value cut-off. The program remotely executes the NCBI BLAST (Altschul et al., 1990) via HTTP and sends each of the sequences in the input file (FASTA format) as an individual query. The results for all queries are stored in a text file containing each query name followed by the description of the sequence, the E-Value, the score and the percentage identity of the BLAST hit. The number of BLAST hits reported are defined by the hit-report cut-off.

For every sequence a BLAST search was performed against the NCBI Nucleotide Collection (`nr`) database with an  $1 e^{-10}$  E-Value cut-off, reporting back three hits per query. Sequences displaying actinopterygic hits were interpreted as host contaminations and excluded from further analysis.

### 3.8.3. Assembly of consensus sequences

Every fragment was sequenced from both orientations generating two partially overlapping sequences. The information contained in both sequences was joined in order to generate a so called consensus sequence, hereafter referred to as  $Seq_c$ . This provides not only an increase in sequence length (due to non-overlapping parts) but an improvement in sequence reliability (within overlapping parts). The improvement arises from the antagonistic, and therefore complementary, decrease in sequencing accuracy towards the ends of fragments.

This task was achieved by using the `seq2con.pl` Perl program (see appendix A.1). Relying on the output of `trace2seq.pl`, the program determines, for each sequence, the high-quality base-calls made by `trace2sew.pl` (through `phred`) and extracts the phred quality score<sup>2</sup> of each high-quality base into a new PHD file. Fragments sequenced in reverse orientation are converted into their reverse complements. Based on sequence name nomenclature (example shown in

---

<sup>2</sup>Phred scores are determined by examining the peaks around each base call assigning a quality score to each base call.

Quality scores range from 4 to about 60, with higher values corresponding to higher quality. The quality scores are logarithmically linked to error probabilities. e.g Phred score = 30 → 99,9% accuracy of base call.

Figure 3.1) the corresponding sequences of a gene are extracted from the generated PHD files and aligned using CLUSTALW 2.0.9 (Larkin et al., 2007).

$$\left. \begin{array}{l} AQT2\_CF \\ AQT2\_CR \end{array} \right\} AQT2\_C$$

**Figure 3.1.:** Nomenclature of sequences illustrated at the example of the *Cox1* sequence of the second nematode sampled from Townsville, Australia. “AQT” is the population prefix, the specimen number is “2” and the gene suffix is “C”

As a further check for contamination or inconsistency regarding the nature of the sequences, the percentage identity (PID) of each alignment is compared against a threshold of 85. Alignments below the threshold are extracted into \*.nomatch.aln (CLUSTALW file) in the alignments/ folder for further inspection by hand. Alignments above the threshold are extracted as \*.aln (CLUSTALW file) in the alignments/ folder and a IUPAC (International Union of Pure and Applied Chemistry) consensus sequence is generated, where ambiguities are shown following the IUPAC rules. Within the alignment the first and last base of each sequence is then determined, giving the positional framework for sequence comparison.

In order to ease further description, a graphical representation of an alignment and the associated Phred quality scores, together with the resulting consensus sequence  $Seq_c$ , are given in Figure 3.2.

		$S_{start}$			$S_{body}$				$S_{end}$		
Alignment position	$S$	1	2	3	4	5	6	7	8	9	10
DNA sequences	<i>A</i>	C	T	G	C	T	A	A	–	–	–
	<i>B</i>	–	–	G	C	A	G	T	T	A	T
Phred quality scores	<i>A</i>	40	40	45	50	50	45	15	–	–	–
	<i>B</i>	–	–	40	45	20	42	40	40	25	40
$Seq_c$		C	T	G	C	T	Y	T	T	N	T

**Figure 3.2.:** Graphical illustration of the framework for the determination of the consensus sequence  $Seq_c$  based on the Phred quality scores of the two sequences *A* and *B*

The program starts at the first position (1) of the alignment. If  $S_{start}$  is defined, it is composed solely of bases of sequence *A* and therefore unambiguous. Hence  $Seq_c$  inherits  $S_{start}$  from sequence *A* including the quality values.

Upon reaching  $S_{body}$  ambiguity can be present. During this automated sequence comparison, two scenarios can occur at any given alignment position ( $x$ ) within  $S_{body}$ . Either sequence *A* equals sequence *B* at the position in question ( $Seq_A(x) = Seq_B(x)$ ) or they differ

( $\text{Seq}_A(x) \neq \text{Seq}_B(x)$ ). In the former case the program determines the highest phred score at position  $x$ , which  $\text{Seq}_c$  inherits. In the latter case additional possibilities can arise at position  $x$ :

- a) If sequence A exhibits a different base than sequence B the Phred quality scores of A ( $\text{Qual}_A(x)$ ) and B ( $\text{Qual}_B(x)$ ) are compared:
  - i) If  $\text{Qual}_A(x) < \text{Qual}_B(x) \wedge |\text{Qual}_A(x) - \text{Qual}_B(x)| > 5 \Rightarrow$  Sequencing error at  $\text{Seq}_A(x)$
  - ii) If  $\text{Qual}_A(x) > \text{Qual}_B(x) \wedge |\text{Qual}_A(x) - \text{Qual}_B(x)| > 5 \Rightarrow$  Sequencing error at  $\text{Seq}_B(x)$
  - iii) If  $\text{Qual}_A(x) \leq \text{Qual}_B(x) \wedge |\text{Qual}_A(x) - \text{Qual}_B(x)| < 5 \Rightarrow$  SNP<sup>3</sup> at position  $x$
- b) One of the two sequences exhibits a gap “.”:
  - i) If  $\text{Seq}_A(x) = \{\cdot\} \wedge \text{Qual}_B(x) > 30 \Rightarrow$  False negative base-call at  $\text{Seq}_A(x)$
  - ii) If  $\text{Seq}_A(x) = \{\cdot\} \wedge \text{Qual}_B(x) < 30 \Rightarrow$  False positive base-call at  $\text{Seq}_B(x)$
  - iii) If  $\text{Seq}_B(x) = \{\cdot\} \wedge \text{Qual}_A(x) > 30 \Rightarrow$  False negative base-call at  $\text{Seq}_B(x)$
  - iv) If  $\text{Seq}_B(x) = \{\cdot\} \wedge \text{Qual}_A(x) < 30 \Rightarrow$  False positive base-call at  $\text{Seq}_A(x)$

In the case of sequencing errors as well as in presence of false negative base-calls  $\text{Seq}_c$  inherits from the sequence with the highest phred score. Upon encountering SNPs the corresponding ambiguous letter from the IUPAC consensus sequence is adopted and in the event of false positive base-calls the program simply proceeds. The relative changes in sequence position for the phred-score sequences due to gaps are accounted for.

After completing the analysis of  $S_{body}$ ,  $S_{end}$ , if defined, is appended to  $\text{Seq}_c$  including it's phred quality scores.

The final output consists of the sequence of  $\text{Seq}_c$  in FASTA format, its base call and quality information in PHD format is saved in the `consensus/` folder and an alignment containing  $\text{Seq}_A$ ,  $\text{Seq}_B$  and  $\text{Seq}_c$  is extracted into the `alignments/` folder. Additionally a log file is created.

### 3.8.4. Construction of consensus sequence alignments

In order to further analyse the sequences, all sequences of the same gene were bundled into one file. For each type of gene, the respective consensus sequences of the samples were imported into Mesquite v2.6 (Maddison and Maddison, 2009) and saved in the NEXUS format. The alignment was constructed executing the implemented CLUSTALW 2.0.9 (Larkin et al., 2007) with the penalties for gap opening and extension of 10 and 0.2, respectively.

Each alignment was inspected by eye and variable regions confirmed by hand visualising the SCF quality trace files of the sequence of interest in 4peaks (Griekspoor and Groothuis, 2006).

---

<sup>3</sup>SNP = Single nucleotide polymorphism

Additionally, a partitioned alignment composed of all the sequences of the different genes for all taxa was constructed. The alignment is referred to as the “supermatrix”.

The alignments were saved using NEXUS, FASTA and PHYLIP 3.4 format.

### 3.8.5. Inclusion of additional anguillicoloid taxa in the datasets

In order to increase sampling size, the 18S rDNA dataset was expanded by including five additional *A. crassus* sequences (Hirose et al., 1998; Blaxter and Roche, 2005; Wijová et al., 2006) and one *A. globiceps* sequence (Hirose et al., 1998) identified through a BLAST search.

All *A. crassus* *Cox1* sequences from Wielgoss et al. (2008), as well as an unpublished dataset of Geiß and Sures (2010a), containing *Cox1* sequences from *A. crassus*, sampled near Essen, Germany (C) and in Lake Bracchiano, Italy (LB); and sequences from *A. novaezealandiae*, sampled in New Zealand (AN, N), were added to the *Cox1* dataset. The exact sampling sites of the specimens sequenced by Wielgoss et al. (2008) are listed in the corresponding publication (Wielgoss et al., 2008, Table 1).

## 3.9. Phylogenetic Analysis

### 3.9.1. Search for outgroups

An outgroup is a taxon or group of taxa closely related to the members of the ingroup, while being more distantly related than any of the ingroup members among each other. This makes it possible to infer a rooted phylogeny.

In the case of the 18S rDNA dataset, outgroup taxa were selected based upon literature (Wijová et al., 2006; Nadler et al., 2007). Additionally, sequences from an unpublished data set of 13 sequences of spirurine nematodes provided by Nadler (2010) were included.

Regarding the 28S rDNA D2-D3 and the *Cox1* dataset, BLAST searches were performed in order to identify outgroup taxa. Also, for the ITS1 dataset a BLAST search was performed but, due to the lack of close hits within spirurine genera, no outgroup was chosen. The outgroups that were selected for the phylogenetic analyses are listed in Table 3.4.

Species name	Source	18S rDNA	“Supermatrix”
<i>Plectus aquatilis</i>	GenBank® GQ892827.1	+	-
<i>Teratocephalus lirellus</i>	GenBank® AF036607.1	+	-
Unidentified nematode SNR-118	Nahoon River (South Africa)	+	+
<i>Linstowinema</i> sp.	N536 (Nadler, 2010)	+	+

continued on next page

### 3. Materials and Methods

---

Species name	Source	18S rDNA	“Supermatrix”
<i>Tanqua tiara</i>	N691 (Nadler, 2010)	+	+
<i>Echinocephalus overstreeti</i>	N411 (Nadler, 2010)	+	+
Unidentified nematode MAD-19	Marovoay (Madagascar)	+	-
<i>Proleptus trygonorrhini</i>	N533 (Nadler, 2010)	+	-
<i>Ascarophis arctica</i>	GenBank® DQ094172.1	+	-
<i>Neoascarophis macrouria</i>	GenBank® DQ442660.1	+	-
<i>Rhabdochona guerreroensis</i>	N644 (Nadler, 2010)	+	-
<i>Rhabdochona denudata</i>	GenBank® DQ442659	+	-
<i>Physaloptera thalacomys</i>	N661 (Nadler, 2010)	+	-
<i>Physaloptera alata</i>	GenBank® AY702703	+	-
<i>Breinlia mundayi</i>	N414 (Nadler, 2010)	+	-
<i>Wuchereria bancrofti</i>	GenBank® AF227234.1	+	-
<i>Brugia malayi</i>	GenBank® AF036588.1	+	-
<i>Dirofilaria immitis</i>	GenBank® AF036638.1	+	-
<i>Loa loa</i>	GenBank® DQ094173.1	+	-
<i>Acanthocheilonema viteae</i>	GenBank® DQ094171.1	+	-
<i>Onchocerca cervicalis</i>	GenBank® DQ094174.1	+	-
<i>Setaria digitata</i>	GenBank® DQ094175.1	+	-
<i>Serratospiculum tendo</i>	GenBank® AY702704.1	+	-
<i>Dracunculus lutrae</i>	N478 (Nadler, 2010)	+	-
<i>Dracunculus medinensis</i>	GenBank® AY947720.1	+	-
<i>Dracunculus insignis</i>	GenBank® AY947719.1	+	-
<i>Dracunculus oesophageus</i>	GenBank® AY852269.1	+	-
<i>Micropleura australiensis</i>	GenBank® DQ442678.1	+	-
<i>Philometra fujimotoi</i>	GenBank® DQ076680.2	+	-
<i>Philometra ovata</i>	GenBank® DQ442677.1	+	-
<i>Philometra obturans</i>	GenBank® AY852267.1	+	-
<i>Skrjabillanus scardinii</i>	GenBank® DQ442669.1	+	-
<i>Molnaria intestinalis</i>	GenBank® DQ442668.1	+	-
<i>Philonema oncorhynchi</i>	GenBank® DQ442670.1	+	-
<i>Procamallanus philippinensis</i>	N689 (Nadler, 2010)	+	-
<i>Procamallanus pacificus</i>	GenBank® DQ442665.1	+	-
<i>Spirocammallanus rarus</i>	GenBank® DQ494195.1	+	-
<i>Camallanus lacustris</i>	RuBheimer Altrhein (Germany)	+	-
<i>Camallanus lacustris</i>	GenBank® DQ442663.1	+	-
<i>Terranova scoliodontis</i>	GenBank® DQ442661.1	+	-
<i>Ascaris lumbricoides</i>	GenBank® U94366.1	+	-
<i>Toxocara canis</i>	GenBank® U94382.1	+	-
<i>Toxocara cati</i>	GenBank® EF180059.1	+	-
<i>Anisakis pegreffii</i>	GenBank® EF180082.1	+	-
<i>Brumptaemilius justini</i>	GenBank® AF036589.1	+	-
<i>Macropoxyuris</i> sp.	N537 (Nadler, 2010)	+	-
<i>Enterobius vermicularis</i>	N427 (Nadler, 2010)	+	-
<i>Rondonia rondoni</i>	GenBank® DQ442679.1	+	-
<i>Gnathostoma neoprocyonis</i>	GenBank® Z96947.1	+	+
<i>Gnathostoma turgidum</i>	GenBank® Z96948.1	+	+

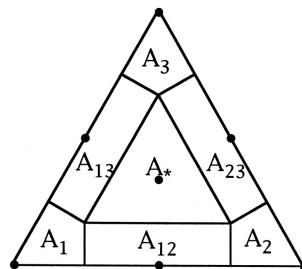
continued on next page

Species name	Source	18S rDNA	“Supermatrix”
<i>Gnathostoma binucleatum</i>	GenBank® Z96946.1	+	+
<i>Dichelyne mexicanus</i>	N626 (Nadler, 2010)	+	-
<i>Culcullanus robustus</i>	N624 (Nadler, 2010)	+	-
Species name	Source	28S rDNA D2-D3	“Supermatrix”
<i>Toxocara vitulorum</i>	GenBank® FJ418790.1	+	+
<i>Contraeicum spiculigerum</i>	GenBank® AB189984.1	+	+
Unidentified nematode SNR-118	Nahoon River (South Africa)	+	+
Species name	Source	CoxI	“Supermatrix”
<i>Toxocara cati</i>	GenBank® Mitochondrial genome (6055-7632)	+	+
<i>Contraeicum cf. spiculigerum</i>	GenBank® U57031.1	-	+
<i>Strongylida</i> sp.	GenBank® FJ172978.1	+	-

**Table 3.4.:** Sequences included as outgroup taxa in the phylogenetic analyses of the different datasets (18S rDNA, 28S rDNA, *CoxI* and the “supermatrix”)

### 3.9.2. Assessment of phylogenetic utility of the alignments

To visualise the utility of a set of aligned sequences for phylogenetic analysis, the likelihood-mapping method (Strimmer and von Haeseler, 1997) implemented in TreePuzzle 5.2 (Schmidt et al., 2002) was used.



**Figure 3.3.:** Framework for likelihood mapping.  $A_1$ ,  $A_2$  and  $A_3$  show the tree-like regions.  $A_{12}$ ,  $A_{23}$  and  $A_{13}$  represent the net-like regions.  $A_*$  displays the star-like region. (see Strimmer and von Haeseler (1997))

Likelihood-mapping is a graphical method based on an analysis of the maximum likelihoods for the three fully resolved tree topologies that can be computed for four sequences. The three likelihoods are represented as one point inside an equilateral triangle (Figure 3.3), that is partitioned into regions. If the point falls into region  $A_*$  a star tree is the optimal tree. The regions  $A_{12}$ ,  $A_{23}$  and  $A_{13}$  represent the situation in which it is difficult to distinguish between two trees, while the remaining three regions,  $A_1$ ,  $A_2$  and  $A_3$ , contain the points where the phylogeny can be completely resolved. For every subset of 4 sequences in a given alignment the likelihoods are mapped onto the triangle and the resulting distribution provides information regarding the phylogenetic utility of the alignment.

The alignments in the PHYLIP 3.4 format were imported into TreePuzzle 5.2 and a likelihood map was constructed under the HKY model of nucleotide substitution with parameters estimated from the dataset and using all possible quartets of taxa.

### 3.9.3. Model selection for DNA evolution

On a molecular level, evolution manifests itself as changes within a DNA molecule.

One form of change, the point mutation, is the substitution of one nucleotide by another. Since it is statistically possible that a given nucleotide experiences multiple point mutations, the actual number of point mutations likely exceeds the number of observable point mutations. Therefore, upon calculating genetic distances between sequences, models for DNA evolution are used for the correction of the observed genetic distance.

The program MrModeltest2 (Nylander, 2004) was chosen to perform hierarchical likelihood ratio tests (hLRT) and calculate the approximate AIC (Akaike Information Criterion) values of 24 nucleotide substitution models implemented in MrBayesv3.2.

The best performing models were used in the Bayesian inference and maximum likelihood analyses and are listed in Table 3.5.

Dataset	AIC	hLRTs
18S rDNA	GTR+Γ	SYM
ITS1	F81	F81
28S rDNA	HKY+Γ	HKY+Γ
<i>Cox1</i>	HKY+I+Γ	GTR+I+Γ

**Table 3.5.:** Results from the calculation of the approximate AIC (Akaike Information Criterion) values and the hierarchical likelihood ratio tests (hLRT) performed by MrModeltest2 for the different datasets.

### 3.9.4. Indel Coding

Apart from nucleotide substitutions, a DNA molecule can experience other forms of alterations during the course of evolution, such as the addition or removal of nucleotides. These are termed insertions and deletions, and referred to as “indels” since a distinction is usually problematic.

In general, the evolutionary information embodied in these phenomena is not used by software for phylogenetic inference, since the sequences are provided in the form of an alignment and the information is parsed column by column.

However, several methods for including indels by coding them within the data are available and the method of Simmons and Ochoterena (2000) implemented in the program SeqState 1.4 (Müller, 2005) was used.

To illustrate the method an example alignment of 5 sequences is shown in Figure 3.4 with its respective indel coding matrix depicted in Figure 3.5. Within the coding matrix the gaps in the alignment, representing the insertion/deletion events, are coded as either present, absent or inapplicable. The latter is chosen upon encountering subset gaps of larger gaps, since it is not possible to infer the absence or presence of a subset gap in larger gap.

	1	5	10	15	20	25
Sequence A	G G - - -	1 - - -	C C T T - - -	3 - - - - -	- G G	
Sequence B	G G - - -	1 - - -	C C T T - - -	3 - - - - -	- G G	
Sequence C	G G A A A - -	2 - - -	T T - 4 - A C - 5 - -	A A A G G		
Sequence D	G G A A A - -	2 - - -	T T - 4 - A C - 5 - -	A A A G G		
Sequence E	G G A A A C C C C C	C T T C A A A C C C C	A A A G G			

**Figure 3.4.:** Example alignment from Simmons and Ochoterena (2000); **1** = gap from positions 3 to 9; **2** = gap from positions 6 to 11; **3** = gap from positions 14 to 25; **4** = gap from positions 14 to 16; **5** = gap from positions 19 to 22.

Gap	1	2	3	4	5
Sequence A	1	0	1	?	?
Sequence B	1	0	1	?	?
Sequence C	0	1	0	1	1
Sequence D	0	1	0	1	1
Sequence E	0	0	0	0	0

**Figure 3.5.:** Simple indel coding data matrix of the alignment in Figure 3.4. “0” = gap is absent, “1” = gap is present, “?” = inapplicable (see Simmons and Ochoterena (2000)).

### 3.9.5. Phylogenetic analysis using Bayesian Inference, as implemented in MrBayes

MrBayes (Ronquist and Huelsenbeck, 2003; Huelsenbeck and Ronquist, 2005) is a program for the Bayesian inference of phylogeny. The Bayes’ theorem, applied to the phylogeny problem, states

$$f(\tau, \nu, \theta | D) = \frac{f(\tau, \nu, \theta) f(D | \tau, \nu, \theta)}{f(D)} \quad [3.1]$$

where  $D$  is the data matrix,  $\tau$  the topology of the tree,  $\nu$  is a vector of branch (or edge) lengths on the tree, and  $\theta$  is a vector of substitution model parameters.

The tree is defined through  $\tau$  and  $\nu$ ; and the model for DNA evolution is contained in  $\theta$ . The distribution  $f(\tau, \nu, \theta)$  is referred to as the *prior*, and specifies the prior probability of different parameter values;  $f(D | \tau, \nu, \theta)$  is the *likelihood function*, describing the probability of the data

under different parameter values, and  $f(D)$  is the total probability of the data summed and integrated over the parameter space.

Bayesian inference is based on the so-called *posterior probability distribution*  $f(\tau, \nu, \theta | D)$ . Usually it is not possible to calculate the *posterior probability* of a phylogeny analytically. However, the *posterior probability* can be approximated by sampling trees from the posterior probability distribution, for which a variant of the Markov chain Monte Carlo (MCMC) - algorithm is used.

A random tree  $T_i$  (with random values for  $\tau$  and  $\nu$ ) serves as initial state of the Markov chain. The parameters  $\tau$  and  $\nu$  are then changed randomly leading to the new tree  $T_j$ .  $T_j$  is accepted as new state of the chain, if

$$A = \frac{P(T_j | D)}{P(T_i | D)} = \frac{\frac{P(D|T_j)P(T_j)}{P(D)}}{\frac{P(D|T_i)P(T_i)}{P(D)}} = \frac{P(D|T_j)P(T_j)}{P(D|T_i)P(T_i)} \geq 1 \quad [3.2]$$

To maximise sampling of the parameter space, MrBayes features two independent runs at 4 Markov chains each, which simultaneously grow towards the tree with the highest *posterior probability*. The number of generations needed for the Markov chains to reach stationarity is referred to as *burn-in* and was discarded before the computation of the majority-rule consensus trees.

For the datasets of 18S rDNA, 28S rDNA D2-D3 and *Cox1*, as well as for the “supermatrix”, phylogenies were reconstructed using MrBayes. The parameters of each analysis are specified in section 3.9.9. All files were named according to the models and parameters applied. Whenever mixed data types, i.e. different types of DNA sequences or DNA sequences combined with binary data, were analysed or mixed models were applied, the data was divided into partitions. Each partition was then entitled to its own parameters. Data gained from coding indels in the alignment according to Simmons and Ochoterena (2000) using SeqState was incorporated in the analysis as a binary partition under the restriction site model with the ascertainment bias set to variable, as suggested by Ronquist et al. (2005). Files containing this type of data bear the acronym `sic` for “simple indel coding”.

### 3.9.6. Assessment of convergence of Markov chains

If the two runs of an analysis have not reached stationarity after the total number of generations, it is assumed that the Markov chains of at least one of the runs has converged on a local optimum in the parameter space.

One method for assessing convergence of two MrBayes runs, is the PSFR (potential scale reduction factor) (Gelman and Rubin, 1992) implemented in MrBayes which should approach 1 for every parameter as runs converge.

In addition, the program Tracer v1.4 (Rambaut and Drummond, 2007) was used to visualise the results of the MrBayes runs in order to assess convergence and ease the setting of the burn-in.

### 3.9.7. Comparison of the models proposed by MrModeltest using MrBayes output

Apart from hLRTs and AIC, nucleotide substitution models can be compared pair-wise using Bayes factors. The Bayes factor for a comparison of the two models  $M_1$  and  $M_2$  is

$$B_{1,2} = \frac{P(D|M_1)}{P(D|M_2)} \quad [3.3]$$

where  $D$  is the dataset and  $P(D|M_i)$  is the *marginal likelihood* of model  $M_i$ .

The Bayes factor is a summary of the evidence provided by the data in favour of one scientific theory as opposed to another and can be viewed as the posterior odds of  $M_2$  to  $M_1$  in a Bayesian inference problem, starting with equal probability of the two models being true (Kass and Raftery, 1995).

Unlike likelihood ratio tests (LRTs), there is no requirement for the models to be nested. Furthermore, there is no need to correct for the parameters in the model, in contrast to comparisons based on the Akaike Information Criterion (AIC).

The *marginal likelihood* of a model is difficult to calculate accurately, but can be estimated through the harmonic mean of the associated likelihoods of MCMC samples from the posterior distribution (Newton and Raftery, 1991). The logarithm of the marginal likelihoods of different models was estimated through MrBayes as the total harmonic mean of the likelihood values of the MCMC samples from the stationary phase of the MrBayes runs.

The logarithm of the Bayes factor for two models  $M_1$  and  $M_2$  was calculated as the difference between the logarithm of the total harmonic mean of the two models.

$$\ln(B_{1,2}) = \ln(HM(M_1)) - \ln(HM(M_2)). \quad [3.4]$$

The resulting  $\ln(B_{1,2})$  was interpreted following Table 3.6.

### 3.9.8. Tests of taxon monophyly using MrBayes

In order to test phylogenetic hypotheses, such as the support for the monophyly of specific taxa, the output from the MrBayes runs was used and Bayes factors were calculated as described in section 3.9.7.

The total harmonic mean likelihood of models in which taxon monophyly was constrained were compared to models in which there was no such constraint.

$\ln(B_{1,2})$	$\log_{10}(B_{1,2})$	$B_{1,2}$	Support for $M_1$ (or evidence against $M_2$ )
0 – 1	0 – 0.5	1 – 3.2	Not worth more than a bare mention
1 – 3	0.5 – 1	3.2 – 10	Substantial
3 – 5	1 – 2	10 – 100	Strong
> 5	> 2	> 150	Very Strong

**Table 3.6.:** Critical values for Bayes factors comparison (see Kass and Raftery, 1995). Comparisons were performed using  $\ln(B_{1,2})$ .

### 3.9.9. Phylogenetic analysis of the different datasets using MrBayes

#### 18S rDNA

For the 18S rDNA dataset phylogenetic analyses were carried out on two levels.

First, an analysis (18S.GTR+I.clade8.\*) was performed on all 18S rDNA sequences, specified in Table 3.4, in addition to the five anguillicolid 18S rDNA sequences. The free-living nematode *Plectus aquatilis* was chosen as outgroup taxon. Under the GTR+I model of DNA evolution, the number of generations of the MCMC runs was set to  $5 \times 10^6$ . Trees were sampled from the MCMC runs every 1,000 generations leading to 5,000 trees, from which 500 were discarded as burn-in.

Those taxa that grouped closest to the Anguillicolidae were subsequently incorporated in the second level of the phylogenetic analysis of the 18S rDNA dataset for hypotheses testing.

Six separate phylogenetic analyses were conducted on the 18S rDNA sequences of *G. neoprocyonis*, *G. turgidum*, *G. binucleatum*, the unidentified nematode SNR-118, *Linstowinema* sp., *E. overstreeti*, *T. tiara* and the five anguillicolid nematodes. *G. neoprocyonis* was selected as outgroup taxon in each of the analyses. In all cases the number of generations was set to  $5 \times 10^6$  and trees were sampled every 1,000 generations, from which 500 were discarded as burn-in. The analyses were performed under the SYM (18S.SYM.\*), GTR+Γ (18S.GTR+G.\*.) and GTR+I+Γ (18S.GTR+I+G.\*.) models of DNA evolution, respectively, and in each case a second analysis under the same model was carried out in which a topological constraint concerning the monophyly of the genus *Anguillicoloides* was enforced (18S.SYM.CA.\*., 18S.GTR+G.CA.\* and 18S.GTR+I+G.CA.\*).

The Bayes factors were calculated and interpreted as specified in section 3.9.7 and 3.9.8.

#### 28S rDNA D2-D3

Six analyses were performed on the 28S rDNA D2-D3 dataset, comprised by nine anguillicolid sequences and the 28S sequences of *T. vitulorum*, *C. spiculigerum* and the unidentified nem-

atode SNR-118. In each case, *T. vitulorum* was selected as outgroup taxon, the number of generations was set to  $7.5 \times 10^6$ , trees were sampled every 1,000 generations and 750 trees were discarded as burn-in.

One analysis was carried out under the HKY+ $\Gamma$  model (28S.HKY+G.\*.) suggested by MrModeltest. In addition, a second analysis was performed under the parametric richer model GTR+I+ $\Gamma$  (28S.GTR+I+G.\*.).

For the remaining four analyses, indels were coded into the dataset (see section 3.9.4) and partitions were defined (see section 3.9.5). Analyses were performed under each model, HKY+ $\Gamma$  and GTR+I+ $\Gamma$ , with and without enforcing a topological constraint concerning the monophyly of the genus *Anguillicoloides* (28S.HKY+G.sic.\*., 28S.GTR+I+G.sic.\*., 28S.HKY+G.sic.CA.\* and 28S.GTR+I+G.sic.CA.\*.)

The Bayes factors were calculated and interpreted as specified in section 3.9.7 and 3.9.8.

### **Cox1**

Since the mitochondrial *Cox1* gene codes for a functional protein, the dataset, containing the sequences of *Strongylida* sp. and *T. cati* in addition to 100 anguillicolid sequences, was partitioned according to the first, second and third codon position.

In all of the seven analyses performed, *Strongylida* sp. was selected as outgroup, the number of generations was set to  $5 \times 10^6$ , trees were sampled every 1,000 generations and the first 300 trees were discarded as burn-in.

For each of the models proposed by MrModeltest, HKY+I+ $\Gamma$  and GTR+I+ $\Gamma$ , one analysis was performed (COX1.HKY+I+G.\* and COX1.GTR+I+G.\*).

Under each model two additional analyses were performed enforcing topological constraints. One concerning the monophyly of the genus *Anguillicoloides* (COX1.HKY+I+G.CA.\* and COX1.GTR+I+G.CA.\*), while the other assumed the species *A. novaezelandiae* to be monophyletic (COX1.HKY+I+G.CN.\*.) and COX1.HKY+I+G.CN.\*).

The Bayes factors were calculated and interpreted as specified in section 3.9.7 and 3.9.8.

Furthermore, an analysis was carried out under the GTR+I+ $\Gamma$  model from which the third codon position was excluded (COX1.GTR+I+G.no3rd.\*).

### **The “supermatrix”**

The “supermatrix”, including sequences of nine clade 8 nematodes in addition to those from 62 anguillicolid specimens, was analysed under the GTR+I+ $\Gamma$  model of DNA evolution. The

number of generations was set to  $10 \times 10^6$ , trees were sampled every 1,000 generations and 1,000 trees were discarded as burn-in.

For the 28S rDNA D2-D3 sequences within the “supermatrix”, indels were coded into the dataset (see section 3.9.4). Partitions were defined as explained in section 3.9.5.

### **3.9.10. Phylogenetic analysis using the maximum likelihood method as implemented in RAxML**

RAxML 7.2.6 (Randomized a(x)ccelerated maximum likelihood for high performance computing) (Stamatakis, 2006) is a program for maximum likelihood (Felsenstein, 1981) based inference of large phylogenetic trees. It was originally derived from fastDNAML, which itself was derived from dnaml which is part of the PHYLIP package (Felsenstein, 2004b).

It enables the user to analyse DNA sequences by conducting a full maximum likelihood analysis combined with a novel rapid bootstrapping (Felsenstein, 1985) algorithm (Stamatakis et al., 2008).

The analysis was performed under the GTR+ $\Gamma$  of DNA evolution on the 18S rDNA dataset, comprising all 18S rDNA sequences, specified in Table 3.4 and the five anguillicolid sequences. *Plectus aquatilis* was assigned as outgroup taxon, a random number seed (12345) was specified for bootstrapping and 1000 rapid bootstrap searches, followed by a full maximum likelihood analysis, were carried out. Bootstrap support values were drawn on the best-scoring maximum-likelihood-tree.

### **3.9.11. Interpretation of branch support**

Since the 1990s, the statistical interpretation of bootstrap support values (BS) has been subject of an intense debate (Felsenstein, 2004a, pp. 335–362). Usually, instead of fixed values, a “rule of thumb” is applied, whereby robust support is indicated by bootstrap values above 70 (Hillis and Bull, 1993).

Similarly, no consensus exists on the statistical interpretation of Bayesian posterior probabilities, although several analyses have shown that the values tend to be significantly higher than corresponding bootstrap values (e.g. Erixon et al., 2003).

In the scope of this thesis, bootstrap values and Bayesian posterior probabilities are interpreted as listed in Table 3.7.

### **3.9.12. Drawing and labelling of phylogenetic trees**

Phylogenetic trees, based on the results of section 3.9.9 and 3.9.10, were drawn, labelled and coloured using FigTree 1.3.1 (Rambaut, 2006–2009).

The tree describing the host-parasite relationships between the genus *Anguilla* and the anguillicolid nematodes was drawn using CoRe-PA 0.3a (Merkle et al., 2009), a software-tool for co-phylogeny reconstruction. Based on the tree and the associated host-parasite-relationships, a tree representing a possible host-parasite co-phylogeny was calculated using specified absolute values (COSPECIATION = -3, SORTING = 1, DUPLICATION = 2 and HOSTSWITCH = 1) and options (HOSTSWITCH = “all host switches are permitted”, SORTING = “direct host switch, no additional sorting costs” and FULLHOSTSWITCH = “full host switches are permitted”).

## 3.10. Analyses of population structure

### 3.10.1. Statistical parsimony haplotype networks

#### Identification of unique sequences and generation of input files for further analyses

In order to quantify the occurrence of each unique sequence (hereafter referred to as “haplotypes” and defined as a unique set of single nucleotide polymorphisms) within each dataset and to simplify further analyses for the user, the Perl program `hap2info.pl` was written.

After specifying the title for the analysis, an input file in FASTA format containing an aligned dataset, an input file containing the sequence prefixes associated with the population (e.g. “MAD Madagascar”), the output-path for the haplotype-list file and the output-path for the Arlequin project file (\*.ARP), the program iterates through the alignment finding all samples sharing the same sequence.

Two output files are created. The haplotype-list file contains the name of each sample followed by its identifier (e.g. “h1”, for the first haplotype found) separated through a tab character. The second output file is written as an Arlequin project file (ARP) consisting of a profile block (general information about data and sample size) and a data block. In the data block the haplotypes are listed with their respective sequences and each population is displayed with the haplotypes sampled within.

BS	BP	Interpretation
< 50	< 0.65	Weak support
50 – 70	0.65 – 0.90	Moderate support
> 70	> 0.90	Robust support

**Table 3.7.:** Interpretations of the support values calculated for branch phylogenetic trees. BS = bootstrap values, BP = Bayesian posterior probabilities.

## TCS

The program TCS 1.2.1 (Clement et al., 2000) implements the estimation of gene genealogies within populations as described by (Templeton et al., 1992). This cladogram estimation is also known as statistical parsimony.

Estimating genealogical relationships among genes at the population level presents a number of difficulties to traditional methods of phylogeny reconstruction, such as parsimony and maximum likelihood, since these methods make assumptions that are invalid at the population level. For instance, these methods assume that ancestral haplotypes are no longer in the population, yet coalescent theory predicts that ancestral haplotypes will be the most frequent sequences sampled from a population (Crandall and Templeton, 1993).

Although network analysis is traditionally used to show genealogical relationships between haplotypes within populations (Templeton et al., 1992), it also provides a useful visual indication of the number of mutations (nucleotide differences) between sequences.

A haplotype network consist of nodes and edges, in which the haplotypes are represented as nodes of a size proportional to their frequency and mutational changes are shown as edges connecting the nodes.

After importing sequences provided in the NEXUS format, the sequences were collapsed into haplotypes, whereby gaps were regarded as a fifth state, and haplotypes were linked if they had a 95% probability of being justified by the parsimony criterion.

The networks were saved in a modified version of the Graph Modelling Language (GML) file format.

## Processing of TCS output

Since the output format of TCS 1.2.1 cannot be read from common graph editor programs, the Perl program `gml2graph.pl` was written.

The program requires the haplotype-list file generated through `hap2info.pl` and the GML file generated through TCS 1.2.1. First, the haplotype-list file is read in order to allow the renaming of the nodes, which TCS 1.2.1 labelled according to the sample name that displayed the haplotype for the first time in the dataset. Then the program parses the GML file for the network, node and edge data and creates a TCS graph (`*.tcs.graph`), in which nodes are represented as circles, instead of ovals.

From this output, another GML file (`*.yed.graph.gml`) is created, which can be read by common graph editors. Additionally, a file (`*.attrib.txt`) is generated, listing each node of the network, its width in pixels and all sample names associated with the haplotype which the node represents.

## Graphical representation of the geographical distribution of the haplotypes

To represent the geographical distribution of the haplotypes graphically, the Perl program `piemaker.pl` was written.

The program was inspired by the two programs `buildPie.pl` and `compositePieIndicators.pl` (Harrington, 2007). It calculates and draws a pie chart for every node specified by the `*.attrib.txt` file created by `gml2graph.pl`, where the pie sectors represent the proportional distribution of the haplotypes among the sampled populations.

For each node, an image file in the Portable Network Graphics (PNG) format proportional to the corresponding size (in pixel) is created. The colours of the pie sectors, represented as hex triplets, are hard-coded into the program.

For running the program a legend file associating sample prefixes with populations separated by the tab character has to be provided. The results have been saved (see appendix A.3).

## yEd

The graph editor yEd<sup>4</sup> makes it possible to colour and apply layouts to networks, generating publishable figures.

After opening the `*.yed.graph.gml` files in yEd, the representation of the nodes was changed by importing the pie charts generated by `piemaker.pl`.

Nodes and edges were manually arranged to provide a clear visual experience. The networks were saved as GRAPHML and PNG files (see appendix A.3).

## Translation of amino-acid sequences from DNA sequences

In order to be able to draw changes in the amino-acid code of *Cox1* on the statistical parsimony networks, DNA sequences were imported into HyPhy Version 2 (Kosakovsky Pond et al., 2005) and translated into the respective amino-acid sequences using the invertebrate mitochondrial genetic code.

The amino acid sequences were verified using the amino acid sequences of *Cox1* from Wielgoss et al. (2008) deposited on GenBank®.

For each of the four species, *A. crassus*, *A. papernai*, *A. australiensis* and *A. novaezelandiae* a character plot of the amino-acid sequences was calculated and saved as a PDF file (see appendix A.3).

---

<sup>4</sup>yEd version: 3.4.2 ©2000–2010 yWorks GmbH. All rights reserved.

### 3.10.2. Calculation of simple population genetic parameters using Arlequin

The Arlequin Project files (\*.arp) generated by `hap2info.pl` were imported into Arlequin 3.5.1.2 (Excoffier and Lischer, 2010) and analyses were performed on each gene following the protocol of Drown (2004).

Molecular diversity indices, such as the number of transitions and transversions, the number of haplotypes and the amount of indels, were calculated within groups (i.e. species or populations) and between groups (i.e. populations), in order to be able to interpret the results from the statistical parsimony networks.

Furthermore, an AMOVA (Analysis of molecular variance) (Weir and Cockerham, 1984; Excoffier et al., 1992; Weir, 1996) was conducted on each of the *Cox1* datasets of the anguillicoloid species and  $\phi_{st}$  - values, the proportion of nucleotide diversity (measured as the expected squared evolutionary distance between alleles) due to differences among populations, was calculated.

The parameter  $\phi_{st}$  is analogous to the fixation index  $F_{st}$ , which measures the extent of genetic differentiation among sub-populations.

$\phi_{st}$ - values	Interpretation
0 – 0.05	negligible genetic differentiation
0.05 – 0.15	moderate genetic differentiation
0.15 – 0.25	strong genetic differentiation
> 0.25	very strong genetic differentiation
1	complete isolation

**Table 3.8.:** Ranges of values used for the interpretation of  $\phi_{st}$  (Wright, 1978).

The values are always positive and range from 0 = panmixis (no subdivision) to 1 = complete isolation. Negative values can occur in the absence of population structure, due to calculation artefacts, and are treated as 0. The ranges of the values were interpreted following the suggestions of Wright (1978) and are listed, together with their implications in Table 3.8.

The  $\phi_{st}$  - values were calculated using 10,000 permutations, at an significance level of 0.05. Additionally, the average pair-wise differences, based on the nucleotide diversity  $\pi$  (Nei and Li, 1979), within and between populations, both corrected and uncorrected, were calculated.

All output files from the Arlequin-Analyses were saved (see appendix A.4).

## 4. Results and Discussion

### 4.1. Laboratory results

The protocol for the tissue extraction was developed through an initial test analysis. It involved the extraction of 8 tissue samples from sample TCU-175 (*Anguillicoloides crassus*, differing in amount and host contamination, and the subsequent lysis, DNA amplification (18S rDNA, 28S rDNA D2-D3 and *Cox1*), gel electrophoresis and sequencing in order to visualise the importance of both factors.

A second tissue extraction protocol was designed to extract, amplify and sequence DNA from nematodes encapsulated in the intestinal wall of Japanese eel (*Anguilla japonica*) specimens in order to diagnose an infection with *A. crassus* using the 28S rDNA D2-D3 barcoding region. The mechanical destruction of the collagen sheath of the capsules proved to be crucial for the diagnosis. These results have been published (Heitlinger et al., 2009).

Gel electrophoresis showed that bands of unspecific fragments occur only in samples bearing higher levels of host contamination. However, host contamination does not influence the results of all of the sequencing reactions in the same manner. Primers for the 18S rDNA and the *Cox1* gene were much more susceptible for binding to host DNA than the primers for the 28S rDNA D2-D3 and the ITS1 region.

The amount of samples in which actinopterygic DNA was sequenced is given for each of the primer pairs in Table 4.1.

Primer pair of	Host DNA count
18S rDNA	10
ITS1	0
28S rDNA D2-D3	2
ITS2	(2)
<i>Cox1</i>	3

**Table 4.1.:** Number of samples in which actinopterygic DNA was sequenced (“Host DNA count”) for each of the primer pairs, based on BLAST searches. Numbers in parenthesis represent non-nematode sequences.

For eight samples JPN-K4B2 (*Anguillicoloides globiceps*), JPN-K4B3 (*A. globiceps*), JPN-W4A1 (*A. globiceps*), JPN-W4B1 (*A. globiceps*), JPN-K4A1 (*A. globiceps*), JPN-K4A2 (*A. globiceps*),

SSD-107b (*Anguillicoloides papernai*), JPN-WK6 (*A. crassus*) no *Cox1* sequence was obtained under the standard protocol.

However, after TOPO cloning of the *Cox1* PCR product from each sample, followed by DNA extraction, gel electrophoresis and gel extraction, it was possible to sequence the *Cox1* gene of JPN-W4B1 and JPN-WK6.

#### **4.2. Analysis of the 18S ribosomal DNA (18S rDNA) dataset**

After sequencing and bioinformatic analysis, 127 samples displayed a high-quality-sequence of the amplified partial 18S rDNA fragment. Within the samples, five haplotypes, each being characteristic to one of the anguillicolid species, were found. The sequence lengths range from 682 to 865 bases. The largest sequence for each species, shown in Table 4.2, was chosen as type sequence and will be stored on GenBank® in the near future.

Species	Max. Length (bp)	GC (%)
<i>Anguillicoloides crassus</i>	868	46.889
<i>Anguillicoloides papernai</i>	868	47.005
<i>Anguillicola globiceps</i>	836	47.235
<i>Anguillicoloides novaezelandiae</i>	865	47.177
<i>Anguillicoloides australiensis</i>	865	47.052

**Table 4.2.:** Partial 18S rDNA type sequences of the Anguillicolidae

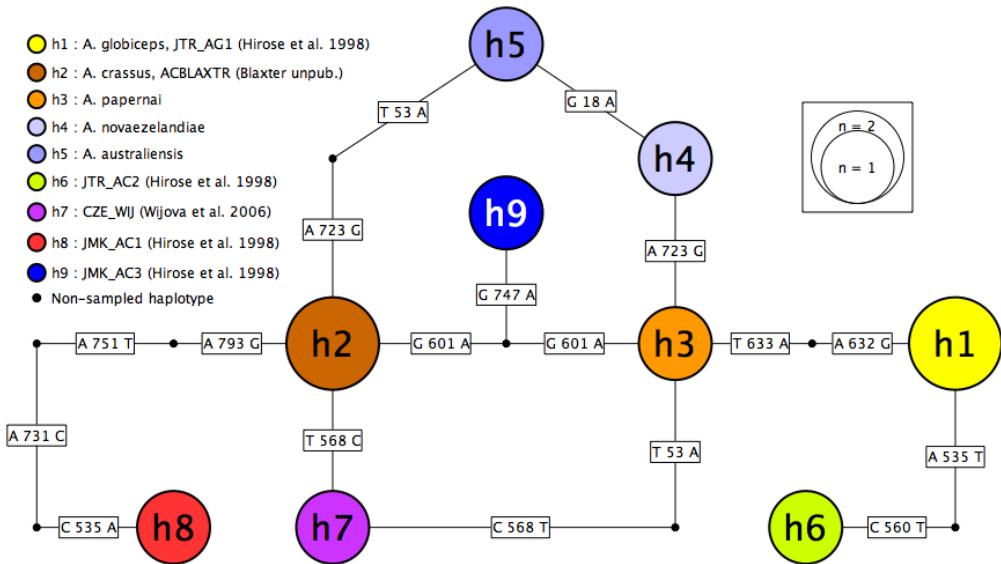
The closest hits for the *A. crassus* sequence revealed by BLAST search referred to the following GenBank® entries of 18S rDNA sequences of *A. crassus*: DQ490223.1 (Wijová et al., 2006), DQ118535.1 (Blaxter and Roche, 2005), AB009317.1 and AB009315.1 (Hirose et al., 1998).

Results of the BLAST search for *A. globiceps* indicated the highest percentage identity to the nucleotide sequences of AB009318.1 (JTR-AG1, *A. globiceps*) and AB009316.1 (JMK-AC2, *A. crassus*) (Hirose et al., 1998).

The relationships between the 18S rDNA sequences from this study and their closest BLAST hits are shown in the statistical parsimony network in Figure 4.1.

As in all statistical parsimony networks calculated within the scope of this study, nodes represent unique haplotypes and are coloured according to the population they were sampled from. Each edge connecting two nodes represents the mutational change that lead from one haplotype (node) to another. The label on the edge indicates the position and nature of the mutational change.

Although the differences among the type sequences are small, the observed sequence uniformity at the species level, the total length of the type-sequences and the sample size of each



**Figure 4.1.:** Statistical parsimony network constructed using all available anguillicolid 18S rDNA sequences

of the species can be used as an argument to re-evaluate some of the 18S rDNA entries on GenBank® of the Anguillicolidae.

Hirose et al. (1998) deposited the first set of anguillicolid sequences on GenBank®. 18S rDNA from three *A. crassus* specimens and one *A. globiceps* specimen were amplified using the primer pairs 18S-1/18S-2 (Ellis et al., 1986) with the synthesised cDNA of *A. crassus* and the genomic DNA of *A. crassus* and *A. globiceps* as templates. The amplified fragments were cloned into two plasmid vectors and used for the transformation of *E. coli* cells. For each specimen, four clones were sequenced and compared to ensure sequence agreement.

Hirose et al. (1998) explains the high percentage sequence identity between JMK-AG1 and JMK-AC2 as an example of the limit of 18S rDNA to discriminate against closely related species within nematode genera. However, apart from the use of quadruplicates no sequence quality assessment was mentioned which opens the possibility that sequencing errors may account for the differences between the type sequence of *A. crassus* of this study and JMK-AC1 and JMK-AC3 from Hirose et al. (1998).

Furthermore JMK-AC2 appears to have a higher sequence identity with the *A. globiceps* type sequence, which is identical to JTR-AG1. This suggests either a morphological misidentification of the specimen JMK-AC2 as *A. crassus* in addition to sequencing errors or a correct identification of the specimen and a considerable number of sequencing errors.

The *A. crassus* 18S rDNA sequence provided by Wijová et al. (2006) displays only one substitution when compared to the corresponding type-sequence, which may be caused by a poor base call at the position in question.

The sequence of *A. crassus* by Blaxter and Roche (2005) is identical to the *A. crassus* type-sequence from this study.

With regard to the three additional clade 8 nematodes *sensu* Holterman et al. (2006) sequenced, the BLAST searches yielded useful information for species identification in all of the three cases.

The specimen identified morphologically as *Camallanus lacustris* proved to be identical (100% identity) to DQ442663, the 18S rDNA sequence of *Camallanus lacustris* sequenced by Wijová et al. (2006).

Similarly, for MAD-19, a nematode isolated from *A. mossambica* in Madagascar, the best BLAST results, namely *Ascarophis arctica* DQ094172 (97 PID) and *Neoascarophis macrouri* DQ442660 (97 PID), suggest that this nematode belongs to the family Cystidicolidae Skrjabin, 1946, which is known to parasitise in the swim bladder of several physostomous fish. Based on the sampling location and the distribution of members of the Cysticolidae, one could hypothesise that this specimen belongs either to the genera *Ascarophis* (marine, worldwide), *Spinitectus* (fresh-water, worldwide) or *Pseudoproleptus* (fresh-water, Africa) (Chabaud and Bain, 1994).

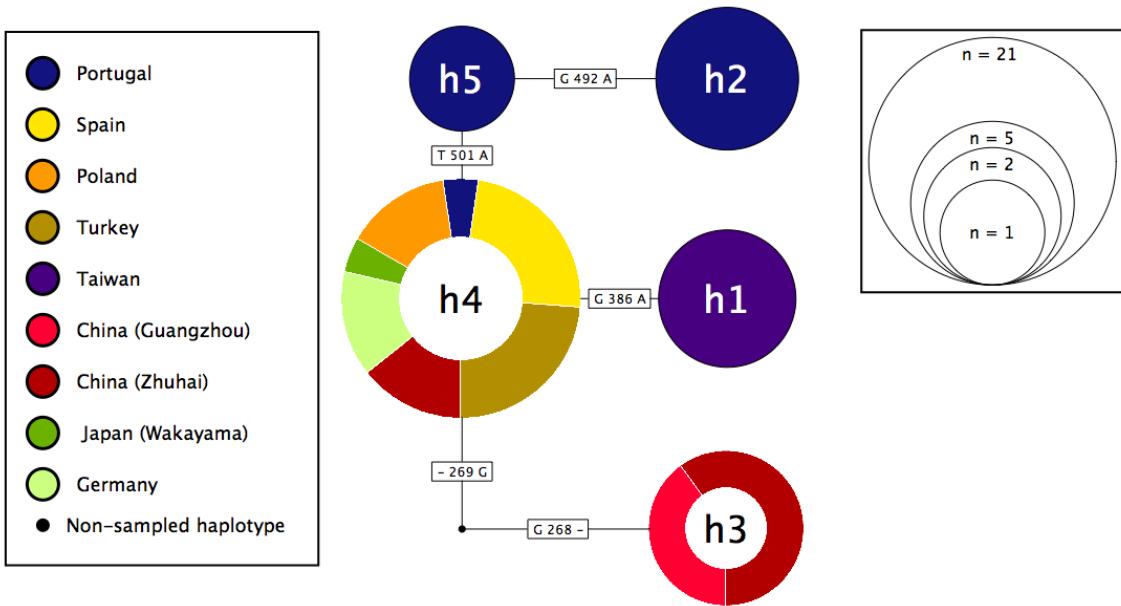
In the case of SNR-118, no conclusive information can be gained from searching against the nucleotide database on GenBank®. However, by comparing the sequence to the 18S rDNA dataset provided by Nadler (2010), SNR-118 showed a 99% identity to both *Tanqua tiara* (Gnathostomatinae Railliet, 1895) and *Linstowinema* sp. (Echinonematinae Inglis, 1967), which is puzzling since *Tanqua* is known from reptiles with a Gondwanan distribution (Chabaud and Bain, 1994) and *Linstowinema* sp. has so far only been documented from dasyurid marsupials and bandicoots (Peramelidae) in Australia (Smales, 2000).

#### **4.3. Analysis of the internal transcribed spacer I (ITS1) dataset**

For 32 out of 56 specimens of *A. crassus* a high quality sequence of the ITS1 region, with an average length of 543 bp, was obtained.

The protocol for amplification and sequencing of this region was applied in order to reveal possible cryptic species within *A. crassus*, since this species has the widest geographical distribution (see section 1.3.1) and is known to parasitise the greatest amount of host species (see section 1.3.2).

Unfortunately, no evidence was found supporting this hypothesis, since very few differences among the five haplotypes, namely two transversions, one transition and two sites with indels, could be detected. Because of the few differences, no phylogenetic analysis was performed on this dataset (see also section 4.7), however, it was included in the “supermatrix”. The observed mutational changes leading from one haplotype to another are depicted in the statistical parsimony network in Figure 4.2.



**Figure 4.2.:** Statistical parsimony network based on the ITS1 sequences of 32 specimens of *A. crassus*

The population of *A. crassus* in which the greatest amount of haplotypic diversity (3) was found is that from Portugal, from which two are restricted only to this population.

The other population showing an unique haplotype was the one from Taiwan, which is counterintuitive since this country was pointed out as the most likely source of introduction of *A. crassus* into Europe (Køie, 1991; Wielgoss et al., 2008). However, only one transition leads from the Taiwanese haplotype to the the haplotype "h4", which is found in all European populations. Therefor it should not be viewed as evidence against the theory of European colonisation of *A. crassus*.

The Japanese and Chinese (Zhuhai) specimens share the haplotype "h4" with all sampled populations from Europe except, with the two haplotypes limited to Portugal.

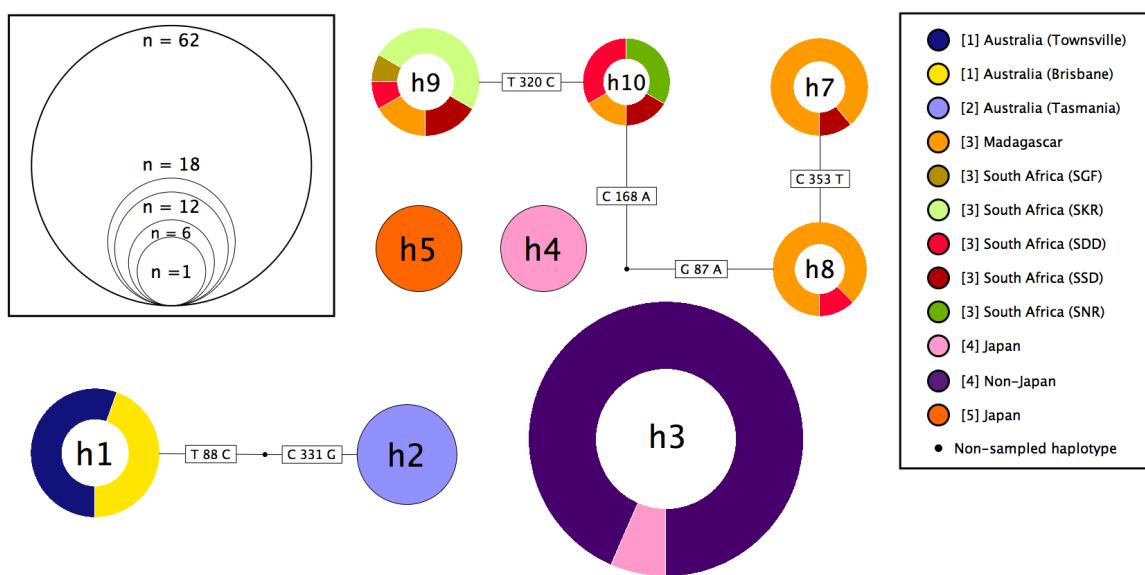
Interestingly the Chinese haplotype "h3" shows a deletion of 2 contiguous bases in comparison to the other haplotypes. This may have been caused by replication slippage, a process responsible for the deletion or insertion of repeating units during replication (Lewin, 2008, p. 125).

#### 4.4. Analysis of the 28S ribosomal DNA D2-D3 (28S rDNA D2-D3) dataset

This barcoding gene showed not only the best capacity to make distinctions at the species level, since private haplotypes are found for each of the five anguillicolid species (see Heitlinger et al., 2009), but turns out to be the least susceptible to host contamination.

#### 4. Results and Discussion

After bioinformatic analysis, 136 high quality sequences were obtained. In Figure 4.3 the statistical parsimony network of all haplotypes is shown.



**Figure 4.3.:** Statistical parsimony network based on the 28S rDNA D2-D3 sequences of 136 anguillicolid specimens. The numbers in square brackets in the caption indicate the species: [1] = *A. australiensis*, [2] = *A. novaezelandiae*, [3] = *A. papernai*, [4] = *A. crassus*, [5] = *A. globiceps*.

In total, 53 polymorphic sites, consisting in 29 transitions, 14 transversions and 18 indels, could be identified among the sampled specimens. Due to the large amount of differences between the individual species, their respective nodes are not interconnected, except for the case of the two Oceanian species *A. australiensis* and *A. novaezelandiae* which are connected by two mutational events indicating a recent divergence of the two species.

The main reason most nodes are not connected is a region of indels, which is depicted in Figure 4.4. This region is of great importance since it allows the design of species-specific PCR primers.

In the case of *A. crassus*, six samples of the Japanese population, represented by the haplotype “h4” in the network, show an indel of six bases at position 312 of the alignment, followed by a duplication of the four bases preceding the indel. This indel is unique to this six Japanese

<i>A. australiensis</i>	A	T	G	C	G	T	G	T	T	A	T	G	T	T	-	-	-	-	-	G	T	T	G	T	T	A	T	A	C	G	C	C	T	C	G	T	G	T	T		
<i>A. novaezelandiae</i>	A	T	G	C	G	T	G	T	T	A	T	G	T	T	-	-	-	-	-	G	T	T	G	T	T	A	T	A	C	G	C	C	T	C	G	T	G	T	T		
<i>A. papernai</i>	A	T	G	C	G	T	-	-	-	-	-	-	-	-	-	-	-	-	-	T	G	T	G	T	T	A	T	A	C	G	C	C	T	C	G	T	G	T	T		
<i>A. crassus "h5"</i>	A	T	G	C	G	T	G	T	T	A	T	G	T	T	G	A	G	T	T	A	T	-	-	-	-	A	T	A	T	A	C	G	C	C	T	C	G	T	G	T	T
<i>A. crassus "h3"</i>	A	T	G	C	G	T	G	T	T	A	T	-	-	-	-	-	-	-	-	A	T	A	T	A	T	A	C	G	C	C	T	C	G	T	G	T	T				
<i>A. globiceps</i>	A	T	G	C	G	T	-	-	-	-	-	-	-	-	T	G	T	G	T	A	T	A	G	G	C	C	T	C	G	T	G	T	T								

**Figure 4.4.:** The alignment of the 28S rDNA D2-D3 dataset starting at position 301 for each of the anguillicolid species showing variation in this segment.

samples and, apart from three transitions (at position 574, 649 and 689), the only polymorphisms that distinguish them from all other specimens of this species.

The sequences from *A. papernai* show the greatest number of haplotypes. Interestingly, all four haplotypes can be found within the population from Madagascar, which opens the hypothesis that the South African populations might have originated in Madagascar. Haplotype "h7" and "h8" were both found in only one South African nematode, respectively. This could be a result from an colonisation process of *A. papernai* that started upstream (SSD) in the Sunday's River spreading downstream (SDD), from which this species then colonised other river systems in South Africa.

Another interesting fact, involving the sequences from *A. papernai*, is that, in five cases several specimens were collected from within the same host specimen and only in three of the five cases the nematode specimens display the same haplotype (MAD-27-1, MAD-27-2, MAD-27-4, MAD-27-7 in haplotype "h7", SKR-31, SKR-31a, SKR-31b in haplotype "h8" and MAD-36-2, MAD-36-3 in haplotype "h10"). The sequences from SSD-107a and SSD-107b, however, share the haplotype "h8", while SSD-107 shows the haplotype "h9"; and the specimens MAD-44-1 and MAD-44-2 display the haplotypes "h10" and "h9", respectively. This proves that *A. papernai* displaying multiple haplotypes of 28S rDNA D2-D3 can parasitise within the same host specimen.

Compared to the 18S rDNA dataset, the search for outgroup taxa for the phylogenetic analysis of this dataset proved to be less fruitful, as for only very few representatives of the clade 8 nematodes this region has been sequenced. BLAST searches yield among the best hits the Anisakidae *Contraeacum spiculigerum* (AB189984) with ~85% identity and the toxocarid nematode *Toxocara vitulorum* (FJ418790) with ~84% identity. Although the partial sequence of the 28S rDNA region of several species of the Gnathostomatidae are available, these differ too much from the Anguillicolidae to include them in the analyses.

Additionally, sequences were obtained for *C. lacustris*, SNR-118 and MAD-19. However, as with the anguillicolid sequences, BLAST searches yielded no close hits. SNR-118 shares 86% identity with the rhabditid *Chiloplacus* sp. (DQ145634.1) as well as with the anisakid *Pseudoterranova decipiens* (AY821763), while MAD-19 displays 88% identity with *T. vitulorum* (FJ418790.1). A BAST search for the sequence of SNR-118 against the anguillicolid sequences revealed only ~75% identity, but showed a greater coverage than other sequences since the same primer pairs were used. For this reason, SNR-118 was included as an additional outgroup taxon in the phylogenetic analysis of the 28S rDNA D2-D3 sequences of the Anguillicolidae.

#### 4.5. Analysis of the internal transcribed spacer 2 (ITS2) dataset

For 33 out of 56 specimens of *A. crassus* an ITS2 sequence was obtained.

Unfortunately, due to the low quality base calls during sequencing and the subsequent quality check implemented in the bioinformatic analysis, the observed differences in the dataset are mainly caused by bases coded as ambiguous letters (IUPAC nucleic acid code).

Hence, this dataset was excluded from population structure and phylogenetic analyses.

#### 4.6. Analyses of the Cytochrome oxidase I (Cox1) dataset

As explained in section 4.1, certain problems were encountered while sequencing *Cox1*.

All specimens of *A. globiceps* could not be sequenced using the standard protocol. Finally, one specimen (JPN-W4B1) yielded a high quality *Cox1* sequence, while the others yielded sequences containing several insertions of one and two bases. The base calls of the insertions have been checked by eye and no inferior quality was observed.

Although insertions within protein coding genes are not uncommon, the fact, that these insertions are not multiples of three, is puzzling, since this would imply the disruption of the reading frame during translation. This problem has been reported previously (S Wielgoss, pers. comm. 08.12.08) and possible explanations are the existence of pseudogenes, in the form of non-functional mitochondrial DNA inserted within the nuclear genome, and a molecular process called RNA editing, in which case the messenger RNA would be altered prior to translation and the gene sequence must not directly reflect the amino-acid sequence of the protein. Although the latter seems unlikely, it has been shown to occur in the nematode *Teratocephalus lirellus* (Vanfleteren and Vierstraete, 1999), whose sequence of the cytochrome *b* gene displays a similar pattern as the one observed in *Cox1* of *A. globiceps*. Since for *A. globiceps* this gene could only be sequenced once, no statistical parsimony network was constructed for this species.

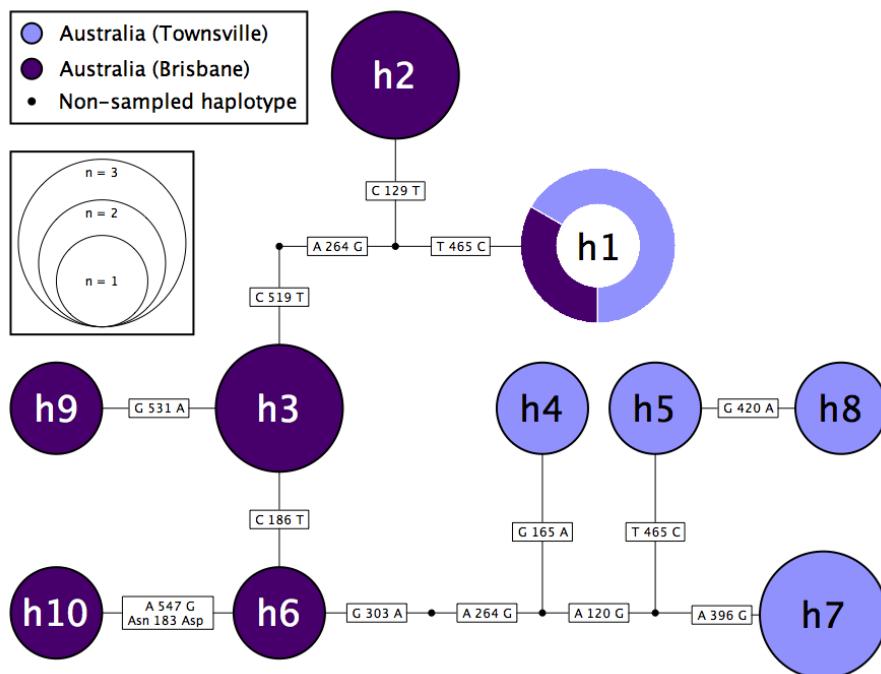
For the anguillicoloid species, 99 high quality sequences with a length of 532 bp were obtained. Additionally, *Cox1* was successfully sequenced from SNR118, which displayed 84% identity with the corresponding gene in both the mitochondrial genomes of *Ascaris suum* (X54253.1) and *Toxocara malayensis* (AM412316.1).

As mentioned in section 1.5, one of the great advantages of this molecular marker is its variability, which was confirmed by the diversity of haplotypes observed within the genus *Anguillicoloides*. For this reason, the statistical parsimony networks are depicted separately for *A. australiensis* (section 4.6.1), *A. novaezelandiae* (section 4.6.2), *A. papernai* (section 4.6.3) and *A. crassus* (section 4.6.4).

Since the *Cox1* gene codes for a functional protein, both synonymous substitutions and non-synonymous substitutions, leading to changes in the amino acid sequence, are specified on

the labels of the edges. The information for the latter was obtained from the character plots of the amino-acid sequences, which have been calculated using HyPhy (see section 3.10.1).

#### 4.6.1. Analysis of the *CoxI* sequences from *A. australiensis*



**Figure 4.5.:** Statistical parsimony network of the *CoxI* sequences from *A. australiensis*

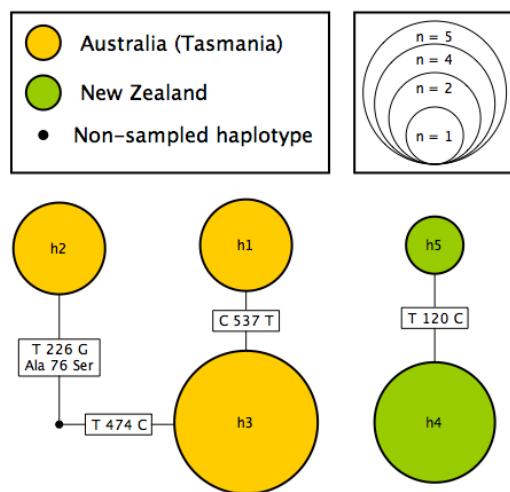
The analysis of the 15 sequences of *A. australiensis* from the two populations within Australia reveals 10 haplotypes due to 12 transversions. Almost all haplotypes are either confined to the population from Townsville or to that from Brisbane. The only one shared between both populations is haplotype “h1”, which in the statistical parsimony network in Figure 4.5 is shown as being more closely connected to the population from Brisbane.

The calculation of the pair-wise  $\phi_{st}$  - value resulted in 0.3177 (P-value =  $0.00396 \pm 0.0006$ ) which indicates a strong genetic differentiation among the two populations of *A. australiensis* in Townsville and Brisbane.

#### 4.6.2. Analysis of the *CoxI* sequences from *A. novaezelandiae*

Geiß and Sures (2010a) provided forward (LCO1490) and reverse (HC02198) sequences of the *CoxI* gene for six specimens of *A. novaezelandiae*. These specimens were part of a laboratory population and descendants from two gravid females collected in New Zealand.

During bioinformatic analysis of the sequences, five of the consensus sequences showed high quality base pairs and have been included in this dataset, as well as in the “supermatrix”, increasing the total sample size for this species to 14. Among the 14 samples of *A. novaezelandiae* five haplotypes were found and are displayed in Figure 4.6.



**Figure 4.6.:** Statistical parsimony network of the *CoxI* sequences from *A. novaezelandiae*

The nodes representing the haplotypes found in Tasmania and the ones found in New Zealand are not connected in the network.

This is due to the large number of differences between the haplotypes of these two populations. While the number of mutational changes within the New Zealand and the Tasmanian populations are one (a transition) and three (one transversion and two transitions), respectively, the detected amount of substitutions within the whole dataset is 64, with 44 transitions and 20 transversions, at 63 substitution sites, which leads to an corrected average pair-wise difference between the two populations ( $\pi_{NZ,TAS} - \frac{(\pi_{NZ} + \pi_{TAS})}{2}$ ) of  $61.94 \pm 0.25$  (see section 3.10.2).

Furthermore, the calculations related to population structure resulted in an  $\phi_{st}$  - value of 0.9828 (P-value  $0.00020 \pm 0.00014$ ), which indicates that the two populations are completely isolated from each other.

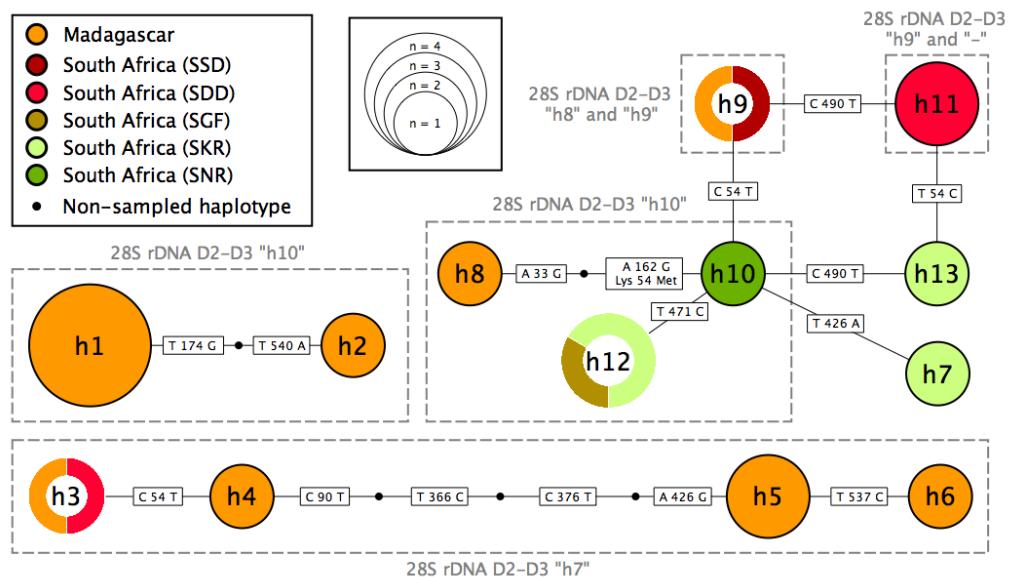
The comparison of the sequences from the two populations using BLAST reveals 88% identity, which are comparable to values gained when comparing both populations of *A. novaezelandiae* to *A. crassus* (see section 4.6.4).

Therefore, based on the results from population structure analysis and sequence comparison using BLAST, the hypothesis is posed that *A. novaezelandiae*, although displaying a homogeneous morphology, is likely to include one cryptic species.

#### 4.6.3. Analysis of the *CoxI* sequences from *A. papernai*

From the 22 *CoxI* sequences for *A. papernai* that were successfully sequenced, twelve belong to specimens sampled in Madagascar while the remaining ten correspond to South African populations. The relationships between the 13 haplotypes identified among the sequences are shown in the statistical parsimony network in Figure 4.7.

Keeping the statistical parsimony network of the 28S rDNA D2-D3 barcoding region (Figure 4.3) in mind, the corresponding haplotypes are drawn onto this figure.



**Figure 4.7.:** Statistical parsimony network of the *CoxI* sequences from *A. papernai*. The haplotypes of the 28S rDNA D2-D3 region are drawn as dashed rectangles around the respective *CoxI* haplotypes. For the specimens SKR-27a (“h7”), SDD-54 (“h11”) and SKR-41 (“h14”) no 28S rDNA D2-D3 sequences have been obtained.

Similar to the samples from *A. novaezelandiae*, the differences among some of the haplotypes were too high to be justified by the parsimony criterion and therefore three individual haplotype networks were generated.

The fact that the three *A. papernai* networks almost completely reflect the pattern found while examining the 28S rDNA D2-D3 haplotype network, is interesting and does not contradict the hypothesis, that colonisation of South Africa originated from Madagascar.

Compared to the molecular diversity indices of the sequences from *A. novaezelandiae*, the ones from *A. papernai* are slightly higher with 79 substitutions, namely 54 transitions and 25 transversions, at 74 sites. The molecular diversity indices of *A. papernai* at the intra-population level, although in part visible in the haplotype network, are listed in Table 4.3.

Population	MAD	SNR	SGF	SKR	SDD	SSD
Sequences	12	1	1	4	3	1
Haplotypes	8	1	1	3	2	1
Transitions	53 (50)	0	0	2 (2)	30 (30)	0
Transversions	25 (25)	0	0	1 (1)	17 (17)	0
Substitutions	78 (73)	0	0	3 (3)	47 (47)	0
Private substitution sites	25	0	0	0	0	0
Nucleotide diversity $\pi$	31.121	0	0	1.667	31.333	0

**Table 4.3.:** Molecular diversity indices calculated using Arlequin for the populations of *A. papernai*. Numbers in brackets refer to the number of sites of which substitutions are observed. The order of the populations was according to their distance from Madagascar and abbreviated according to Table 3.1: MAD = Madagascar, SNR = Nahoon River, SGF = Great Fish River, SKR = Koonap River, SDD = Sunday's River (Darlington Dam), SSD = Sunday's River (Slagboom Dam).

What immediately becomes apparent from the table as well as from the statistical parsimony networks is that all South African populations are likely to be subsets of that from Madagascar, since it is the only population displaying private haplotypes. However, the lack of comparable sampling sizes for the South African population makes it impossible to draw further conclusions. Table 4.4 lists the pair-wise  $\phi_{st}$  - values for the populations of *A. papernai*.

	MAD	SDD	SGF	SKR	SNR	SSD
<b>MAD</b>	-					
<b>SDD</b>	0.1544	-				
<b>SGF</b>	0.2734	(-)	-			
<b>SKR</b>	0.4835 *	0.1847 *	(-)	-		
<b>SNR</b>	0.2561	(-)	1.0000	(-)	-	
<b>SSD</b>	0.2677	(-)	1.0000	0.1667	1.000	-

**Table 4.4.:** Table of the pairwise  $\phi_{st}$  - values calculated using Arlequin for the populations of *A. papernai*.

- = 0.000, (-) = negative  $\phi_{st}$  - values, \* = significant P-values.

Only two  $\phi_{st}$  - values were statistically significant. According to the calculations, a very strong population divergence can be observed between the populations from Madagascar and that from Koonap River (SKR) ( $\phi_{st} = 0.4835$ ).

In contrast, only minor genetic differentiation ( $\phi_{st} = 0.1847$ ) can be detected between Koonap River (SKR) and the Great Fish River (SGF), although both sampling sites are ~75 kilometres apart.

However, these estimates should be considered with care due to the small sampling sizes within the South African populations.

#### 4.6.4. Analysis of the *CoxI* sequences from *A. crassus*

The number of *CoxI* sequences of *A. crassus* used for the analyses exceeds the number of sequences from the other anguillicolid species by far. This is in part because of the 48 high quality sequences obtained from eleven populations during this study. Additionally, 419 *CoxI* sequences provided by Wielgoss et al. (2008) and thirteen sequences supplied by Geiß and Sures (2010a) have been included in the dataset, increasing the sample size to 480 from 28 different populations.

Since Paggi et al. (1982) reported *A. novaezelandiae* from Lake Bracchiano, the consensus among studies of the Anguillicolidae has been that this population has either gone extinct or at least been assimilated, due to the introduction of *A. crassus* in this area.

Geiß and Sures (2010b) compared the sequences of *CoxI* of specimens of *A. novaezelandiae* from New Zealand to the anguillicolid nematodes of the population found in Lake Bracchiano (Italy) and found a very low resemblance, i.e. 88% identity using BLAST.

By comparing them to the specimens of *A. novaezelandiae* from Tasmania form this study a similar picture can be seen, only 87% sequence identity can be observed. In contrast, a BLAST search against *CoxI* sequences from *A. crassus* reveals 97 - 99% identity. Hence, the specimens from Lake Bracchiano definitely belong to *A. crassus*.

In order to display the *CoxI* haplotypes of *A. crassus* in a statistical parsimony network, the sampling sites were grouped according to Wielgoss et al. (2008). The groupings are listed in Table 4.5.

Group	Prefix	Population	No. of sequences		Source
			In population	In group	
North-Eastern Europe	ALA	Åland Islands (Finland)	16		Wielgoss et al. (2008)
	OER	Kullen, Øresund / Kattegat (Sweden)	30		Wielgoss et al. (2008)
	COR	Slapton Ley, Cornwall (Great Britain)	15		Wielgoss et al. (2008)
	NEA	Lake Neagh (Great Britain)	31		Wielgoss et al. (2008)
	SHA	Lough Deragh, Shannon (Ireland)	30	135	Wielgoss et al. (2008)
	GRA	Rußheimer Altrhein (Germany)	4		present study
	GST	Steinfeld (Germany)	1		present study
	POL	Sniardwy Lake, Mikolajki (Poland)	5		present study
Brittany	C	Essen (Germany)	3		Geiß and Sures (2010a)
	FRE	Bois Joli, Frémur (France)	31	61	Wielgoss et al. (2008)
South-West Europe	VIL	Brain-sur-Vilaine (France)	30		Wielgoss et al. (2008)
	LOI	Angers, Loire (France)	32		Wielgoss et al. (2008)
	RHO	Camargue, Rhône	30		Wielgoss et al. (2008)
	ORI	Oria (Spain)	30		Wielgoss et al. (2008)
	EAV	Albufera de Valencia (Spain)	3	136	present study
	POR	Ribeira das Lampreias (Portugal)	1		present study
	TIB	Roma, Tiber (Italy)	30		Wielgoss et al. (2008)
	LB	Lake Bracchiano (Italy)	10		Geiß and Sures (2010a)
Turkey	TUR	Asi River, Hatay (Turkey)	4	4	present study
USA	STJ	St. Jones River (New Jersey, USA)	32	32	Wielgoss et al. (2008)
Taiwan	KAO	Tung-Chiang, Kao-Ping River (Taiwan)	46		Wielgoss et al. (2008)
	TKR	Sinyuan, Kaoping River (Taiwan)	4	55	present study
	TCU	Eel culturing pond, Budai, (Taiwan)	5		present study

continued on next page

Group	Prefix	Population	No. of sequences		Source
			In population	In group	
China (Zhuhai)	CGZ	Zhuhai, Guangdong (China)	14	<b>14</b>	present study
China (Guangzhou)	CGG	Guangzhou, Guangdong (China)	2	<b>2</b>	present study
Japan (Mikawa)	MIK	Mikawa Bay (Japan)	29	<b>29</b>	Wielgoss et al. (2008)
Japan (Yamaguchi)	YAM	Yamaguchi, Fushino (Japan)	7	<b>7</b>	Wielgoss et al. (2008)
Japan (Wakayama)	JPN	Natural water system, Wakayama (Japan)	5	<b>5</b>	present study

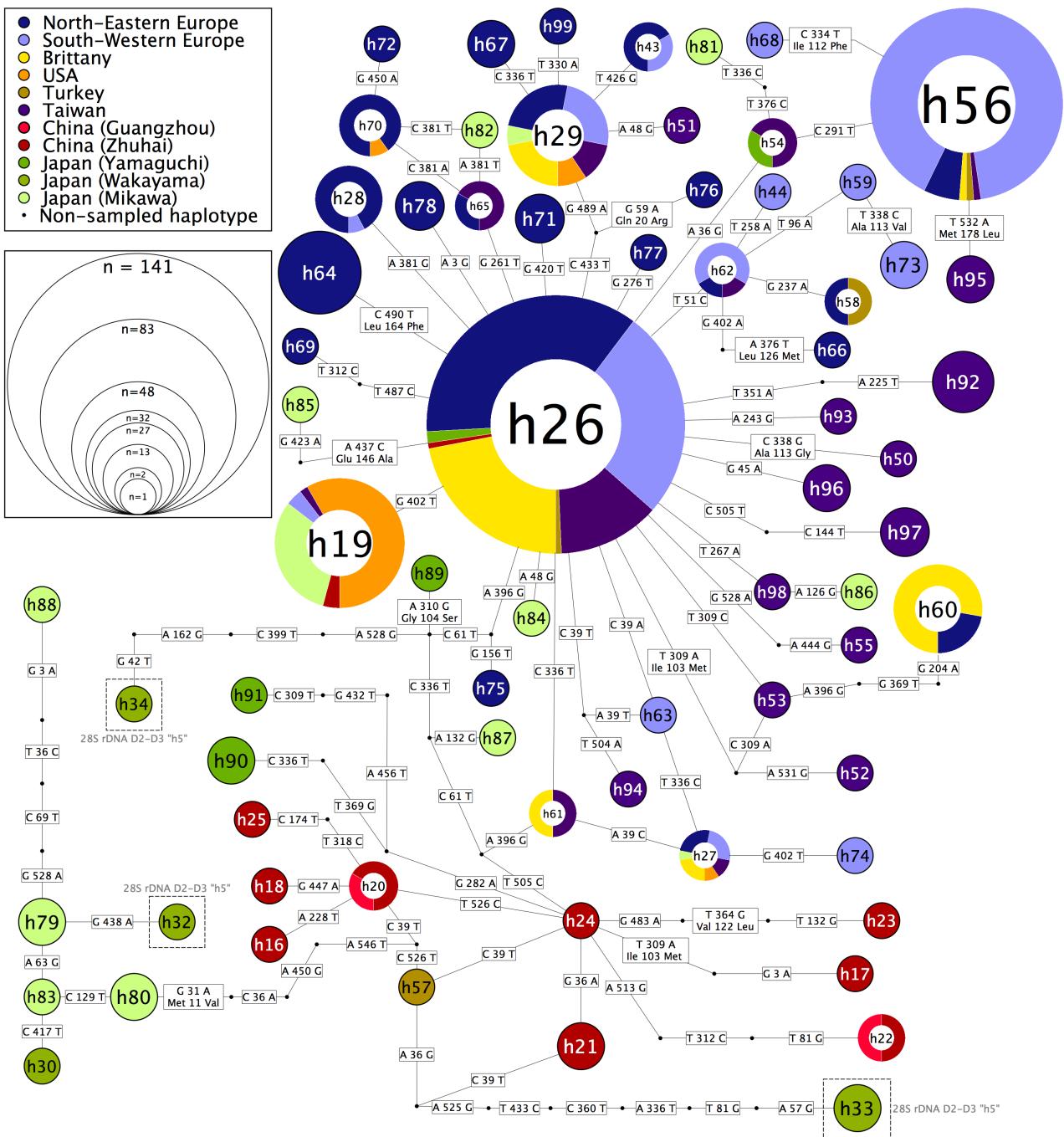
**Table 4.5.:** Table of the geographical grouping applied to the populations of *A. crassus* prior to the calculation of the statistical parsimony network. Grouping was performed according to Wielgoss et al. (2008).

The statistical parsimony network of all *Cox1* sequences of *A. crassus* is depicted in Figure 4.8. A complete discussion of this network and of the results of the calculations of the population parameters would go beyond the scope of this diploma thesis. Nevertheless, some features of interest, will be discussed.

Among the 480 sequences, 71 haplotypes are found which display 75 polymorphic sites in which 53 transitions and 31 transversions are observed. This leads to an average nucleotide diversity  $\pi$  of  $0.005366 \pm 0.003112$ .

It should be mentioned that given the number of haplotypes, this statistical parsimony network is probably just one of several parsimonious solutions. The fact that all haplotypes from Taiwan are nested between European populations supports the fact that the origin of colonisation of the European continent by this species was Taiwan. Furthermore, almost all European haplotypes are connected to another European haplotype by no more than two mutational changes. The only exceptions are the haplotype “h60”, displayed by specimens from Lake Neagh (Great Britain), Bois Joli and Brain-sur-Vilaine (France); and haplotype “h57” which is found in only one specimen from Turkey. Japanese samples from Yamaguchi and Mikawa, as well as Chinese samples from Zhuhai can be found close to European haplotypes as well as very distantly connected (lower part of the network). However, all Japanese samples from Wakayama and all Chinese samples from Guangzhou, are located in the lower part of the network, being connected to Taiwanese and European samples through a high amount of mutational steps.

The nematodes from Wakayama also revealed the highest number of private substitutions, namely eight, as well as the highest nucleotide diversity ( $\pi = 11.6$ ) compared to all other populations, despite the fact that only five specimens were sequenced. Additionally, the pair-wise comparison of the population  $\phi_{st}$  values revealed, in almost all cases, significant genetic divergence between the Japanese population from Wakayama from all other populations, ranging from 0.2683 and 0.2936 for the other two Japanese populations (Yamaguchi and Mikawa) to 0.7435 for the American population. This suggests that the previously unsampled population of *A. crassus* from Wakayama is likely to be isolated from other populations, and that the colonisation of North American eels did not originate from this population.

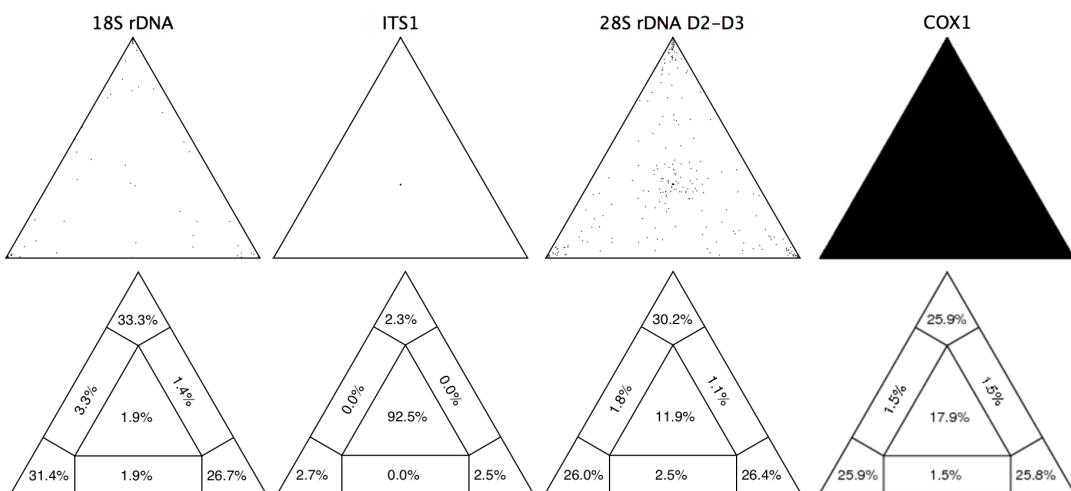


**Figure 4.8.**: Statistical parsimony network of the *CoxI* sequences from *A. crassus*. All *CoxI* haplotypes share the same 28S rDNA D2-D3 haplotype “h3”, except the four specimens JPN-K4B1 (“h32”), JPN-WK6 (“h34”), JPN-K4B4 and JPN-K4B4b (both “h33”) which show the 28S rDNA D2-D3 haplotype “h5”.

#### 4.7. Assessment of the phylogenetic utility of the different barcoding genes

The usefulness of a data set for phylogenetic analyses can be evaluated, based on the distribution of the likelihood vectors calculated from all possible quartets of a set of aligned sequences.

The plots from the results of the likelihood-mapping analysis of each of the four molecular markers are depicted in Figure 4.9. Likelihood-mapping was performed for all  $\binom{n}{4}$  quartets of each dataset, where  $n$  is the number of sequences in the alignments ( $n_{18S rDNA} = 10$ ,  $n_{ITS1} = 5$ ,  $n_{28SrDNA} = 13$ ,  $n_{Cox1} = 103$ ).



**Figure 4.9.: Plots from the results of the likelihood-mapping analysis showing the phylogenetic utility of the four molecular markers.** The upper row shows the distribution of each of the possible quartets of taxa for each marker. The lower row indicates the proportion of quartets in each region. The regions at the vertices of each triangle ( $A_1, A_2, A_3$ ) represent the quartets that could be completely resolved, the central region ( $A_*$ ) indicates the percent of quartets that could not be resolved and the regions between the vertices of each triangle ( $A_{12}, A_{23}, A_{13}$ ) contain the quartets for which no distinction could be made between two trees. The black colouring of the upper plot of *Cox1* resides in the fact that  $\binom{103}{4} = 4,421,275$  quartets were analysed.

The 18S rDNA dataset used for the phylogenetic analysis of the Anguillicolidae shows the smallest amount of star-likeness with 1.9% of the quartet points in region  $A_*$  and the highest amount of tree-likeness as 91.8% of all quartets being placed within the areas  $A_1$ ,  $A_2$  and  $A_3$ . The distribution of the quartets therefore suggests that the 18S rDNA dataset provides the highest phylogenetic utility among the datasets of this study, although it is more likely that this resulted from the outgroups included in the dataset, since only a small numbers of mutational changes are observed between the 18S sequences of the Anguillicolidae (see 4.1).

The likelihood-map for ITS1 revealed a reasonable amount of star-likeness and only 7.5% of the quartets could be resolved. Again, this is caused by the high sequence identity between

the ITS1 sequences of *A. crassus*, as well as by the fact that no suitable outgroup sequence could be identified.

In contrast, the plots for 28S rDNA D2-D3 and *Cox1* display only 11.9% and 17.9% of quartets positioned in  $A_*$ , respectively. The amount of completely resolved trees is higher for 28S rDNA D2-D3, namely 82.6%, compared to 77.6% of *Cox1*. Although these are reasonable amounts of tree-likeness for both datasets, these datasets show further advantages. Due to the work of Wielgoss et al. (2008) and Geiß and Sures (2010a), the *Cox1* dataset shows the largest number of taxa (see section 4.6.4).

Unfortunately, for the 28S rDNA D2-D3 dataset no sequences apart from those of this study could be obtained. However, it bears a feature which none of the other datasets displays to this extent: during the evolutionary history of the Anguillicolidae multiple insertion and deletion events (indels) occurred within this region (see Figure 4.4). This is not reflected by the likelihood plot as only the nucleotide sequences were used as input for the analysis.

For both phylogenetic analyses, of 28S rDNA D2-D3 sequences alone and the “supermatrix”, indels were coded into the dataset as described in section 3.9.4, increasing the information contained.

## 4.8. Phylogenetic analyses

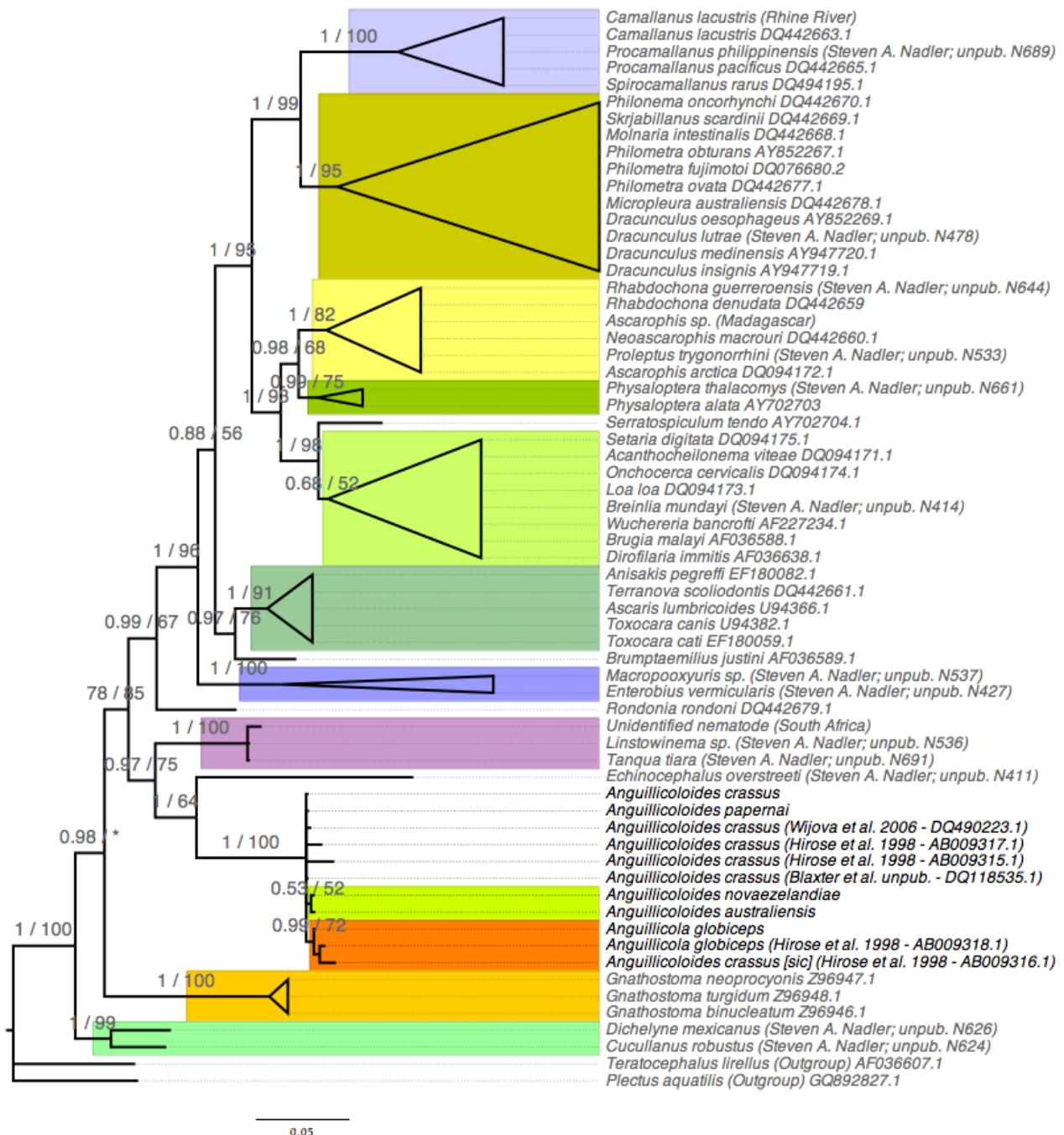
The branch supports, abbreviated as BPP for Bayesian posterior probabilities and BS for bootstrap values, of the trees inferred through phylogenetic analyses are interpreted as described in section 3.9.11.

### 4.8.1. Phylogenetic analyses of the 18S rDNA sequences

A phylogenetic tree of the clade 8 nematodes was constructed with the idea in mind, that although the basal position of *A. crassus* has been determined by Wijová et al. (2006) and Nadler et al. (2007), this position may not be necessarily shared with the other anguillicolid species. Furthermore, since new sequences are available, the hypothesis, that the basal position of *A. crassus* might be an artefact due to the lack of pivotal taxa, can be tested.

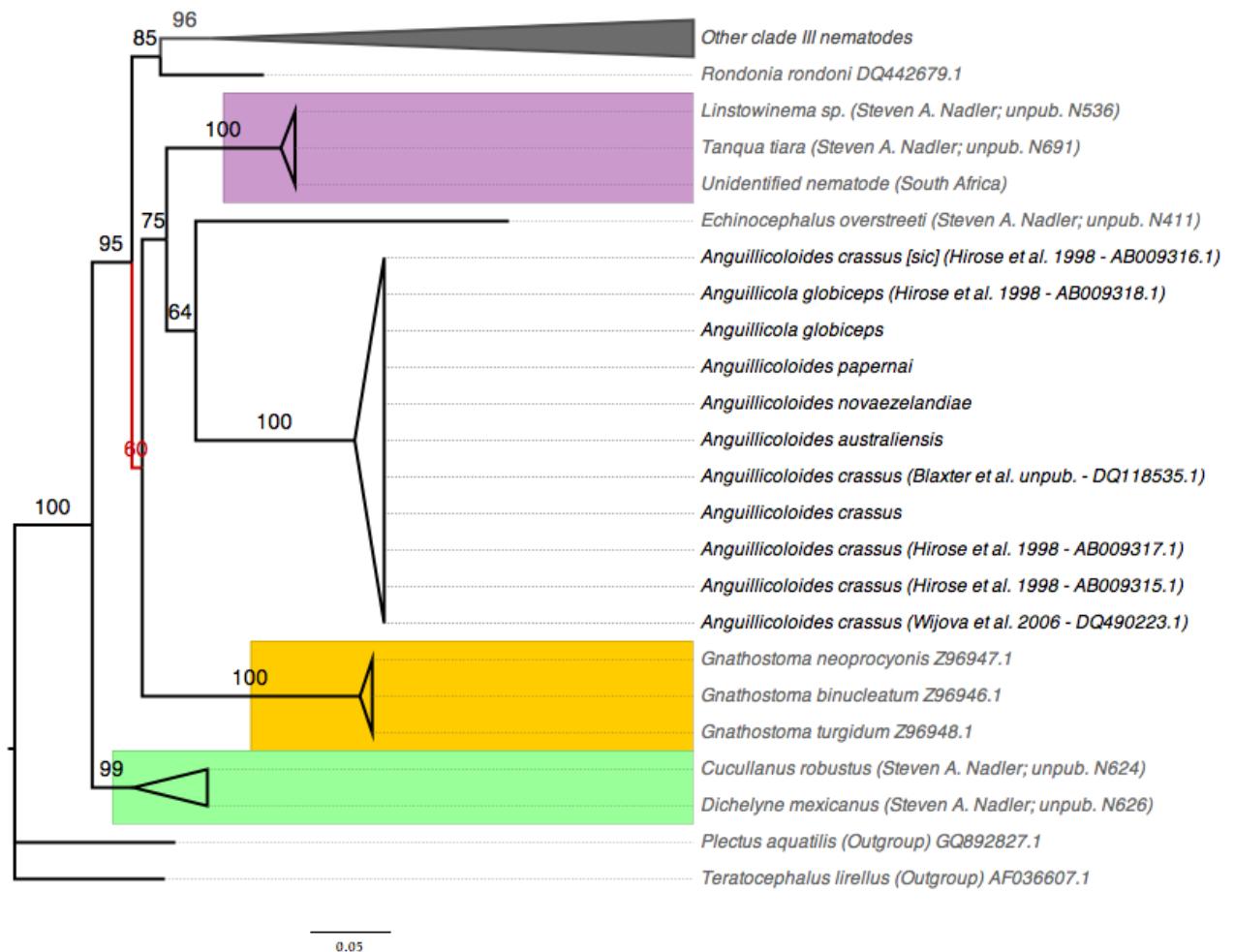
The phylogenetic tree was constructed using Bayesian inference (MrBayes) and maximum likelihood (RAxML) methods. The trees show few differences and, for the sake of simplicity, both are depicted in Figure 4.10.

Between the two trees only one difference in the branching order can be observed. In the Bayesian inference tree, three nematodes of the genus *Gnathostoma* constitute the most basal clade within the clade III nematodes *sensu* Blaxter et al. (1998), while in the maximum like-



**Figure 4.10.: Combined phylogram of the phylogenetic analysis performed on the 18S rDNA sequences from clade 8 nematodes *sensu* Holterman et al. (2006) using Bayesian inference (MrBayes) and maximum likelihood (RAxML) methods. The parameters for both analyses are listed in section 3.9.9 and 3.9.10. Bayesian posterior probabilities (BPP) followed by bootstrap values (BS) are displayed on the internal branches. Deviation in branching structure between the two trees is marked by \* . Branches are collapsed whenever possible, based on taxonomic affiliations.**

lihood analysis clade III is divided into two sub-clades. The lower clade is composed by all members of the two families Gnathostomatidae and Anguillicolidae, whereas the upper clade contains the remaining clade III nematodes. A partially collapsed phylogram of the maximum likelihood analysis with the difference in branching order being highlighted is shown in Figure 4.11. Similar to the results of Nadler et al. (2007), who recovered uniformly the nematode



**Figure 4.11.:** Phylogram of the maximum likelihood analysis performed, using RAxML. The exact parameters of the analysis are listed in section 3.9.10. Bootstrap supports (BS) are displayed on internal branches. The difference in branching order in comparison to the Bayesian inference tree is highlighted in red. Other clade III nematodes = Camallanoidea, Dracunculoidea, Rhabdochonidae, Physalopteridae, Diplotriaenoidea, Filarioidea, Ascarididae, Rhigonematidae, Oxyuroidea, Atracidae.

*Truttaedactinis truttae* (Cucullanidae Cobbold, 1864) as being basal to all clade 8 nematodes *sensu* Holterman et al. (2006), and therefore rendering the order Ascaridida paraphyletic; the two cucullanid nematodes *Dichelyne mexicanus* and *Cucullanus robustus*, both parasitising freshwater fish in Mexico and South Korea, respectively, are found to be the most basal taxa

within clade 8. This supports the hypothesis that the Ascaridida are paraphyletic and members of the Cucullanidae are the most basal clade within clade 8.

Compared to the results of Wijová et al. (2006) and Nadler et al. (2007) no great differences in respect to the overall branching order can be observed, despite the fact that fifteen new taxa have been included. However, four of the new taxa helped considerably to shed light on the phylogeny of the Anguillicolidae, decreasing the great branch length of *A. crassus* encountered in previous analyses. Those taxa are *Linstownema* sp., *Tanqua tiara*, *Echinocephalus overstreeti*, a parasite of stingrays (Dasyatidae), and the nematode from South Africa (SNR-118), which have been discussed in part in section 4.2.

The clade grouping those taxa together with the Anguillicolidae, is found to have robust support ( $BPP = 0.97$ ,  $BS = 75$ ) under both phylogenetic methods. This is a surprising result, since it not only indicates the subfamily Gnathostomatinae, comprising the three genera *Echinocephalus*, *Tanqua* and *Gnathostoma*, being paraphyletic, but also suggests that members of the Seuratoidea, comprising the families Cucullanidae and Seuratidae Hall, 1916, the latter represented by *Linstownema* sp., are positioned together with both the Anguillicolidae and the subfamily Gnathostomatinae at the basal part of the clade 8 *sensu* Holterman et al. (2006).

Additionally, the nematode from South Africa (SNR-118) is likely to be a member either of the Gnathostomatinae or the Seuratidae.

Hence, based on the results of the Bayesian inference and maximum likelihood analysis, instead of a basal clade composed of the genera *Gnathostoma* and *Anguillicoloides* presented by Wijová et al. (2006) and Nadler et al. (2007), the most basal taxa in clade III *sensu* Blaxter et al. (1998) are found to be members of the superfamilies Seuratoidea, Gnathostomatoidea and Anguillicoidea, displaying an immense host diversity which comprises fresh water fish, cartilaginous fish (stingrays), mammals and reptiles.

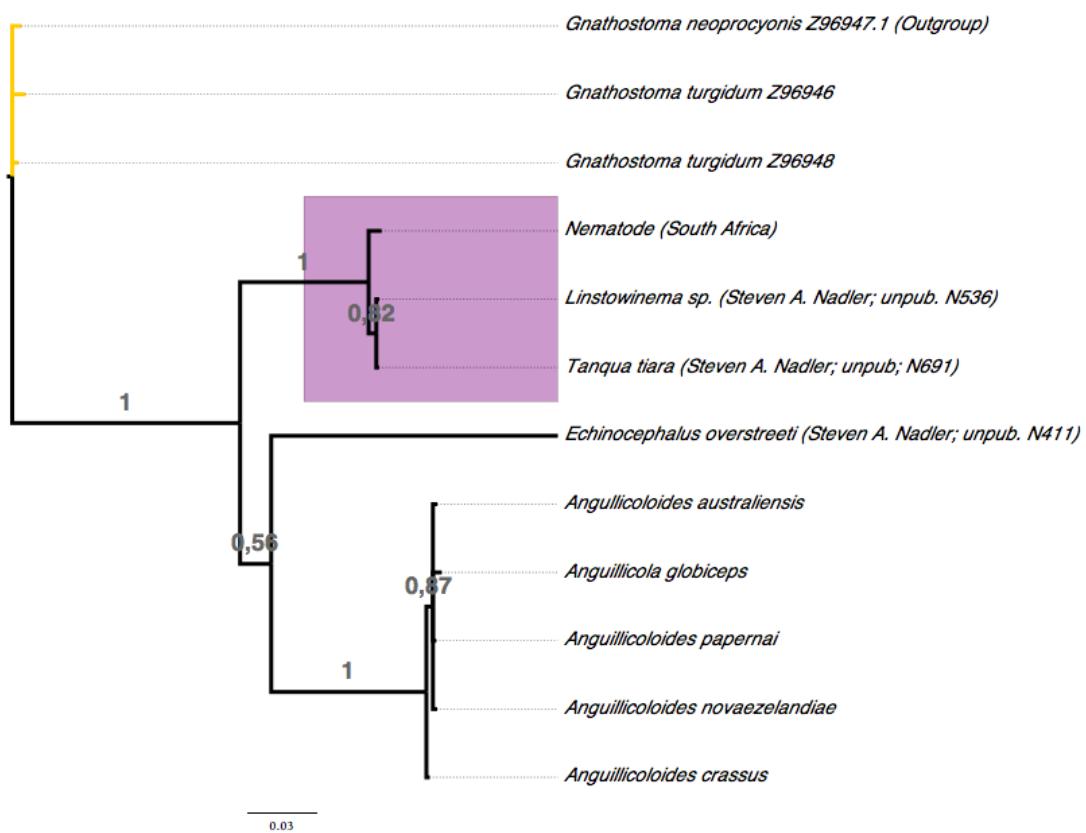
In order to clearly illustrate the relationships among the Anguillicolidae, several additional analysis using Bayesian inference under the two DNA evolution models proposed by MrModeltest, namely GTR+ $\Gamma$  and SYM, as well as under the parametric richer model GTR+I+ $\Gamma$ , were performed. The results from the In Bayes Factor ( $ln(B)$ ) analyses of the MrBayes runs under different models are displayed in Table 4.6.

Concerning the topology no differences are observed under the different models. The tree of inferred under the GTR+I+ $\Gamma$  model, that appears to represent the data the best, is shown in Figure 4.12. The trees inferred under the other two models are depicted in appendix A.5.1.

Although the support for the branch leading to the clade of *E. overstreeti* and all Anguillicolidae is low ( $BPP = 0.56$ ), this clade is recovered in all of the tree analysis, suggesting that the members of the genus *Echinocephalus* and the family Anguillicolidae share a common ancestor.

Model	THM	$\ln(B)$	18S rDNA inference
GTR+Γ	-2326.15	-	-
SYM	-2357.00	32.13	Observed data extremely better represented by GTR+Γ
GTR+I+Γ	-2321.76	-	-
GTR+Γ	-2326.15	4.39	Observed data strongly better represented by GTR+I+Γ
GTR+I+Γ	-2321.76	-	-
<i>Anguillicoloides</i> monophyly (above model)	-2325.62	3.86	Strong evidence against monophyly of the genus <i>Anguillicoloides</i>

**Table 4.6.:** Summary of  $\ln$  Bayes Factor ( $\ln(B)$ ) analyses of alternative models for the 18S rDNA dataset comprising all anguillicolid sequences, as well as seven other clade III nematodes *sensu* Blaxter et al. (1998). THM = Total harmonic means of the negative log likelihoods of the MrBayes runs.



**Figure 4.12.:** Phylogram of the phylogenetic analysis performed on the 18S rDNA sequences of the Anguillicolidae and seven other clade III nematodes *sensu* Blaxter et al. (1998) using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches.

Similarly, but robustly supported (BPP = 1) the species *A. crassus* is found basal to all Anguillicolidae in all analyses. The support for the branch defining the clade of *A. australiensis*,

*A. globiceps*, *A. papernai* and *A. novaezelandiae*, hereafter referred to as AGPN - clade, is moderate ( $BPP = 0.87$ ), and together with the results from the  $\ln$  Bayes factor ( $\ln(B)$ ) analysis (see Table 4.6), indicates strong evidence against the monophyly of the genus *Anguillicoloides*.

#### 4.8.2. Phylogenetic analyses of the 28S rDNA D2-D3 sequences

The selection of outgroup taxa for this analysis resulted to be more difficult than for the 18S rDNA dataset, since fewer 28S rDNA D2-D3 sequences from clade III nematodes *sensu* Blaxter et al. (1998) were available.

Apart from the nine haplotypes found within the Anguillicolidae, the sequences from the South African nematode SNR118, from *Contracaecum spiculigerum* and from *Toxocara vitulorum* were included in the analyses. The latter was defined as outgroup taxon.

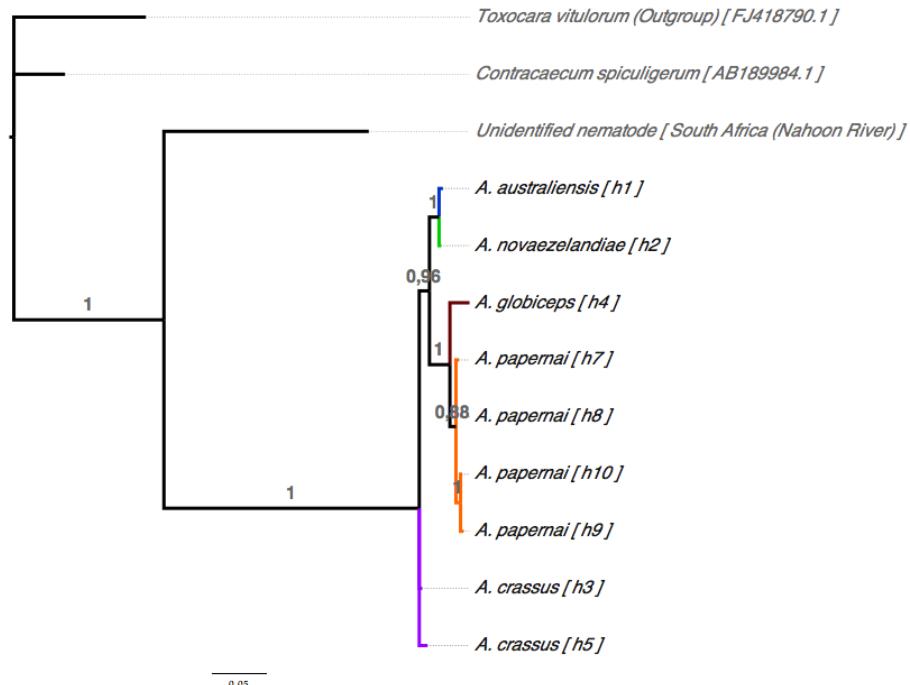
The only model for DNA evolution suggested by MrModeltest was HKY+ $\Gamma$ , however an additional analysis using the parametric richer model GTR+I+ $\Gamma$ , was carried out and was found to better represent the observed data. The models for DNA evolution used, with its respective total harmonic mean of the negative log-likelihoods and the conclusions drawn from them are listed in Table 4.7.

The fact that the GTR+I+ $\Gamma$  model appears to suit the observed data better, is true for both analysis GTR+I+ $\Gamma$ , where indels are treated as gaps, and GTR+I+ $\Gamma$  (SIC), where indels are coded as explained in section 3.9.4. The tree inferred in the GTR+I+ $\Gamma$  (SIC) analysis is depicted in Figure 4.13. The trees inferred in the GTR+I+ $\Gamma$ , HKY+I+ $\Gamma$  and HKY+I+ $\Gamma$  (SIC) analyses can be found in appendix A.5.2.

Although no variation in branching order is observed between the two analyses, GTR+I+ $\Gamma$  and GTR+I+ $\Gamma$  (SIC), in the latter analysis the branch leading to the AGPN - clade is considerably better supported, with an BPP - value of 0.96, in comparison to 0.79 in the GTR+I+ $\Gamma$  analysis.

Model	THM	$\ln(B)$	28S rDNA D2-D3 Inference
GTR+I+ $\Gamma$	-2498.96	-	-
HKY+ $\Gamma$	-2499.36	0.40	Observed data is slightly better represented by GTR+I+ $\Gamma$
GTR+I+ $\Gamma$ (SIC)	-2602.24	-	-
HKY+ $\Gamma$ (SIC)	-2604.47	2.23	Observed data is substantially better represented by GTR+I+ $\Gamma$
GTR+I+ $\Gamma$ (SIC)	2602.24	-	-
<i>Anguillicoloides</i> monophyly (above model)	-2605.79	3.55	Strong evidence against monophyly of the genus <i>Anguillicoloides</i>

**Table 4.7.: Summary of  $\ln$  Bayes Factor ( $\ln(B)$ ) analyses of alternative models for the 28S rDNA D2-D3 dataset. THM = Total harmonic mean of the negative log likelihoods of the MrBayes runs, (SIC) = Simple Indel Coding (see section 3.9.4)**



**Figure 4.13.:** Phylogram of the phylogenetic analysis performed on the nine anguillicolid 28S rDNA D2-D3 haplotypes together with three clade III nematodes *sensu* Blaxter et al. (1998) using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. Indels were coded as explained in section 3.9.4. The exact parameters for the analysis is listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For each anguillicolid taxon the corresponding haplotype in the statistical parsimony network in Figure 4.3 is indicated in square brackets.

As in the phylogenetic analysis of the 18S rDNA sequences, strong evidence was encountered for the monophyly of the genus *Anguillicoloides*, being based both on branch support and In Bayes factor  $\ln(B)$  analyses. (see Table 4.7).

#### 4.8.3. Phylogenetic analyses of the Cox1 sequences

The analyses were carried out using the 100 Cox1 haplotypes found within the Anguillicolidae, together with the sequences of *Strongylida* sp. and *Toxocara cati* as specified in section 3.9.9. The results of the MrBayes runs are displayed and interpreted in Table 4.8.

Again, the analysis performed under the GTR+I+Γ model appears to represent the observed data better than the simple model HKY+I+Γ and no differences in the branching order can be observed. The resulting phylogram inferred under the GTR+I+Γ model is shown in Figure 4.14, while the HKY+I+Γ tree is depicted in appendix A.5.3.

#### 4. Results and Discussion

---

Model	THM	$\ln(B)$	CoxI Inference
GTR+I+Γ	-3507.25	-	-
HKY+I+Γ	-3535.76	28.51	Observed data is substantially better represented by the GTR+I+Γ
GTR+I+Γ	-3507.25	-	-
<i>Anguillicoloides</i> monophyly	-3525.85	21.60	Very strong evidence against monophyly of the genus <i>Anguillicoloides</i>
<i>A. novaezelandiae</i> monophyly	-3510.68	3.43	Strong evidence against monophyly of the species <i>A. novaezelandiae</i>

**Table 4.8.:** Summary of ln Bayes Factor ( $\ln(B)$ ) analyses of alternative models for the *coxI* dataset. THM = Total harmonic mean of the negative log likelihoods of the MrBayes runs.

The overall topology of this tree differs from the topology of the trees inferred using 18S rDNA and 28S rDNA D2-D3 sequences in the point that *A. crassus* is not the most basal species and therefore the AGPN - clade is not recovered. Instead the Anguillicolidae are split into two clades: one comprises the two species *A. papernai* and *A. globiceps*, while the other contains *A. crassus* and a clade formed by the two Oceanian species *A. australiensis* and *A. novaezelandiae*.

All branches display robust support (BPP = 0.95 – 1) apart from the branch defining the clade of the *A. papernai* and *A. globiceps* which is only moderately supported (BPP = 0.83).

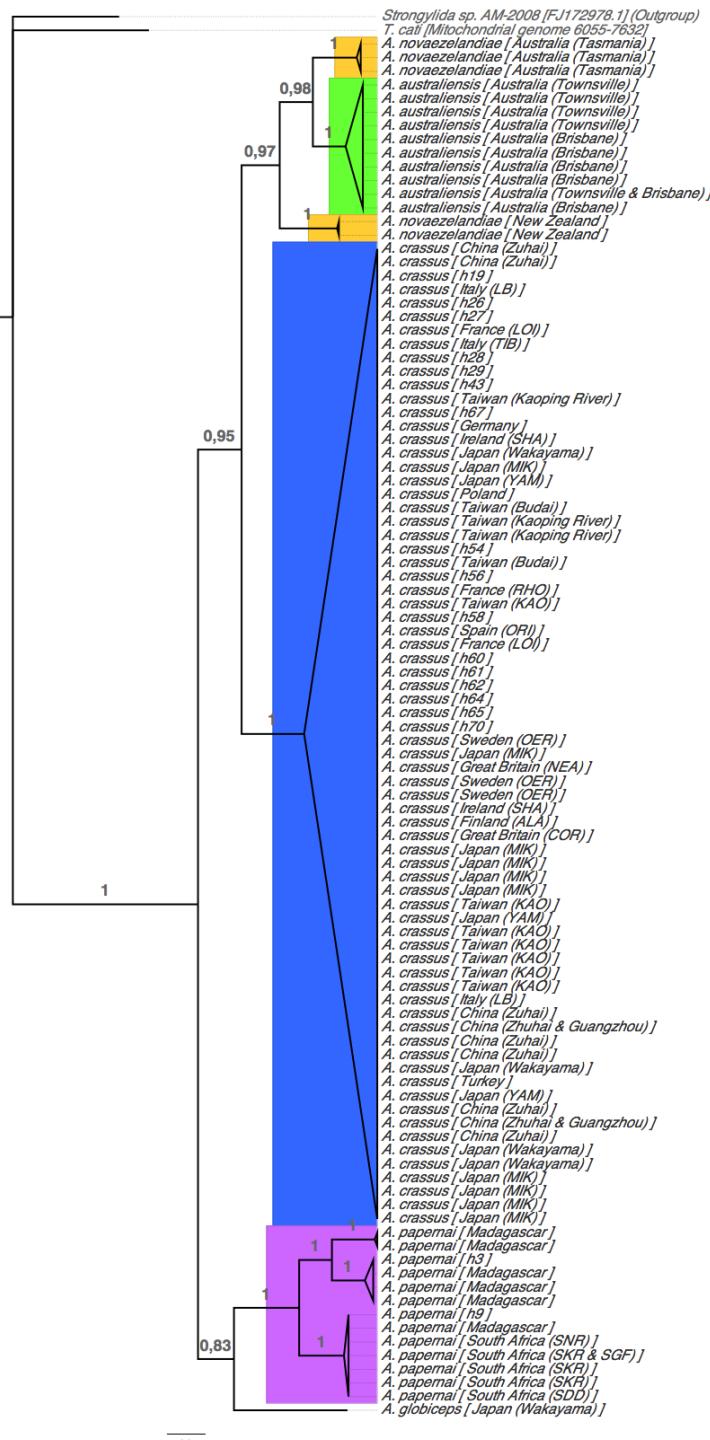
Interestingly, the species *A. novaezelandiae* is recovered always as being paraphyletic. The population from Tasmania appears to be more closely related to the specimens of *A. australiensis* than to the population of *A. novaezelandiae* from New Zealand. Furthermore, the Bayes factor analysis shows strong evidence against the monophyly of this species.

This enforces the hypothesis stated in section 4.6.2, that the species *A. novaezelandiae* is likely to include one cryptic species despite showing a homogenous morphology.

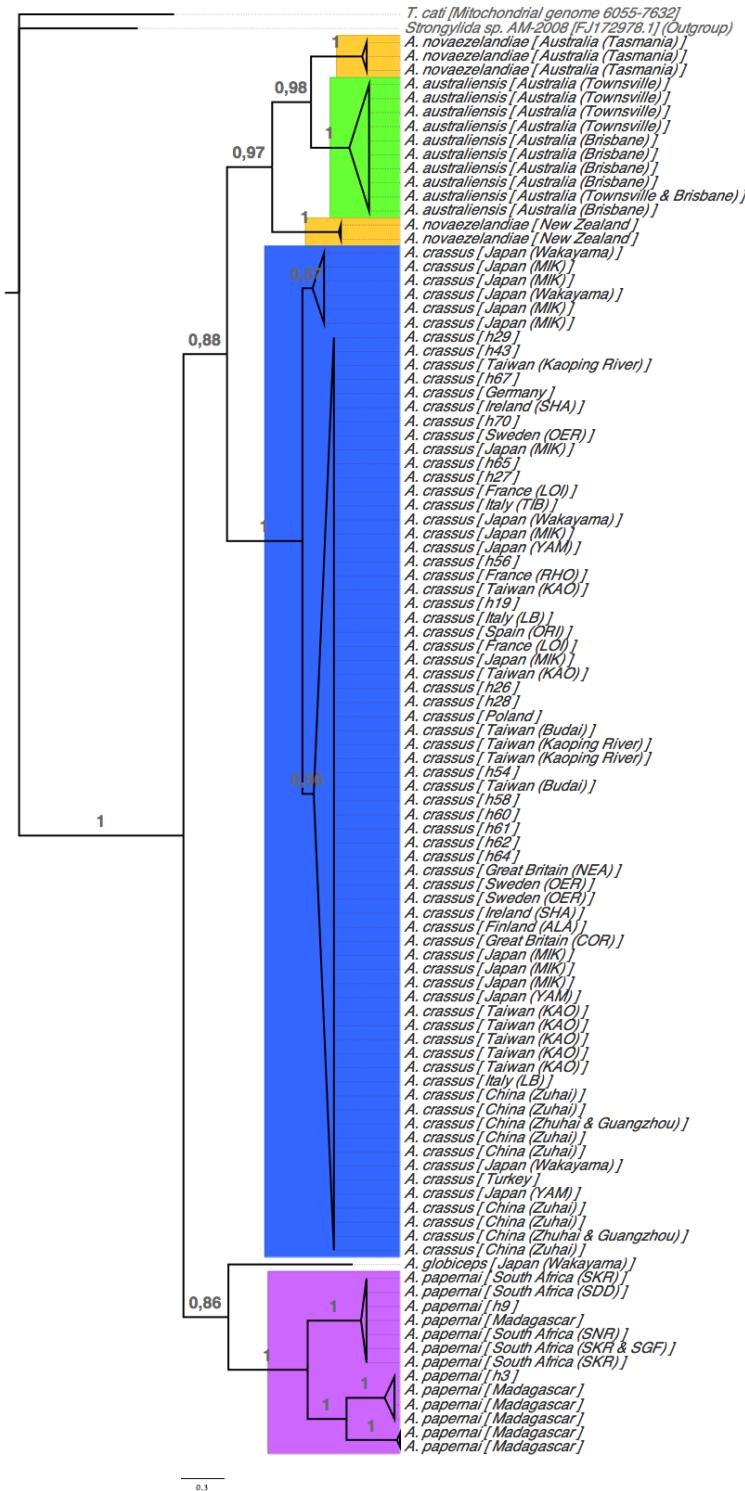
Based on the theory, that the third codon position of a protein-coding gene may contain “noisy” data due to high mutational rates leading to multiple substitutions, a second type of analysis was performed on the *Cox1* dataset, in which the third codon position was excluded from the analysis. The phylogram is depicted in Figure 4.15.

The only effect on topology, compared to the analyses in which the third codon position was included, that can be observed is that the species *A. crassus* is divided into two clades. One comprises Japanese specimens from Wakayama and Mikawa Bay, while the other contains all remaining specimens.

This in part corresponds to the division pattern found in the 28S rDNA D2-D3 dataset, where those two Japanese *A. crassus* from Wakayama displayed the 28S rDNA D2-D3 haplotype “h4” (see section 4.6.4). This proves that the phylogeny inferred from the *Cox1* sequence is not



**Figure 4.14.:** Phylogram of the phylogenetic analysis performed on the 100 *Cox1* anguillicolid haplotypes and two other nematode sequences using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For anguillicolid sequences found in more than one population the corresponding haplotype is indicated in square brackets and its geographical distribution can be checked in the respective statistical parsimony network in section 4.6.



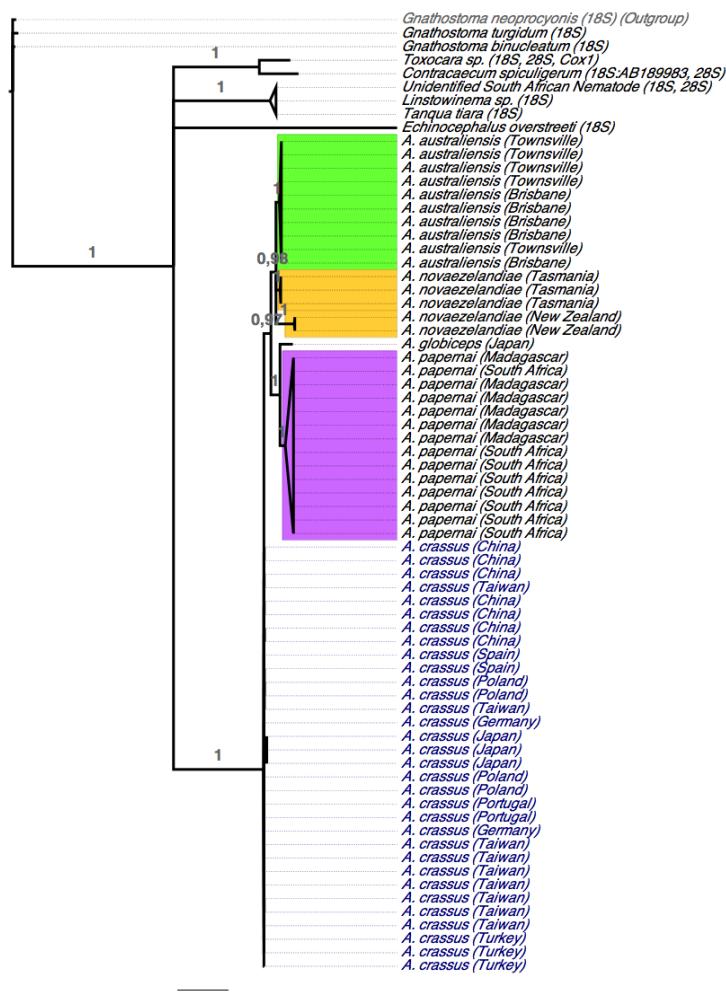
**Figure 4.15.: Phylogram of the phylogenetic analysis performed on the 100 *Cox1* anguillicolid haplotypes and two other nematode sequences using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. The third codon position was excluded from the analysis. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For anguillicolid sequences found in more than one population the corresponding haplotype is indicated in square brackets and its geographical distribution can be checked in the respective statistical parsimony network in section 4.6.**

based on a high number of mutations at the third codon position, but due to mutational changes at the first and second position.

Regarding branch supports, the exclusion of the third codon position increases the support for the *A. papernai* - *A. papernai* clade, while it decreases the support for the *A. crassus* - *A. australiensis* - *A. novaezelandiae* - clade. The paraphyly of the species *A. novaezelandiae* is found again robustly supported.

#### 4.8.4. Phylogenetic analysis of the “supermatrix”

The phylogenetic analysis of the “supermatrix” serves as a summary analysis and comprises the datasets of the four barcoding genes 18S rDNA, ITS1, 28S rDNA and *Cox1*. The phylogram is depicted in Figure 4.16.



**Figure 4.16.:** Phylogram of the phylogenetic analysis performed under the GTR+I+Γ model on the “supermatrix”. The parameters for the analysis are specified in section 3.9.9. The sequences used for the non-anguillicolid taxa are specified in brackets.

However, it should be interpreted with care since for many non-anguillicolid taxa as well as for the specimens of *A. novaezelandiae* from New Zealand, not all sequences for the four genes have been obtained and therefore the data matrix is composed of an high amount of missing data.

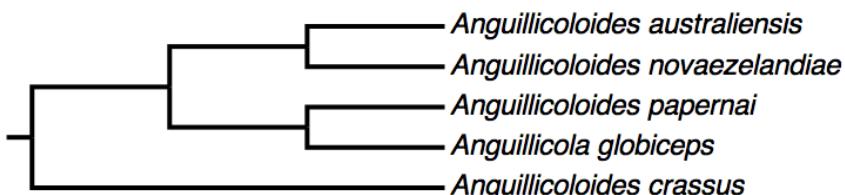
The lack of phylogenetic information, due to missing data, may be reliable for the soft polytomies found in the tree. The first polityomny comprises the taxa *Toxocara* sp., *Contracaecum spiculigerum*, the nematode from South Africa (SNR-118), *Linstownema* sp., *Tanqua tiara* and *Echinocephalus overstreeti* and the root of the Anguillicolidae. The second poltyomy involves *A. australiensis* and *A. novaezelandiae* from New Zealand and Tasmania.

The species *A. crassus* is recovered basal to the other anguillicolid nematodes, as it is the case in the phylogenetic analyses of the 18S rDNA and the 28S rDNA D2-D3 datasets.

The overall branch support was robust ranging from Bayesian posterior probabilities of 0.97 to 1, indicating again that the poltyomies are likely caused by the high amount of missing data within the pivotal taxa.

#### 4.8.5. Summary of the phylogenetic analyses

As a summary of the phylogenetic analyses, a cladogram representing the phylogenetic relationships of the Anguillicolidae is displayed in Figure 4.17.

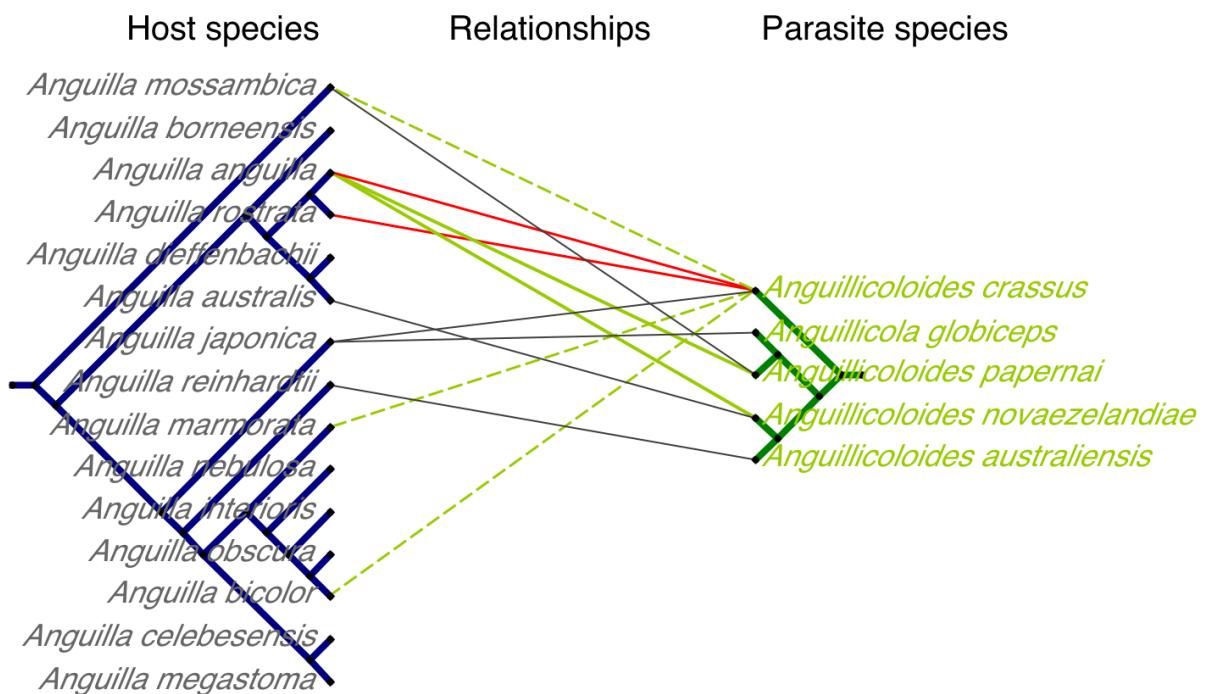


**Figure 4.17.: Cladogram of the evolutionary relationships among the family Anguillicolidae (Nematode: Anguillicoloidea).**

#### 4.9. A co-evolution theory for the host-parasite-system of *Anguilla* infected with anguillicolid nematodes

Since a robust phylogenetic tree for the Anguillicolidae was available, it was compared to the phylogenetic tree of the host genus *Anguilla*, inferred by Minegishi et al. (2004) and based on a total of 15187 sites of mitochondrial DNA sequences.

The two trees are drawn in Figure 4.18, together with lines indicating the traditional host-parasite relationships as well as the novel relationships that have been established due to international live eel trade (see section 1.3.2).

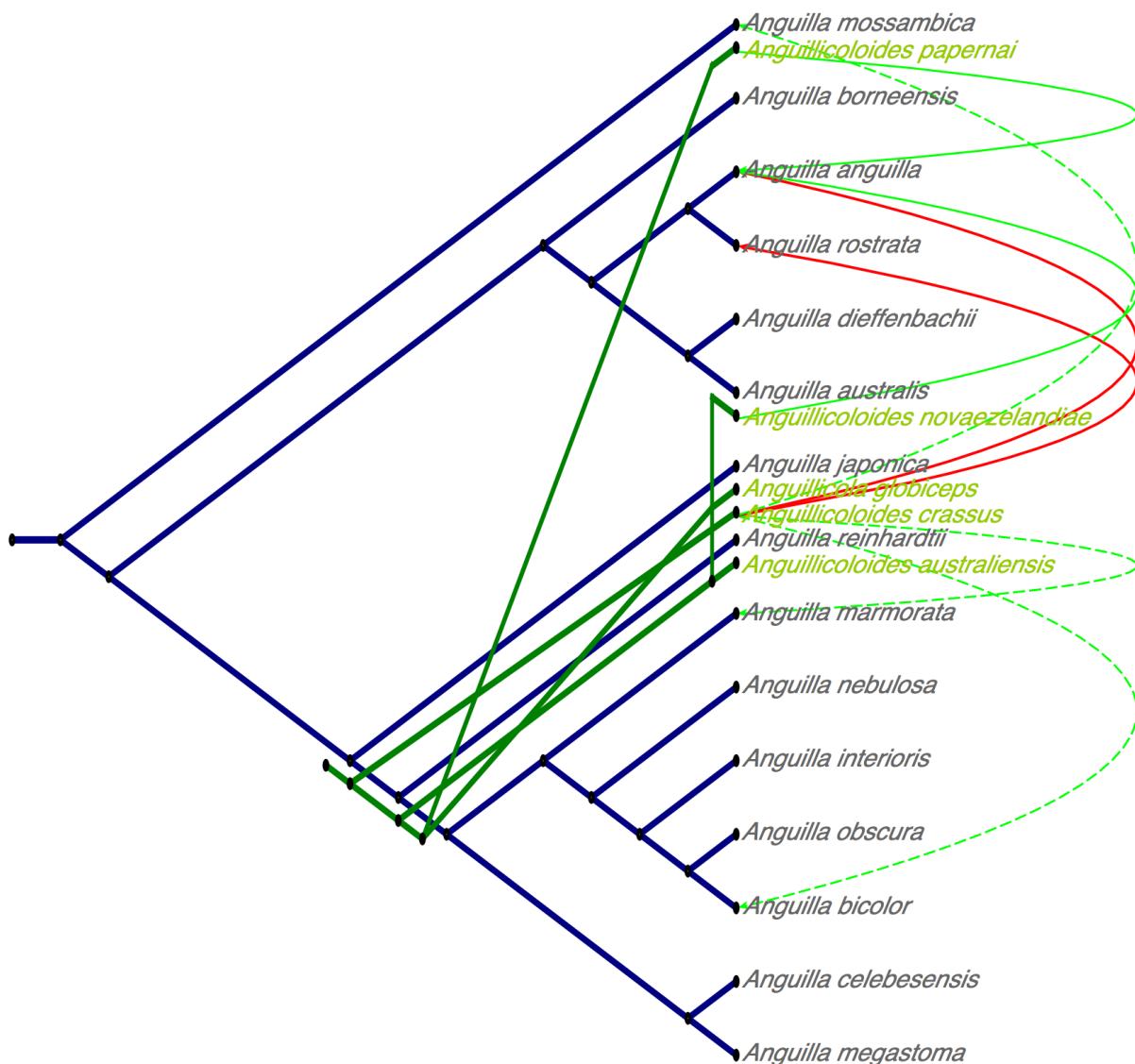


**Figure 4.18:** Cladograms of the genus *Anguilla* (left tree) (Minegishi et al., 2004) and its swim bladder parasites, the family Anguillicolidae (Nematoda: Anguillicoloidea) (right tree). Grey lines = traditional host-parasite relationships, displaying low abundances and low pathogenicity (parasite endemic in host); red lines = novel host-parasite relationships, displaying high abundances and pathogenicity; green lines = novel host-parasite relationship, displaying low abundances and pathogenicity; dashed lines = completion of life cycle has not been demonstrated.

As argued by Minegishi et al. (2004), the dispersal route of the genus *Anguilla* is far from clear and the present geographic distribution could be attributed to, for example, multiple dispersal events, multidirectional dispersion or past extinctions.

However, distinctions based on the evolutionary relationships within the genus *Anguilla* can be made between three major geographical groups: the Atlantic group (*A. anguilla* and *A. rostrata*) and the Oceanian group (*A. dieffenbachii* and *A. australis*), which both diverged early in the evolutionary history of the genus *Anguilla* from the Indo-Pacific group (*A. japonica*, *A. reinhardtii*, *A. marmorata*, *A. nebulosa*, *A. interioris*, *A. obscura*, *A. bicolor*, *A. celebesensis* and *A. megastoma*).

Based on the findings of Wielgoss et al. (2008) and of those of the present study, the common ancestor of all anguillicolid species is likely to have had a South-East Asian distribution since



**Figure 4.19.:** Reconstruction of a possible co-evolution scenario for the genus *Anguilla* (blue) (Minegishi et al., 2004) and its swim bladder parasites, the family *Anguillicolidae* (Nematoda: *Anguillicolidae*) (green). Red lines = novel host-parasite relationships displaying high abundances and pathogenicity; green lines = novel host-parasite relationship displaying low abundances and pathogenicity; dashed lines = completion of life cycle has not been demonstrated..

the highest diversity of anguillicolid species is found in this part of the world. Currently the only species endemic West of the E80° longitude and East of the 180° is *A. papernai*.

Hence, the ancestors of both, the parasites and the hosts might have shared a similar geographical distribution in the past. Based on this assumption, the hypothesis is proposed, that the common ancestor of the *Anguillicolidae* may have been already associated with the common ancestor of the Indo-Pacific clade of the genus *Anguilla*.

A parameter-adaptive event-based approach for reconciliation analysis (see section 3.9.12) of the two phylogenetic trees together with the traditional host-parasite relationships (grey

lines) was carried out in order to draw a tree to illustrate this hypothesis. The host-parasite-relationships established through international eel trade were drawn on the tree and the results are depicted in Figure 4.19.

This hypothesis would imply that the split of *A. crassus* from the other anguillicolid nematodes occurred due to a co-speciation of *Anguilla japonica* and of *A. crassus*, which would explain the lack of pathogenicity observed today for this parasite within this host species.

A second co-speciation event involving the ancestor of *Anguilla reinhardtii* and the common ancestor of *Anguillicoloides australiensis* and *Anguillicoloides novaezelandiae*, would have led to another stable host-parasite relationship. Based on the fact, that the two eel species *Anguilla reinhardtii* and *Anguilla australis* now share a similar geographical distribution, a host switch in the past, from *A. reinhardtii* to *A. australis* may have led to a speciation event that gave rise to the two Oceanian anguillicolid nematodes, *Anguillicoloides novaezelandiae* and *Anguillicoloides australiensis*, that are found today in Australia and New Zealand.

In contrast, under this scenario, the two closely related species *Anguillicola globiceps* and *Anguillicoloides papernai*, would both have experienced relatively recent host-switches to the eel species *Anguilla japonica* and *Anguilla mossambica*, respectively. In the case of *A. globiceps* this is supported by the fact that today high infection rates of this nematode in *A. japonica* leads to a thickening of the swim bladder wall (Yamaguti, 1933; Nagasawa et al., 1994), indicating a relatively recent host-parasite relationship. However, no comparable data is available for the host-parasite system of *A. mossambica* and *A. papernai*.

Under this scenario of co-evolution, the patterns, concerning abundance and pathogenicity, found in those host-parasite systems that have been established due to international live eel trade starting in the early 1980s, can be evaluated.

The low abundances of *A. crassus* in *Anguilla mossambica* may be caused by the co-evolution this eel species has experienced with *A. papernai*. Similarly, the low abundances of *A. crassus* in the eel species *A. marmorata* and *A. bicolor* supports this scenario since both eel species diverged after their ancestors had already experienced a co-evolution with the common ancestor of the AGPN clade (*A. australiensis*, *A. globiceps*, *A. papernai* and *A. novaezelandiae*).

On the contrary, neither the two Atlantic eel species, *Anguilla anguilla* and *Anguilla rostrata*, nor their ancestors would have had contact with anguillicolid nematodes, which would explain the fact that *A. crassus*, *A. papernai* and *A. novaezelandiae* are capable of completing their life cycles within the *naive* host *A. anguilla*. Furthermore, paired with the ability of *A. crassus* to use a broad variety of organisms as intermediate and paratenic hosts, this would explain the high abundances, and the associated pathogenicity, observed for this nematode in the Atlantic eel

#### 4. Results and Discussion

---

species *A. anguilla* and *A. rostrata* and might have been the main factors that led to its infamous role as a global invader *sensu* Colautti and MacIsaac (2004) and Taraschewski (2006).

## 5. Conclusions

Several insights into the evolutionary history and the population structures of the Anguillicolidae (Anguillicoloidea Yamaguti, 1935), a family of nematodes parasitising fresh-water eels, were gained through this diploma thesis. Furthermore, methods for identification of nematode specimens using nuclear and mitochondrial DNA barcoding genes were implemented and evaluated, which allowed identification of specimens of the Anguillicolidae to the species level as well to the population level.

Laboratory protocols were developed for tissue extraction of adult specimens as well as deteriorated larvae encapsulated within the intestine of the host (see Heitlinger et al. (2009) and section 3.2). Protocols for DNA barcoding of the nuclear genes for small subunit ribosomal DNA (18S rDNA), internal transcribed spacer 1 (ITS1), the D2-D3 region of the large subunit ribosomal DNA (28S rDNA D2-D3) and the mitochondrial gene for Cytochrome c oxidase subunit I (*Cox1*); involving tissue lysis, DNA amplification and sequencing, were implemented and adjusted for over 150 specimens of the Anguillicolidae and three other clade 8 nematodes *sensu* Holterman et al. (2006) parasitising fresh-water eels.

Bioinformatic solutions for sequence analysis, such as automated BLAST searches (section 3.8.2) and construction of consensus sequences for the barcoding genes based on phred-quality-scores of forward and reverse sequences (section 3.8.3), were developed, guaranteeing the assembly of high quality barcoding sequences which later were used in population structure and phylogenetic analyses.

The bioinformatic and phylogenetic analyses of the barcoding genes of the three additional clade 8 nematodes *sensu* Holterman et al. (2006) resulted useful for their identification, determining them as *Camallanus lacustris* sampled from *Anguilla anguilla* in Germany; a member of the Cystidicolidae (MAD-19), isolated from *Anguilla mossambica* in Madagascar and for which no further identification up to date is possible; and a member of the Gnathostomatinae or Echinonematinae (SNR-118) discovered on the serosa of the swim bladder of *Anguilla mossambica* in South Africa (see section 4.2 and 4.8.1).

Especially the 18S rDNA sequence of the latter nematode (SNR-118), together with five sequences of *Linstowinema* sp. (Echinonematinae Inglis, 1967), *Tanqua tiara* (Gnathostomatinae Railliet, 1895), *Echinocephalus overstreeti* (Gnathostomatinae Railliet, 1895), *Dichelyne mexicanus* (Cucullaninae Cobbold, 1864) and *Cucullanus robustus* (Cucullaninae Cobbold, 1864), resulted crucial for shedding light on the basal part of the phylogenetic tree of the clade 8

nematodes. Apart from the fact that the family Gnathostomatinae, comprising the three genera *Gnathostoma*, *Echinocephalus* and *Tanqua*, are recovered polyphyletic and well supported in all phylogenetic analyses; the cucullanid nematodes *D. mexicanus* and *C. robustus* were found to occupy the most basal position within clade 8, which was found to be consistent with findings of previous analyses of clade 8 nematodes (*Truttaedactinis truttae* in Nadler et al., 2007; Van Meegen et al., 2009) and was confirmed hereby (see section 4.8.1).

Although the 18S rDNA sequences of the five species of the Anguillicolidae, *Anguillicoloides crassus*, *Anguillicoloides papernai*, *Anguillicola globiceps*, *Anguillicoloides novaezelandiae* and *Anguillicoloides australiensis* were found to display small differences, ranging from one to three mutational substitutions between the species, discrimination on the species level was possible for each of the species using this barcoding gene. The sequences have been compared to available anguillicolid 18S rDNA sequences deposited on GenBank® (see section 4.2), possible misinterpretation of data has been discussed for the work of Hirose et al. (1998) and it was noted that adequate evaluation of sequencing quality is essential for this barcoding gene when working on anguillicolid nematodes.

The position of all five anguillicolid species is recovered within the basal part of clade 8, together with members of the superfamilies Seuratoidea and Gnathostomatoidea. Hence, the parasitic nematodes in the basal part of clade 8 display an enormous host diversity, comprising fresh water fish, cartilaginous fish (stingrays), mammals and reptiles (see section 4.8.1). This result implies that, concerning the early diversification of the entirely parasitic clade 8 of the phylum Nematoda, this clade does not show patterns of host-parasite co-speciation on the level of superfamilies, although a bias towards aquatic vertebrates is observed. The theory of Osche (1956), concerning pre-adaptation of nematode parasites to their host species, may still be true at a small evolutionary scale, but, keeping in mind the accelerated nucleotide substitution rates observed in most parasitic nematodes Holterman et al. (2006), might not be applicable to entire clades *sensu* Holterman et al. (2006).

However, the high branch lengths observed within many of the basal taxa of clade 8 suggest that either some superfamilies show increased substitution rates, because of an elevated production of free radicals attributed to high metabolic rates or to an elevated accumulation of replication errors due to short generation times (Gillooly et al., 2005); or that pivotal taxa are still missing, due to the lack of adequate sampling techniques or to past extinction events among these taxa.

For the four barcoding genes, their utility for inferring a phylogeny for the Anguillicolidae was assessed using likelihood-mapping (see section 4.7), which revealed, after careful interpretation of the results, that the most useful sequences for inferring a phylogeny of the Anguillicolidae are the 28S rDNA D2-D3 region and the *Cox1* barcoding gene. The ITS1 may contain suffi-

---

cient phylogenetic information, but was sequenced only for some specimens of the species *A. crassus*.

The 18S rDNA sequences showed little ability to resolve the phylogeny of the Anguillicolidae, because of the small amount of mutational changes, and suggests that this family of nematodes experienced a relatively recent radiation. Nevertheless, the mutational changes allowed to test for the hypothesis that the genus *Anguillicoloides* is monophyletic, for which strong evidence was found against. In fact, all of the three barcoding genes that have been sequenced for all anguillicolid species show strong, for 18S rDNA (Table 4.6) and 28S rDNA (Table 4.7), and very strong, for *Cox1* (Table 4.8), evidence against the monophyly of the Anguillicolidae, which should encourage taxonomist to re-evaluate the division of the Anguillicolidae into the two genera *Anguillicola* and *Anguillicoloides* since the only member of the genus *Anguillicola*, *Anguillicola globiceps*, is recovered as being closely related to *Anguillicoloides papernai*.

Furthermore, the availability of *Cox1* sequences of *Anguillicoloides novaezelandiae* from New Zealand together with the specimens from Tasmania from this study made it possible to test for the monophyly of this species. The results (Table 4.8) showed that there is strong evidence against the monophyly of this species. In all analyses of the *Cox1* sequences (section 4.8.3), the specimens of Tasmania were recovered to be more closely related to specimens of the species *A. australiensis* than to the specimens of *A. novaezelandiae* from New Zealand. Together with the results of the population structure analyses for *A. novaezelandiae* (section 4.6.2), which indicated a complete genetic isolation of the two populations, this suggests that the population of *A. novaezelandiae* from New Zealand has diverged before the speciation of the species *A. australiensis* and the population of *A. novaezelandiae* from Tasmania, despite displaying a similar morphology to *A. novaezelandiae*, which could be explained through convergent morphological evolution, since both populations of *A. novaezelandiae* parasitise within the same host eel species, *Anguilla australis*. Unfortunately, the lack of 18S rDNA and 28S rDNA D2-D3 sequences for the population of New Zealand, did not allow to test this hypothesis further.

As mentioned above, the species *Anguillicoloides papernai* and *Anguillicola globiceps* were recovered to be closely related and showing branch supports ranging from moderate (Figure 4.14) to high (Figure 4.13 and 4.15). As explained in section 4.6, problems were encountered when sequencing *Cox1* of *A. globiceps*, and only one high quality sequence was obtained. Although it was possible to use this sequence in the phylogenetic analyses, the small sampling size did not allow an evaluation of the population structure of this species.

In contrast, for *A. papernai* several 28S rDNA and *Cox1* sequences from different populations in South Africa and Madagascar were obtained. Interestingly, the 28S rDNA sequences of *A. papernai* displayed four distinct haplotypes (Figure 4.3), while the only other anguillicolid species showing more than one 28S rDNA D2-D3 haplotype was *A. crassus*. The population

structure analysis performed on the *Cox1* dataset of *A. papernai*, revealed a high number of haplotypes which corresponded in part to the pattern found for the 28S rDNA D2-D3 dataset (Figure 4.7). Both barcoding genes suggested that the populations within South Africa are likely to be subsets from the population from Madagascar, which led to the hypothesis that the population of *A. papernai* that colonised South Africa originated in Madagascar. Additionally, it was demonstrated that intra-host populations do not necessarily have to be homogenous, since in two cases specimens of *A. papernai* displaying different 28S rDNA D2-D3 and *Cox1* haplotypes were found inside the same host specimen.

The species *A. australiensis*, sampled from two populations in Queensland (Australia), was found to display only one 28S rDNA D2-D3 haplotype (Figure 4.3), although the analysis of population structure on the *Cox1* sequences, which show ten different haplotypes, indicates a strong genetic differentiation between the two populations (section 4.6.1). Of the ten *Cox1* haplotypes only one is shared between the two populations. This could reflect a recent expansion event of the populations of *A. australiensis*, since mutations in *Cox1* gene accumulate at a much faster rate than in the 28S rDNA D2-D3 region.

For the species *A. crassus* the greatest number of specimens was sequenced. Apart from the 18S rDNA, 28S rDNA D2-D3 and *Cox1* barcoding genes, the ITS region was sequenced from several populations with the idea in mind that this highly variable nuclear region might reveal interesting information concerning the evolutionary history of this species. Unfortunately, this region appeared to be quite stable among the sampled specimens (Figure 4.2), which reflects the recent colonisation of Europe by *A. crassus*. The 28S rDNA D2-D3 region displays an uniform haplotype for all specimens of *A. crassus* except for six specimens of the Japanese population from Wakayama (Figure 4.3). For the population structure analysis, 480 *Cox1* sequences of *A. crassus* were examined. The resulting statistical parsimony network (Figure 4.8) was discussed for the case of the the six Japanese samples displaying a different 28S rDNA D2-D3 haplotype. The results from the AMOVA suggest that this population of is likely to be genetically isolated from almost all other populations of *A. crassus*.

Based on the evolutionary relationships between the Anguillicolidae inferred within the scope of this thesis, a possible scenario of co-evolution of this family of nematodes and their respective hosts of the genus *Anguilla* was hypothesised and implications were discussed in the light of recent alterations of the host-parasite relationships due to anthropogenic influences (section 4.9).

In summary, the three barcoding genes (18S rDNA, 28S rDNA D2-D3 and *Cox1*) resulted extremely useful for the identification of anguillicolid nematodes to the species level. Furthermore, the sequences of 28S rDNA D2-D3 and *Cox1* made distinctions on the population levels of the different species possible. The phylogeny of the Anguillicolidae, based on the three barcoding genes, was inferred and is depicted in Figure 4.17. The division of the the Anguillicolidae

---

into the two genera *Anguillicola* and *Anguillicoloides*, as well as the monophyly of the species *Anguillicoloides novaezelandiae*, was questioned based on the findings of this study.

## 5. Conclusions

---

## 6. Suggestions for further work

Since the barcoding genes sequenced within the scope of this study resulted very useful for species identification and population structure analysis, their use is encouraged in general.

The evolutionary relationships among the basal taxa of clade 8 *sensu* Holterman et al. (2006) could be revealed through broader sampling of parasitic nematodes, especially within the superfamilies Seuratoidea and Gnathostomatoidea, paired with the generation of additional sequences of different barcoding genes through next-generation sequencing techniques.

In regard to anguillicolid nematodes, obtaining additional sequences from certain populations, e.g. *Anguillicoloides novaezelandiae* from New Zealand and *Anguillicoloides papernai* from South Africa, would allow to disprove or to corroborate certain hypothesis stated within the scope of this diploma thesis. By sequencing the 18S rDNA and 28S rDNA D2-D3 regions of *A. novaezelandiae* specimens from New Zealand, in addition to a detailed morphological analysis, the hypothesis, that these specimens constitute a cryptic species, could be tested. Regarding the hypothesis concerning the colonisation route of *A. papernai*, further specimens could be collected East and West of the sampling sites of this study and their 28S rDNA D2-D3 and *Cox1* sequences could be compared to the existing sequences.

The statistical parsimony network of *Anguillicoloides crassus* could be further analysed and compared to alternative networks constructed under different methods and parameters. In addition, the data from the population structure analyses calculated using *Arlequin* could be analysed further to evaluate the relationships among the populations of *A. crassus*.

The *Cox1* barcoding gene of additional specimens of *Anguillicola globiceps* could be sequenced in order to explain the problems encountered during this study. The hypothesis, that this phenomenon is caused by RNA editing, could be tested by conducting Real-Time PCRs (see Vanfleteren and Vierstraete, 1999).

Primers for the internal transcribed spacer (ITS2) region of the Anguillicolidae could be designed, to address further questions regarding the intra-specific relationships of the anguillicolid species. Furthermore, the indels contained in the 28S rDNA D2-D3 sequences of the anguillicolid species allow the design of species-specific PCR primers.

The co-evolution theory for the host-parasite-system of *Anguilla* infected with anguillicolid nematodes could be tested by experimental infections of *Anguilla borneensis* with *A. crassus*.



## **A. Appendix**

### **A.1. Perl programs**

Perl programs written within the scope of this diploma thesis are included in the data DVD (programs/).

Additionally, the author can be contacted through dominik.laetsch@gmail.com for questions or copies of the programs.

### **A.2. DNA Sequences and alignments**

All sequences (raw and processed) and alignments generated within the scope of this diploma thesis are included in the data DVD (sequences/).

### **A.3. Statistical parsimony networks**

All statistical parsimony networks and the associated files created within the scope of this diploma thesis are included in the data DVD (networks/).

### **A.4. Arlequin analyses**

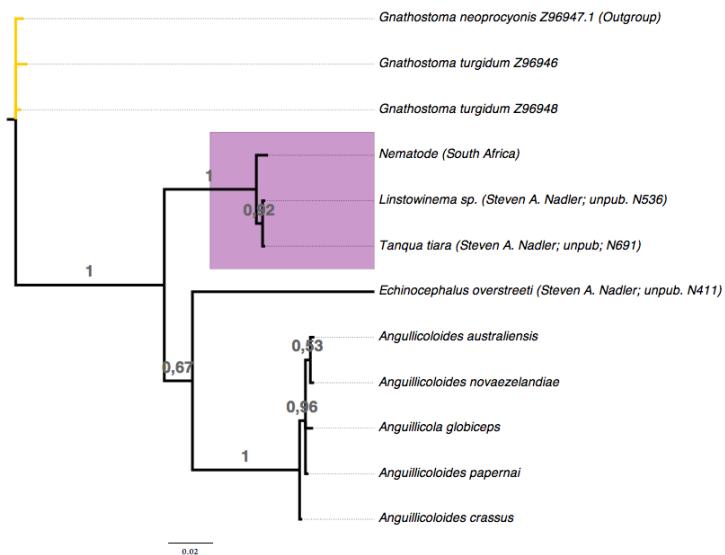
The results of the Arlequin analyses obtained within the scope of this diploma thesis are included in the data DVD (arlequin/).

### **A.5. Phylogenetic analysis**

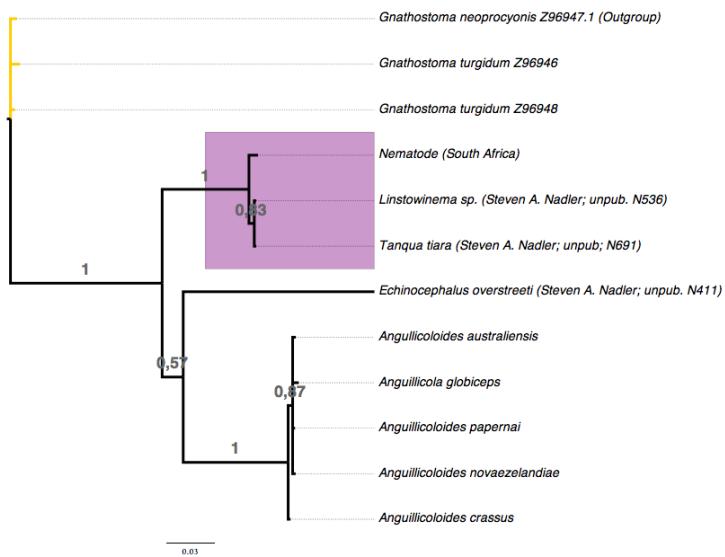
All datasets used in and all results obtained from the phylogenetic analyses can be found on the data DVD (phylogenetic-analysis/).

However, in the following subsections the phylogenograms of the analyses under alternative models, not shown in the result section, are depicted.

### A.5.1. 18S rDNA dataset

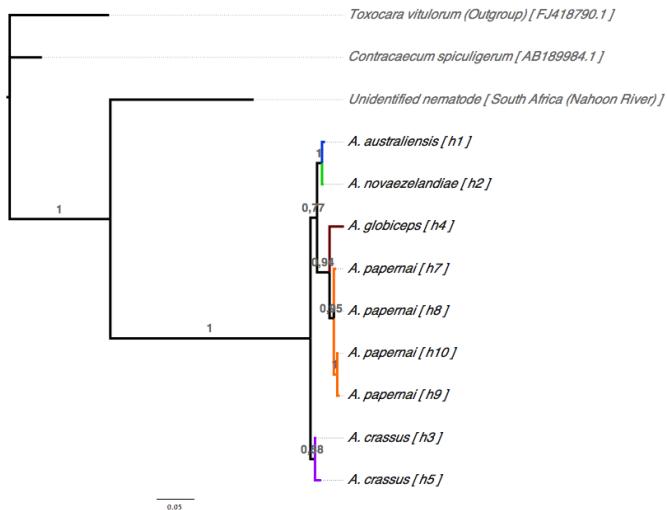


**Figure A.1.:** Phylogram of the phylogenetic analysis performed on the 18S rDNA sequences of the Anguillicolidae and seven other clade III nematode sequences *sensu* Blaxter et al. (1998) using Bayesian inference (MrBayes) under the SYM model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches.

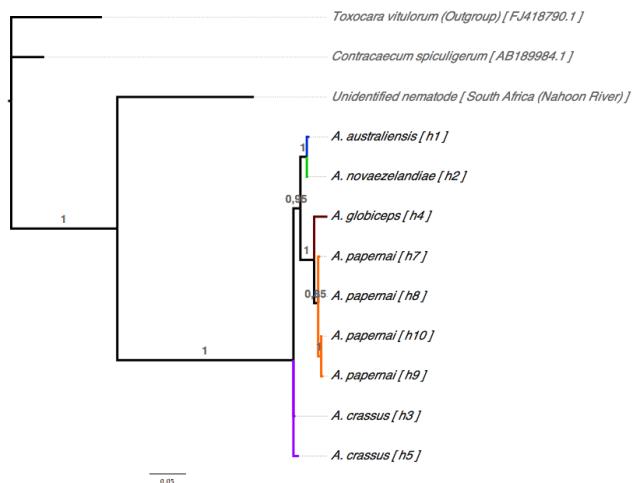


**Figure A.2.:** Phylogram of the phylogenetic analysis performed on the 18S rDNA sequences of the Anguillicolidae and seven other clade III nematode sequences *sensu* Blaxter et al. (1998) using Bayesian inference (MrBayes) under the GTR+ $\Gamma$  model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches.

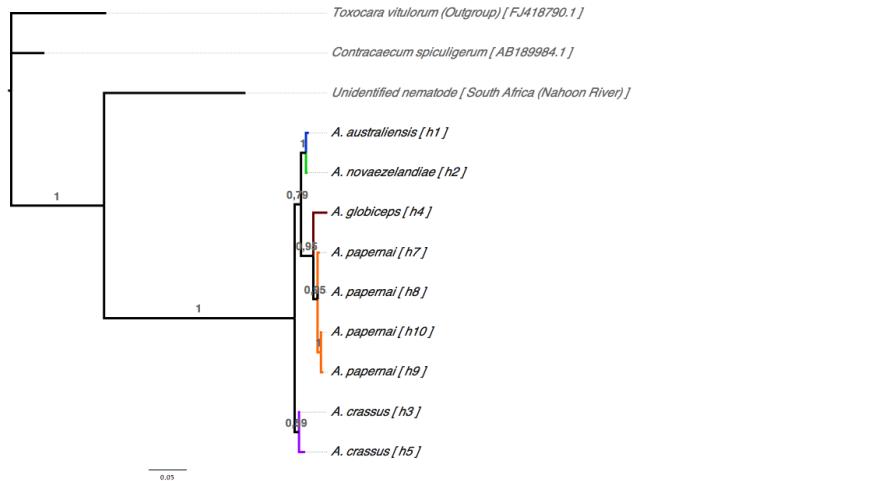
### A.5.2. 28S rDNA D2-D3 dataset



**Figure A.3.:** Phylogram of the phylogenetic analysis performed on the nine 28S rDNA D2-D3 anguillicolid haplotypes and three other clade III nematode sequences *sensu* Blaxter et al. (1998) using Bayesian inference (MrBayes) under the HKY+I+Γ model of DNA evolution. Indels were treated as gaps. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For each anguillicolid taxon the corresponding haplotype in the statistical parsimony network in Figure 4.3 is indicated in square brackets.

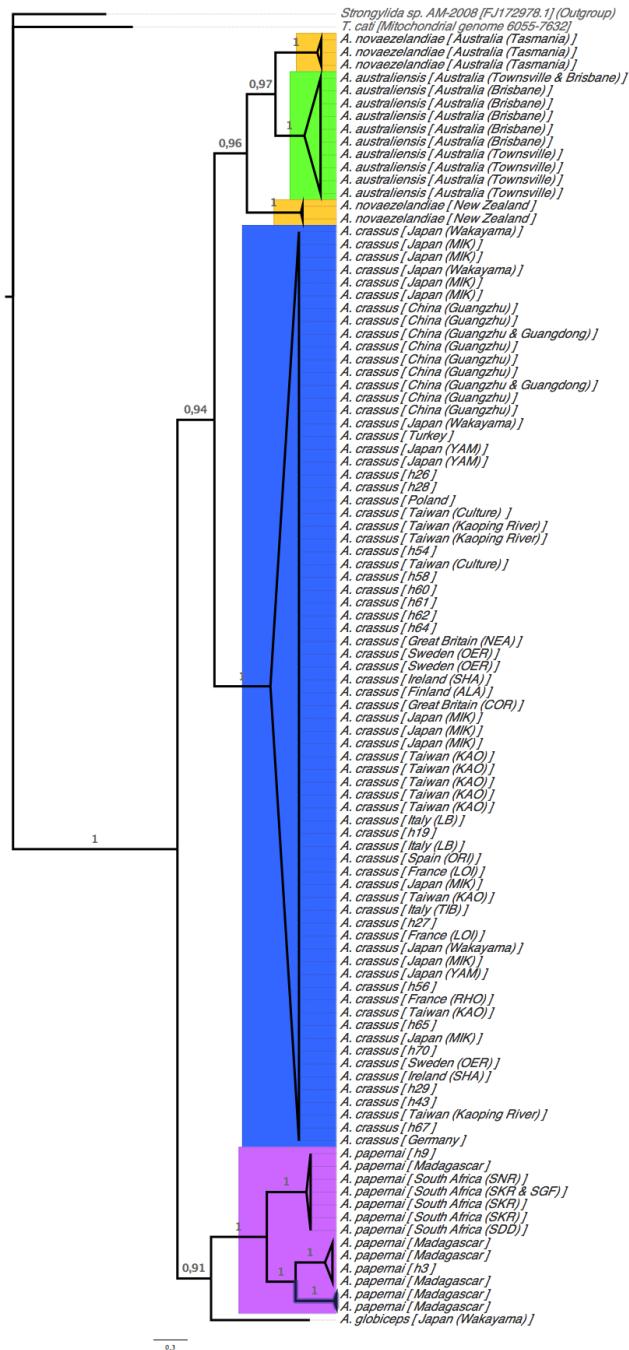


**Figure A.4.:** Phylogram of the phylogenetic analysis performed on the nine 28S rDNA D2-D3 anguillicolid haplotypes and three other clade III nematode sequences *sensu* Blaxter et al. (1998) using Bayesian inference (MrBayes) under the HKY+I+Γ model of DNA evolution. Indels were coded into the dataset as specified in section 3.9.4. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For each anguillicolid taxon the corresponding haplotype in the statistical parsimony network in Figure 4.3 is indicated in square brackets.



**Figure A.5.:** Phylogram of the phylogenetic analysis performed on the nine 28S rDNA D2-D3 anguillicolid haplotypes and three other clade III nematode sequences *sensu* Blaxter et al. (1998) using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. Indels were treated as gaps. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For each anguillicolid taxon the corresponding haplotype in the statistical parsimony network in Figure 4.3 is indicated in square brackets.

### A.5.3. Cox1 dataset



**Figure A.6:** Phylogram of the phylogenetic analysis performed on the 100 anguillicolid *Cox1* haplotypes and the sequences of two other nematode sequences using Bayesian inference (MrBayes) under the HKY+I+Γ model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For anguillicolid sequences found in more than one population the corresponding haplotype is indicated in square brackets and its geographical distribution can be checked in the respective statistical parsimony network in section 4.6.



## B. List of Figures

1.1 An overview of the Nematoda based on SSU phylogenetic analyses. Support values (in percentage) from Bayesian inference and LogDet transformation of distances are given for putative monophyletic groups. Trophic groups are shown as pictograms. Major taxa according to De Ley and Blaxter (2004) are highlighted in the same colour (from Meldal et al., 2007) . . . . .	3
1.2 Systematic classification of the Anguillicolidae <i>sensu</i> Moravec (2006). . . . .	7
1.3 Phylogeny of Spirurina <i>sensu</i> De Ley and Blaxter (2002) based on phylogenetic analyses of the 18S rDNA sequences (see Wijová et al., 2006). . . . .	13
1.4 The nuclear ribosomal DNA (rDNA) cistron. Sizes are approximate and not to scale. Each cistron comprises the small subunit rDNA gene (18S rDNA), the internal transcribed space I (ITS1), the 5.8S gene, ITS2 and the large subunit rDNA gene (28S rDNA). An external non-transcribed spacer (NTS) separates each transcribed cistron. The relative rate of sequence variation observed between taxa is shown schematically above the cistron (from Dorris et al., 1999). . . . .	16
3.1 Nomenclature of sequences illustrated at the example of the <i>CoxI</i> sequence of the second nematode sampled from Townsville, Australia. “AQT” is the population prefix, the specimen number is “2” and the gene suffix is “C” . . . . .	25
3.2 Graphical illustration of the framework for the determination of the consensus sequence <i>Seq<sub>c</sub></i> based on the Phred quality scores of the two sequences <i>A</i> and <i>B</i> . . . . .	25
3.3 Framework for likelihood mapping. <i>A</i> <sub>1</sub> , <i>A</i> <sub>2</sub> and <i>A</i> <sub>3</sub> show the tree-like regions. <i>A</i> <sub>12</sub> , <i>A</i> <sub>23</sub> and <i>A</i> <sub>13</sub> represent the net-like regions. <i>A</i> <sub>*</sub> displays the star-like region. (see Strimmer and von Haeseler (1997)) . . . . .	29
3.4 Example alignment from Simmons and Ochoterena (2000); <b>1</b> = gap from positions 3 to 9; <b>2</b> = gap from positions 6 to 11; <b>3</b> = gap from positions 14 to 25; <b>4</b> = gap from positions 14 to 16; <b>5</b> = gap from positions 19 to 22. . . . .	31
3.5 Simple indel coding data matrix of the alignment in Figure 3.4. “0” = gap is absent, “1” = gap is present, “?” = inapplicable (see Simmons and Ochoterena (2000)). . . . .	31
4.1 Statistical parsimony network constructed using all available anguillicolid 18S rDNA sequences . . . . .	43
4.2 Statistical parsimony network based on the ITS1 sequences of 32 specimens of <i>A. crassus</i>	45

4.3 Statistical parsimony network based on the 28S rDNA D2-D3 sequences of 136 anguillicolid specimens. The numbers in square brackets in the caption indicate the species: [1] = <i>A. australiensis</i> , [2] = <i>A. novaezelandiae</i> , [3] = <i>A. papernai</i> , [4] = <i>A. crassus</i> , [5] = <i>A. globiceps</i> . . . . .	46
4.4 The alignment of the 28S rDNA D2-D3 dataset starting at position 301 for each of the anguillicolid species showing variation in this segment. . . . .	46
4.5 Statistical parsimony network of the <i>Cox1</i> sequences from <i>A. australiensis</i> . . . . .	49
4.6 Statistical parsimony network of the <i>Cox1</i> sequences from <i>A. novaezelandiae</i> . . . . .	50
4.7 Statistical parsimony network of the <i>Cox1</i> sequences from <i>A. papernai</i> . The haplotypes of the 28S rDNA D2-D3 region are drawn as dashed rectangles around the respective <i>Cox1</i> haplotypes. For the specimens SKR-27a (“h7”), SDD-54 (“h11”) and SKR-41 (“h14”) no 28S rDNA D2-D3 sequences have been obtained. . . . .	51
4.8 Statistical parsimony network of the <i>Cox1</i> sequences from <i>A. crassus</i> . All <i>Cox1</i> haplotypes share the same 28S rDNA D2-D3 haplotype “h3”, except the four specimens JPN-K4B1 (“h32”), JPN-WK6 (“h34”), JPN-K4B4 and JPN-K4B4b (both “h33”) which show the 28S rDNA D2-D3 haplotype “h5”. . . . .	55
4.9 Plots from the results of the likelihood-mapping analysis showing the phylogenetic utility of the four molecular markers. The upper row shows the distribution of each of the possible quartets of taxa for each marker. The lower row indicates the proportion of quartets in each region. The regions at the vertices of each triangle ( $A_1$ , $A_2$ , $A_3$ ) represent the quartets that could be completely resolved, the central region ( $A_*$ ) indicates the percent of quartets that could not be resolved and the regions between the vertices of each triangle ( $A_{12}$ , $A_{23}$ , $A_{13}$ ) contain the quartets for which no distinction could be made between two trees. The black colouring of the upper plot of <i>Cox1</i> resides in the fact that $\binom{103}{4} = 4,421,275$ quartets were analysed. . . . .	56
4.10 Combined phylogram of the phylogenetic analysis performed on the 18S rDNA sequences from clade 8 nematodes <i>sensu</i> Holterman et al. (2006) using Bayesian inference (MrBayes) and maximum likelihood (RAxML) methods. The parameters for both analyses are listed in section 3.9.9 and 3.9.10. Bayesian posterior probabilities (BPP) followed by bootstrap values (BS) are displayed on the internal branches. Deviation in branching structure between the two trees is marked by *. Branches are collapsed whenever possible, based on taxonomic affiliations. . . . .	58

4.11 Phylogram of the maximum likelihood analysis performed, using RAxML. The exact parameters of the analysis are listed in section 3.9.10. Bootstrap supports (BS) are displayed on internal branches. The difference in branching order in comparison to the Bayesian inference tree is coloured in red. Other clade III nematodes = Camallanoidea, Dracunculoidea, Rhabdochonidae, Physalopteridae, Diplotriaenoidea, Filarioidea, Ascarididae, Rhigonematidae, Oxyuroidea, Atracidae. . . . .	59
4.12 Phylogram of the phylogenetic analysis performed on the 18S rDNA sequences of the Anguillicolidae and seven other clade III nematodes <i>sensu</i> Blaxter et al. (1998) using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. . . . .	61
4.13 Phylogram of the phylogenetic analysis performed on the nine anguillicolid 28S rDNA D2-D3 haplotypes together with three clade III nematodes <i>sensu</i> Blaxter et al. (1998) using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. Indels were coded as explained in section 3.9.4. The exact parameters for the analysis is listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For each anguillicolid taxon the corresponding haplotype in the statistical parsimony network in Figure 4.3 is indicated in square brackets. . . . .	63
4.14 Phylogram of the phylogenetic analysis performed on the 100 <i>Cox1</i> anguillicolid haplotypes and two other nematode sequences using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For anguillicolid sequences found in more than one population the corresponding haplotype is indicated in square brackets and its geographical distribution can be checked in the respective statistical parsimony network in section 4.6. . . . .	65
4.15 Phylogram of the phylogenetic analysis performed on the 100 <i>Cox1</i> anguillicolid haplotypes and two other nematode sequences using Bayesian inference (MrBayes) under the GTR+I+Γ model of DNA evolution. The third codon position was excluded from the analysis. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For anguillicolid sequences found in more than one population the corresponding haplotype is indicated in square brackets and its geographical distribution can be checked in the respective statistical parsimony network in section 4.6. . . . .	66
4.16 Phylogram of the phylogenetic analysis performed under the GTR+I+Γ model on the “supermatrix”. The parameters for the analysis are specified in section 3.9.9. The sequences used for the non-anguillicolid taxa are specified in brackets. . . . .	67
4.17 Cladogram of the evolutionary relationships among the family Anguillicolidae (Nematode: Anguillicoidea). . . . .	68

4.18 Cladograms of the genus <i>Anguilla</i> (left tree) (Minegishi et al., 2004) and its swim bladder parasites, the family Anguillicolidae (Nematoda: Anguillicoloidea) (right tree). Grey lines = traditional host-parasite relationships, displaying low abundances and low pathogenicity (parasite endemic in host); red lines = novel host-parasite relationships, displaying high abundances and pathogenicity; green lines = novel host-parasite relationship, displaying low abundances and pathogenicity; dashed lines = completion of life cycle has not been demonstrated. . . . .	69
4.19 Reconstruction of a possible co-evolution scenario for the genus <i>Anguilla</i> (blue) (Minegishi et al., 2004) and its swim bladder parasites, the family Anguillicolidae (Nematoda: Anguillicoloidea) (green). Red lines = novel host-parasite relationships displaying high abundances and pathogenicity; green lines = novel host-parasite relationship displaying low abundances and pathogenicity; dashed lines = completion of life cycle has not been demonstrated. . . . .	70
A.1 Phylogram of the phylogenetic analysis performed on the 18S rDNA sequences of the Anguillicolidae and seven other clade III nematode sequences <i>sensu</i> Blaxter et al. (1998) using Bayesian inference (MrBayes) under the SYM model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. . . . .	82
A.2 Phylogram of the phylogenetic analysis performed on the 18S rDNA sequences of the Anguillicolidae and seven other clade III nematode sequences <i>sensu</i> Blaxter et al. (1998) using Bayesian inference (MrBayes) under the GTR+Γ model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. . . . .	82
A.3 Phylogram of the phylogenetic analysis performed on the nine 28S rDNA D2-D3 anguillicolid haplotypes and three other clade III nematode sequences <i>sensu</i> Blaxter et al. (1998) using Bayesian inference (MrBayes) under the HKY+I+Γ model of DNA evolution. Indels were treated as gaps. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For each anguillicolid taxon the corresponding haplotype in the statistical parsimony network in Figure 4.3 is indicated in square brackets. . . . .	83
A.4 Phylogram of the phylogenetic analysis performed on the nine 28S rDNA D2-D3 anguillicolid haplotypes and three other clade III nematode sequences <i>sensu</i> Blaxter et al. (1998) using Bayesian inference (MrBayes) under the HKY+I+Γ model of DNA evolution. Indels were coded into the dataset as specified in section 3.9.4. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For each anguillicolid taxon the corresponding haplotype in the statistical parsimony network in Figure 4.3 is indicated in square brackets. . . . .	83

A.5 Phylogram of the phylogenetic analysis performed on the nine 28S rDNA D2-D3 anguillicolid haplotypes and three other clade III nematode sequences <i>sensu</i> Blaxter et al. (1998) using Bayesian inference (MrBayes) under the GTR+I+ $\Gamma$ model of DNA evolution. Indels were treated as gaps. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For each anguillicolid taxon the corresponding haplotype in the statistical parsimony network in Figure 4.3 is indicated in square brackets. . . . .	84
A.6 Phylogram of the phylogenetic analysis performed on the 100 anguillicolid <i>Cox1</i> haplotypes and the sequences of two other nematode sequences using Bayesian inference (MrBayes) under the HKY+I+ $\Gamma$ model of DNA evolution. The exact parameters for the analysis are listed in section 3.9.9. Bayesian posterior probabilities (BPP) are displayed on internal branches. For anguillicolid sequences found in more than one population the corresponding haplotype is indicated in square brackets and its geographical distribution can be checked in the respective statistical parsimony network in section 4.6. . . . .	85



## C. List of Tables

1.1	Geographical distribution of the major medically relevant nematode species and the estimated number of infections (in millions) caused in Humans. (after Crompton, 1999; De Silva et al., 2003; Lukeš et al., 2005) . . . . .	6
1.2	Geographic distribution and qualitative anguillicolid parasite burden of the 15 species of the genus <i>Anguilla</i> (After Minegishi et al., 2004). Anguillicolid parasites: A = <i>A. australiensis</i> , C = <i>A. crassus</i> , G = <i>A. globiceps</i> , N = <i>A. novaezelandiae</i> , P = <i>A. papernai</i> ; (+) = high abundance, (-) = low abundance. . . . .	10
3.1	Sampling location for the nematodes including label prefix, information on geographic position including latitude ( $\phi$ ) and longitude ( $\lambda$ ) of the sampling site, host species and information about sampling. AK : Albert Keim, BS : Björn Schäffner, EH : Emanuel Heitlinger, EG : Ercüment Genç, HT : Horst Taraschewski, JLC : José Lino Costa, LP : Lea Perseke, OW : Olaf Weyl, PMR : Pilar Muñoz Ruíz, HS : Hiroshi Sato, . . . . .	19
3.2	PCR primers used for amplification of the barcoding genes. . . . .	21
3.3	Cycling profiles of the different PCR reactions. . . . .	22
3.4	Sequences included as outgroup taxa in the phylogenetic analyses of the different datasets (18S rDNA, 28S rDNA, <i>CoxI</i> and the “supermatrix”) . . . . .	29
3.5	Results from the calculation of the approximate AIC (Akaike Information Criterion) values and the hierarchical likelihood ratio tests (hLRT) performed by MrModeltest2 for the different datasets. . . . .	30
3.6	Critical values for Bayes factors comparison (see Kass and Raftery, 1995). Comparisons were performed using $\ln(B_{1,2})$ . . . . .	34
3.7	Interpretations of the support values calculated for branch phylogenetic trees. BS = bootstrap values, BP = Bayesian posterior probabilities. . . . .	37
3.8	Ranges of values used for the interpretation of $\phi_{st}$ (Wright, 1978). . . . .	40
4.1	Number of samples in which actinopterygic DNA was sequenced (“Host DNA count”) for each of the primer pairs, based on BLAST searches. Numbers in parenthesis represent non-nematode sequences. . . . .	41
4.2	Partial 18S rDNA type sequences of the Anguillicolidae . . . . .	42

4.3	Molecular diversity indices calculated using Arlequin for the populations of <i>A. papernai</i> . Numbers in brackets refer to the number of sites of which substitutions are observed. The order of the populations was according to their distance from Madagascar and abbreviated according to Table 3.1: MAD = Madagascar, SNR = Nahoon River, SGF = Great Fish River, SKR = Koonap River, SDD = Sunday's River (Darlington Dam), SSD = Sunday's River (Slagboom Dam). . . . .	52
4.4	Table of the pairwise $\phi_{st}$ - values calculated using Arlequin for the populations of <i>A. papernai</i> . - = 0.000, (-) = negative $\phi_{st}$ - values, * = significant P-values. . . . .	52
4.5	Table of the geographical grouping applied to the populations of <i>A. crassus</i> prior to the calculation of the statistical parsimony network. Grouping was performed according to Wielgoss et al. (2008). . . . .	54
4.6	Summary of ln Bayes Factor ( $\ln(B)$ ) analyses of alternative models for the 18S rDNA dataset comprising all anguillicolid sequences, as well as seven other clade III nematodes <i>sensu</i> Blaxter et al. (1998). THM = Total harmonic means of the negative log likelihoods of the MrBayes runs. . . . .	61
4.7	Summary of ln Bayes Factor ( $\ln(B)$ ) analyses of alternative models for the 28S rDNA D2-D3 dataset. THM = Total harmonic mean of the negative log likelihoods of the MrBayes runs, (SIC) = Simple Indel Coding (see section 3.9.4) . . . . .	62
4.8	Summary of ln Bayes Factor ( $\ln(B)$ ) analyses of alternative models for the <i>coxI</i> dataset. THM = Total harmonic mean of the negative log likelihoods of the MrBayes runs. . . . .	64

## D. Bibliography

- Abawi, G. S. and T. L. Widmer (2000). Impact of soil health management practices on soilborne pathogens, nematodes and root diseases of vegetable crops. *Applied Soil Ecology* 16, 37–47.
- Aleshin, V., O. S. Kedrova, I. A. Milyutina, N. S. Vladychenskaya and N. B. Petrov (1998). Relationships among nematodes based on the analysis of 18S rRNA gene sequences: molecular evidence for monophyly of chromadorian and secernentian nematodes. *Russian Journal of Nematology* 6.
- Altschul, S., W. Gish, W. Miller, E. Myers and D. Lipman (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* 215, 403–410.
- Anderson, R. C., A. G. Chabaud and S. Wilmott (1975). *CIH keys to the nematode parasites of vertebrates*. Farnham Royal, UK.
- Ando, K., H. Tokura, H. Matsuoka, D. Taylor and Y. Chinzei (1992). Life cycle of *Gnathostoma nipponicum* Yamaguti, 1941. *Journal of Helminthology* 66, 53–61.
- Andrássy, I. (1976). *Evolution as a basis for the systematization of nematodes*. Pitman Publishing Ltd.
- Arai, T., T. Otake and K. Tsukamoto (2000). Timing of metamorphosis and larval segregation of the Atlantic eels *Anguilla rostrata* and *A. anguilla*, as revealed by otolith microstructure and microchemistry. *Marine Biology* 137, 39–45.
- Barse, A. M. and D. H. Secor (1999). An exotic nematode parasite of the American eel. *Fisheries* 24, 6–10.
- Blaxter, M. L. (2003). Counting angels with DNA. *Nature* 421, 122–124.
- Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman, A. Vierstraete, J. R. Vanfleteren, L. Y. Mackey, M. Dorris and L. M. Frisse (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature* 392, 71–75.
- Blaxter, M. L. and R. Roche (2005). The small subunit ribosomal RNA sequence of *Anguillicola crassus*. GenBank<sup>TM</sup> entry DQ118535.1.

- Blouin, M. S., C. A. Yowell, C. H. Courtney and J. B. Dame (1998). Substitution bias, rapid saturation and the use for mtDNA for nematode systematics. *Molecular Biology and Evolution* 15, 1719–1727.
- Bolliet, V., P. Lambert, J. Rives and A. Bardonnet (2007). Rhythmic swimming activity in *Anguilla anguilla* glass eels: Synchronisation to water current reversal under laboratory conditions. *Journal of Experimental Marine Biology and Ecology* 344, 54–66.
- Bovien, P. (1937). *Some types of association between nematodes and insects*. Bianco Luno (Copenhagen).
- Carvajal, J., L. González and S. Toledo (1995). New record of *Hysterotylacium aduncum* (Rudolphi, 1802) (Nematoda: Anisakidae) in salmonids cultured in sea farms from southern Chile. *Research and Reviews in Parasitology* 55, 195–197.
- Chabaud, A. G. and O. Bain (1994). The evolutionary expansion of the Spirurida. *International Journal for Parasitology* 24, 1179–1201.
- Charlesworth, B., P. Sniegowski and W. Stephan (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, 215–220.
- Chilton, N., H. Hoste, G. Hung, I. Beveridge and R. Gasser (1997a). The 5.8 S rDNA sequences of 18 species of bursate nematodes (order Strongylida): comparison with rhabditid and tylenchid nematodes. *International Journal for Parasitology* 27, 119–124.
- Chilton, N. B., R. B. Gasser and I. Beveridge (1995). Differences in ribosomal DNA sequence of morphologically indistinguishable species within the *Hypodontus macropi* complex (Nematoda: Strongyoidea). *International Journal for Parasitology* 25, 647–651.
- Chilton, N. B., R. B. Gasser and I. Beveridge (1997b). Phylogenetic relationships of Australian strongyloid nematodes inferred from ribosomal DNA sequence data. *International Journal for Parasitology* 27, 1481–1494.
- Chitwood, B. and M. Chitwood (1950). *An Introduction to Nematology*. Monumental Printing Company, Baltimore, MD, second Ed.
- Chitwood, B. G. (1933). A revised classification of the Nematoda. *Journal of Parasitology* 20, 131.
- Chitwood, B. G. (1937). *A revised classification of the Nematoda*. Papers on helminthology. All-Union Lenin Academy of Agricultural Sciences, Moscow, Russia.
- Chitwood, B. G. (1958). The designation of official names for higher taxa of invertebrates. *Bulletin of Zoological Nomenclature* 15.

- Clement, M., D. Posada and K. A. Crandall (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology* 9, 1657–1660.
- Colautti, R. I. and H. J. MacIsaac (2004). A neutral terminology to define invasive species. *Diversity and Distributions* 10.
- Crandall, K. A. and A. R. Templeton (1993). Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134, 959—969.
- Crofton, H. D. (1971). A model of host-parasite relationships. *Parasitology* 63, 343–364.
- Crompton, D. W. T. (1999). How much Human helminthiasis is there in the World? *Journal of Parasitology* 85, 397–403.
- De Charleroy, D., L. Grisez, K. Thomas, C. Belpaire and F. Ollevier (1990). The life cycle of *Anguillicola crassus*. *Diseases of Aquatic Organisms* 8, 77–84.
- De Ley, P. and M. L. Blaxter (2002). *The Biology of Nematodes*, Chap. Systematic position and phylogeny. Taylor & Francis (London), 1–30.
- De Ley, P. and M. L. Blaxter (2004). A new system for the Nematoda: combining morphological characters with molecular trees, and translating clades into ranks and taxa. *Proceedings of the fourth International Congress of Nematology*, (Eds.) R. Cook and D. J. Hunt. Brill (Leiden, Netherlands), Tenerife, Spain, 865.
- De Ley, P., I. T. De Ley, K. Morris, E. Abebe, M. Mundo-Ocampo, M. Yoder, J. Heras, D. Wau-mann, A. Rocha-Olivares, A. h. Jay Burr, J. G. Baldwin and W. Thomas (2005). An integrated approach to fast and informative morphological vouchering of nematodes for applications in molecular barcoding. *Philosophical Transactions of the Royal Society. Series B: Biological Sciences* 360, 1945–1958.
- De Silva, N. R., S. Brooker, P. J. Hotez, A. Montresor, D. Engels and L. Savioli (2003). Soil-transmitted helminth infection: updating the global picture. *Trends in Parasitology* 12, 547–551.
- Dobson, A., K. D. Lafferty, A. M. Kuris, R. F. Hechinger and W. Jetz (2008). Hommage to Linnaeus : How many parasites ? How many hosts ? *Proceedings of the National Academy of Sciences* 105, 11482–11489.
- Dorris, M., P. De Ley and M. L. Blaxter (1999). Molecular Analysis of Nematode Diversity and the Evolution of Parasitism. *Parasitology Today* 15, 188 – 193.
- Dougherty, E. C. (1951). Evolution of Zooparasitic Groups in the Phylum Nematoda, with Special Reference to Host-Distribution. *Journal of Parasitology* 37, 353–378.

- Douglas, A. E. (1994). *Symbiotic Interactions*. Oxford Science Publications.
- Drown, D. (2004). *Arlequin: Step by step instructions*. Barber Lab, Dept. of Ecology and Evolutionary Biology, UCLA.
- Dunn, C. W., A. Hejnol, D. Q. Matus, K. Pang, W. E. Browne, S. A. Smith, E. Seaver, G. W. Rouse, M. Obst, G. D. Edgecombe et al. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452, 745–749.
- Ege, V. (1981). *A revision of the genus Anguilla Shaw: a systematic, phylogenetic and geographical study*. Oxford Univ Press.
- Egusa, S. (1978). Anguillicolosis of Eels. *Infectious Diseases of Fish* , 501 – 506.
- El Hilali, M., A. Yahyaoui, A. Sadak, M. Maachi and Z. Taghy (1996). Première données épidémiologiques sur l'anguillicolose au Maroc. *Bulletin Français de la Pêche et de la Pisciculture* 340, 57–60.
- Ellis, R. E., J. E. Sulston and A. R. Coulson (1986). The rDNA of *C. elegans*: sequence and structure. *Nucleic Acids Research* 14, 2345–2364.
- Elton, C. (1927). *Animal Ecology*. Macmillan Co. (New York).
- Erixon, P., B. Svensson, T. Britton and B. Oxelman (2003). Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Systematic Biology* 52, 665–673.
- Excoffier, L. and H. L. Lischer (2010). Arlequin suite Version 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources (In press)* .
- Excoffier, L., P. Smouse and J. Quattro (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- FAO Report (2001). Aquaculture production: quantities 1970-2000. Food and Agriculture Organization of the UN. URL <ftp://ftp.fao.org/fi/stat/windows/fishplus/aquaq.zip>, <ftp://ftp.fao.org/fi/stat/windows/fishplus/>
- FAO Report (2002). Commodities trade and production of fishery products 1976-1999. The Food and Agricultural Organization of the UN. URL <ftp://ftp.fao.org/fi/stat/windows/fishplus/fishcomm.zip>, [ftp://ftp.fao.org/fi/stat/windows/fishplus/fst\\_plus.zip](ftp://ftp.fao.org/fi/stat/windows/fishplus/fst_plus.zip).
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* 17, 368–376.

- Felsenstein, J. (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution*, 783–791.
- Felsenstein, J. (2004a). *Inferring Phylogenies*. Sinauer Associates, Inc. (Massachusetts).
- Felsenstein, J. (2004b). Phylip (phylogeny inference package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.
- Floyd, R., E. Abebe, A. Papert and M. L. Blaxter (2002). Molecular barcodes for soil nematode identification. *Molecular Ecology* 11, 839–850.
- Folmer, O., M. Black, W. Hoeh, R. Lutz and R. Vrijenhoek (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3, 294.
- Fries, L., D. Williams and S. Johnson (1996). Occurrence of *Anguillicola crassus*, an Exotic Parasitic Swim Bladder Nematode of Eels, in the Southeastern United States. *Transactions of the American Fisheries Society* 125, 794–797.
- Gasser, R., N. Chilton, H. Hoste and I. Beveridge (1993). Rapid sequencing of rDNA from single worms and eggs of parasitic helminths. *Nucleic Acids Research* 21, 2525–2525.
- Geiß, K. and B. Sures (2010a). Cytochrome Oxidase I sequences of *Anguillicoloides novaezelandiae* from New Zealand. Pers. comm. (unpub.).
- Geiß, K. and B. Sures (2010b). The swim bladder nematode *A. novaezelandiae*. *Proceedings of the Joint Meeting of the German Societies of Parasitology and Protozoology, Düsseldorf, Germany*.
- Gelman, A. and D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–472.
- Genç, E., M. K. Sangun, M. Dural, M. F. Can and C. Altunhan (2008). Element concentrations in the swimbladder parasite *Anguillicola crassus* (Nematoda) and its host the European eel, *Anguilla anguilla* from Asi River (Hatay, Turkey). *Environmental Monitoring and Assessment* 141, 59–65.
- Gillooly, J. F., A. P. Allen, G. B. West and J. H. Brown (2005). The rate of DNA evolution: effects of body size and temperature on the molecular clock. *Proceedings of the National Academy of Sciences* 102.
- Grieksma, A. and T. Groothuis (2006). 4Peaks. Version 1.7.2. <http://mekentosj.com>.
- Griesbach, J. A. and A. R. Maggenti (1989). Vector capability of *Xiphinema americanum* sensu lato in California. *Journal of Nematology* 21, 517–523.

- Hanke, M. and M. Wink (1994). Direct DNA sequencing of PCR-amplified vector inserts following enzymatic degradation of primer and dNTPs. *Biotechniques* 17, 858.
- Harrington, N. (2007). Create custom data charting tools using Perl and GD. Techn. Rep., Technical Library (IBM).
- Hassouna, N., B. Michot and J. P. Bachellerie (1984). The complete nucleotide sequence of mouse 28S rRNA gene. Implications for the process of size increase of the large subunit rRNA in higher eukaryotes. *Nucleic Acids Research* 12, 3563–3583.
- Hebert, P. D. N., A. Cywinska, S. L. Ball and J. R. de Waard (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society. Series B: Biological Sciences* 270, 313–321.
- Heitlinger, E. H., D. R. Laetsch, U. Weclawski, Y. Han and H. Taraschewski (2009). Massive encapsulation of larval *Anguillicoloides crassus* in the intestinal wall of Japanese eels. *BMC Parasites and Vectors*.
- Hillis, D. M. and J. J. Bull (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analyses. *Systematic Biology* 42, 182–192.
- Hirose, H., T. Yabu, I. Hirono and T. Aoki (1998). The phylogeny of *Anguillicola crassus* and *A. globiceps* based on partial 18S ribosomal RNA sequences. *Journal of Fish Diseases* 21, 265–271.
- Holterman, M., A. Van Der Wurff, S. Van Den Elsen, H. Van Megen, T. Bongers, O. Holovachov, J. Bakker and J. Helder (2006). Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Molecular Biology and Evolution* 23, 1792–1800.
- Huelsenbeck, J. P. and F. Ronquist (2005). *Statistical Methods in Molecular Evolution*, Chap. Bayesian analysis of molecular evolution using MrBayes. Springer, New York, 183–232.
- Hugot, J.-P., P. Baujard and S. Morand (2001). Biodiversity in helminths and nematodes as a field of study: an overview. *Nematology* 3, 199–208.
- Hurst, G. D. D. and F. M. Jiggins (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society. Series B: Biological Sciences* 272, 1525–1534.
- Ivashkin, V. M., A. A. Sobolev and L. A. Khromoca (1971). Camallanata of Animals and Man and the Diseases Caused by Them. *Osnovy Nematologii* 22, 388.

- Johnston, T. H. and P. Mawson (1940). Some nematodes parasitic in Australian freshwater fish. *Transactions of the Royal Society of South Australia* 64.
- Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773–795.
- Kennedy, C. R. (1994). The distribution and abundance of the nematode *Anguillicola australis* in eels *Anguilla reinhardtii* in Queensland, Australia. *Folia Parasitologica* 41, 279–285.
- Kennedy, C. R. (2007). The pathogenic helminth parasites of eels. *Journal of Fish Diseases* 30, 319–334.
- Kettle, A. and K. Haines (2006). How does the European eel (*Anguilla anguilla*) retain its population structure during its larval migration across the North Atlantic Ocean? *Journal of Fisheries and Aquatic Sciences* 63, 90–106.
- Kiontke, K. and W. Sudhaus (2006). *Ecology of Caenorhabditis species*. The *C. elegans* Research Community, Wormbook.
- Kirk, R. S. (2003). The impact of *Anguillicola crassus* on European eels. *Fisheries Management and Ecology* 10, 385–394.
- Kleckner, R. C. and J. D. McCleave (1988). The Northern limit of spawning by Atlantic eels (*Anguilla spp.*) in the Sargasso Sea in relation to the thermal fronts and surface water masses. *Journal of Marine Research* 46, 647–667.
- Koenning, S. R., C. Overstreet, J. W. Noling, P. A. Donald, J. O. Becker and B. A. Fortnum (1999). Survey of crop losses in response to phytoparasitic nematodes in the United States for 1994. *Supplement to the Journal of Nematology* 31, 587–618.
- Køie, M. (1991). Swimbladder nematodes (*Anguillicola spp.*) and gill monogeneans (*Pseudodactylogyrus spp.*) parasitic on the European eel (*Anguilla anguilla*). *Journal du Conseil International pour l'exploration de la Mer* 476, 391–398.
- Koops, H. and F. Hartmann (1989). *Anguillicola*-Infestations in Germany and in German eel imports. *Journal of Applied Ichthyology* 1, 41–45.
- Kosakovsky Pond, S. L., S. D. W. Frost and S. V. Muse (2005). Hyphy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679.
- Kristmundsson, Á. and S. Helgason (2007). Parasite communities of eels *Anguilla anguilla* in freshwater and marine habitats in Iceland in comparison with other parasite communities of eels in Europe. *Folia Parasitologica* 54, 141–153.

- Lafferty, K. D., A. P. Dobson and A. M. Kuris (2006). Parasites dominate food web links. *Proceedings of the National Academy of Sciences* 30, 11211–11216.
- Lambson, P. J. D. and G. Boucher (2003). Marine nematode deep-sea biodiversity — hyper-diverse or hype ? *Journal of Biogeography* 30, 475–485.
- Lambson, P. J. D., J. Tietjen, T. Ferrero and P. Jensen (2000). Latitudinal diversity gradients in the deep sea with special reference to North Atlantic nematodes. *Marine Ecology Progress Series* 194, 159–167.
- Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGgettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez et al. (2007). Clustal W and Clustal X Version 2.0. *Bioinformatics* 23, 2947.
- Lecomte-Finiger, R. (1994). The early life of the European eel. *Nature* 370, 424.
- Lecomte-Finiger, R. (2003). The genus *Anguilla* Schrank, 1798: current state of knowledge and questions. *Reviews in Fish Biology and Fisheries* 13, 265–279.
- Lefebvre, F., P. Contournet and A. J. Crivelli (2007). Interaction between the severity of the infection by the nematode *Anguillicola crassus* and the tolerance to hypoxia in the European eel *Anguilla anguilla*. *Acta Parasitologica* 52, 171–175.
- Lemire, B. (2005). *Mitochondrial genetics*. The *C. elegans* Research Community, Wormbook.
- Lewin, B. (2008). *Genes IX*. Jones and Bartlett (London).
- Lorenzen, S. (1981). Entwurf eines phylogenetischen Systems der freilebenden Nematoden. *Veröffentlichungen des Institut für Meeresforschungen Bremerhaven Supplement* 7, 1–472.
- Lu, L. H., M. R. Lin, W. M. Choi, K. P. Hwang, Y. H. Hsu, M. J. Bair, J. D. Liu, T. E. Wang, T. P. Liu and W. C. Chung (2006). Human intestinal capillariasis (*Capillaria philippinensis*) in Taiwan. *The American Journal of Tropical Medicine and Hygiene* 74, 810–813.
- Lukeš, J., A. Horák and T. Scholz (2005). Helminth genome projects: all or nothing. *Trends in Parasitology* 21, 265–266.
- Maamouri, F., L. Gargouri, M. Ould Daddah and G. Bouix (1999). Occurrence of *Anguillicola crassus* (Nematoda, Anguillicolidae) in the Ichkeul Lake (Northern Tunisia). *Bulletin of The European Association of Fish Pathologists* 19, 17–19.
- Maddison, W. P. and D. R. Maddison (2009). Mesquite: a modular system for evolutionary analysis. Version 2.6. <http://mesquiteproject.org>.

- McGuire, J. M. (1973). Retention of tobacco ringspot virus by *Xiphinema americanum*. *Phytopathology* 63, 324–326.
- Meldal, B. H. M., N. J. Debenham, P. De Ley, I. T. De Ley, J. R. Vanfleteren, A. Vierstraete, W. Bert, G. Borgonie, T. Moens, P. A. Tyler, M. C. Austen, M. L. Blaxter, A. D. Rogers and P. J. D. Lambshead (2007). An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa. *Molecular Phylogenetics and Evolution* 42, 622–636.
- Merkle, D., M. Middendorf and N. Wieseke (2009). *CoRe-PA – Softwaretool for Cophylogeny Reconstruction: A Short Tutorial*. Dept. of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark.
- Minegishi, Y., J. Aoyama, J. G. Inoue, M. Miya, M. Nishida and K. Tsukamoto (2004). Molecular phylogeny and evolution of the freshwater eels genus *Anguilla* based on the whole mitochondrial genome sequences. *Molecular Phylogenetics and Evolution* 34, 134–146.
- Miscampbell, A., M. Lankester and M. Adamson (2004). Molecular and morphological variation within swim bladder nematodes, *Cystidicola* spp. *Canadian Journal of Fisheries and Aquatic Sciences* 61, 1143–1152.
- Molnár, K., F. Baska, G. Csaba, R. Glávits and C. Székely (1993). Pathological and histopathological studies of the swimbladder of eels *Anguilla anguilla* infected by *Anguillicola crassus* (Nematoda: Dracunculoidea). *Diseases of Aquatic Organisms* 15, 41–50.
- Moravec, F. (2004). Some aspect of the taxonomy and biology of dracunculoid nematodes parasitic in fishes: a review. *Folia Parasitologica* 51, 1–13.
- Moravec, F. (2006). *Dracunculoid and anguillicoloid nematodes parasitic in vertebrates*. Academia (Prague).
- Moravec, F., D. Di Cave, P. Orecchia and L. Paggi (1994). Present occurrence of *Anguillicola novaezelandiae* (Nematoda: Dracunculoidea) in Europe and its development in the intermediate host. *Folia Parasitologica* 41, 203–208.
- Moravec, F. and K. Rohde (1992). Three species of nematodes of the superfamily Dracunculoidea from Australian fishes. *Acta Societas Zoologicae Bohemicae* 56, 187–195.
- Moravec, F. and Škoríková (1988). Amphibians and larvae of aquatic insects as new paratenic hosts of *Anguillicola crassus* (Nematoda: Dracunculoidea), a swimbladder parasite of eels. *Diseases of Aquatic Organisms* 34, 217–222.
- Moravec, F. and H. Taraschewski (1988). Revision of the genus *Anguillicola* Yamaguti, 1935 (Nematoda: Anguillicolidae) of the swimbladder of eels, including descriptions of two new species, *A. novaezelandiae* sp. n. and *A. papernai* sp. n. *Folia Parasitologica* 35, 125–146.

Moriarty, C. and W. Dekker (1997). Management of the European Eel. *Fisheries Bulletin (Dublin)*, 1–125.

Muchiut, S., F. Gallet, D. Aubin, L. Baranger, V. Le Bihan and Y. Perraudeau (2002). Principaux facteurs à prendre en compte pour une meilleure gestion de l'anguille européenne *Anguilla anguilla*. Techn. Rep., AGLIA, Observatoire des Pêches et des Cultures Marines du Golfe de Gascogne, France.

Müller, K. (2005). SeqState - primer design and sequence statistics for phylogenetic DNA data sets. *Applied Bioinformatics* 4, 65–69.

Nadler, S. A. (2010). 18s sequences of spirurine nematodes. Pers. comm. (unpub.).

Nadler, S. A., R. A. Carreno, H. Mejía-Madrid, J. Ullberg, C. Pagan, R. Houston and J.-P. Hugot (2007). Molecular phylogeny of clade III nematodes reveals multiple origins of tissue parasitism. *Parasitology* 134, 1421–1442.

Nadler, S. A., P. De Ley, M. Mundo-Ocampo, A. B. Smythe, S. P. Stock, D. Bumbarger, B. J. Adams, I. T. De Ley, O. Holovachov and J. G. Baldwin (2006). Phylogeny of Cephalobina (Nematoda): Molecular evidence for recurrent evolution of probolae and incongruence with traditional classifications. *Molecular Phylogenetics and Evolution* 40, 696–711.

Nadler, S. A. and D. S. S. Hudspeth (2000). Phylogeny of the Ascaridoidea (Nematode: Ascaridida) based on three genes and morphology: hypotheses of structural and sequence evolution. *Journal of Parasitology* 86, 380–393.

Nagasawa, K., Y.-G. Kim and H. Hirose (1994). *Anguillicola crassus* and *A. globiceps* (Nematoda: Dracunculoidea) parasitic in the swimbladder of eels (*Anguilla japonica* and *A. anguilla*) in East Asia: a review. *Folia Parasitologica* 41, 127–137.

Nei, M. and W.-H. Li (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences* 76, 5269–5273.

Neumann, W. (1985). Schwimmblasenparasit *Anguillicola* bei Aalen. *Fischer und Teichwirt* 36, 322.

Newton, M. A. and A. E. Raftery (1991). Approximate bayesian inference by the weighted likelihood bootstrap. Techn. Rep., Department of Statistics, University of Washington, Seattle, Washington 98195 USA.

Nielsen, T. and P. Prouzet (2008). Capture-based aquaculture of the wild European eel (*Anguilla anguilla*): a global overview. Techn. Rep. 508, FAO Fisheries Technical Paper.

- Nunn, G. B. (1992). *Nematode Molecular Evolution : An Investigation of Evolutionary Patterns Among Nematodes Based on DNA Sequences*. PhD Thesis, University of Nottingham, Nottingham.
- Nylander, J. A. A. (2004). MrModeltest Version 2. Evolutionary Biology Centre, Uppsala University.
- Osche, G. (1956). Die Praäadaptation freilebender Nematoden an den Parasitismus. *Zoologischer Anzeiger* 19 (Suppl.), 391—396.
- Paggi, L., P. Orecchia, R. Minervini and S. Mattiucci (1982). Sulla comparsa di *Anguillicolaustraliensis* Johnston e Mawson, 1940 (Dracunculoidea: Anguillicolidae) in *Anguilla anguilla* del Lago di Bracciano. *Parasitologia* 24, 139–144.
- Palstra, A. P., D. F. M. Heppener, V. J. T. van Ginneken, C. Székely and G. E. E. J. M. van den Thillart (2007). Swimming performance of silver eels is severely impaired by the swim-bladder parasite *Anguillicola crassus*. *Journal of Experimental Marine Biology and Ecology* 352, 244–256.
- Platt, H. (1994). *The phylogenetic systematics of free-living nematodes*, Chap. Foreword. The Royal Society (London).
- Powers, T. (2004). Nematode molecular diagnostics: from bands to barcodes. *Annual Review of Phytopathology* 42, 367–383.
- Price, P. W. (1980). *Evolutionary Biology of Parasites*. Princeton University Press (Princeton).
- Rambaut, A. (2006–2009). FigTree: Tree Figure Drawing Tool Version 1.3.1. <http://tree.bio.ed.ac.uk/>.
- Rambaut, A. and A. J. Drummond (2007). Tracer v1.4. <http://beast.bio.ed.ac.uk/Tracer>.
- Reimer, L. W., A. Hildebrand, D. Scharberth and U. Walter (1994). *Anguillicola crassus* in the Baltic Sea: Field data supporting transmission in brackish waters. *Diseases of Aquatic Organisms* 18, 77–79.
- Rigby, M., W. Font and T. Deardorff (1997). Redescription of *Camallanus cotti* Fujita, 1927 (Nematoda: Camallanidae) from Hawai'i. *Journal of Parasitology* 83, 1161–1164.
- Rolbiecki, L. and J. Rokicki (2005). *Anguillicola crassus* - An alien nematode species from the swim bladders of eel *Anguilla anguilla* in the Polish Zone of the Southern Baltic and in the Waters of Northern Poland. *International Journal of Oceanography and Hydrobiology* 35, 121–136.
- Ronquist, F. and J. P. Huelsenbeck (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.

- Ronquist, F., J. P. Huelsenbeck and P. van der Mark (2005). *MrBayes 3.1 manual*.
- Rubinoff, D. and B. S. Holland (2005). Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Systematic Zoology* 54, 952–961.
- Sasal, P., H. Taraschewski, P. Valade, H. Grondin, S. Wielgoss and F. Moravec (2008). Parasite communities in eels of the Island of Reunion (Indian Ocean): a lesson in parasite introduction. *Parasitology Research* 102, 1343–1350.
- Savolainen, V., R. S. Cowan, A. P. Vogle, G. K. Roderick and R. Lane (2005). Towards writing the encyclopedia of life: an introduction to DNA barcoding. *Philosophical Transactions of the Royal Society. Series B: Biological Sciences* 360, 1805–1811.
- Schmidt, H. A., K. Strimmer and A. von Haeseler (2002). TREE-PUZZLE: maximum likelihood phlogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504.
- Schmidt, J. (1923). The breeding places of the eel. *Philosophical Transactions of the Royal Society. Series B: Biological Sciences* 211, 179–208.
- Schoth, M. and F. W. Tesch (1982). Spatial distribution of small 0-group *Anguilla* larvae in the Sargasso Sea. *Helgoländer Wissenschaftliche Meeresuntersuchungen* 35, 309–320.
- Simmons, M. P. and H. Ochoterena (2000). Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49, 369–381.
- Skryabin, K. I. (Ed.) (1954). *Key to parasitic nematodes. Volume 4: Camallanata, Rhabditata, Tylenchata, Trichocephalata, Dioctophymata and distribution of parasitic nematodes in different hosts*.
- Smales, L. R. (2000). *Durettechina beveridgei* n. g., n. sp. (Nematoda: Seuratidae) from *Antechinus* spp. (Dasyuridae: Marsupialia) from Australia. *Systematic Parasitology* 45, 25–28.
- Stamatakis, A. (2006). RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. *Bioinformatics* 22, 2688–2690.
- Stamatakis, A., P. Hoover and J. Rougemont (2008). A rapid bootstrap algorithm for the RAxML Web Servers. *Systematic Biology* 57, 758–771.
- Strimmer, K. and A. von Haeseler (1997). Likelihood-mapping: a simple method to visualise phylogenetic content of a sequence alignment. *Proceedings of the National Academy of Sciences* 94, 6815–6819.

- Subbotin, S. A., D. Sturhan, N. Vovlas, P. Castillo, J. Tanyi Tambe, M. Moens and J. G. Baldwin (2007). Application of the secondary structure model of rRNA for phylogeny: D2-D3 expansion segments of the LSU gene of plant-parasitic nematodes from the family Hoplolaimidae Filipjev, 1934. *Molecular Phylogenetics and Evolution* 43, 881–890.
- Suzuki, J., R. Murata and J. Hosaka, M. and Araki (2010). Risk factors for human *Anisakis* infection and association between the geographic origins of *Scomber japonicus* and anisakid nematodes. *International Journal of Food Microbiology* 137, 88–93.
- Tanaka, H., H. Kagawa, H. Ohta, T. Unuma and K. Nomura (2003). The first production of glass eel in captivity: fish reproductive physiology facilitates great progress in aquaculture. *Fish Physiology and Biochemistry* 28, 493–497.
- Taraschewski, H. (2006). Host and parasites as aliens. *Journal of Helminthology* 80, 99–128.
- Taraschewski, H., J. Boomker and F. Moravec (2005). Studies on the morphology and ecology of *Anguillicola papernai* (Nematoda: Anguillicolidae) parasitizing the swimbladder of the African longfin eel, *Anguilla mossambica*, and on the helminth community in this eel. *Diseases of Aquatic Organisms* 63, 185–195.
- Taraschewski, H., F. Moravec, T. Lamah and K. Anders (1987). Distribution and morphology of two helminths recently introduced into European eel populations: *Anguillicola crassus* (Nematoda, Dracunculoidea) and *Paratenuisentis ambigous* (Acanthocephala, Tenuisentidae). *Diseases of Aquatic Organisms* 3, 167–176.
- Tatsukawa, K. (2001). The present state of population dynamics studies in unagi, *Anguilla japonica*. *Aquabiology* 133, 108–114.
- Templeton, A. R., K. A. Crandall and C. F. Sing (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132, 619–633.
- Tesch, F. W. (1977). *The Eel: Biology and Management of Anguillid Eels*. Chapman & Hall.
- Tesch, F. W. (2002). *The Eel, 5th edition*. Blackwell, Oxford, UK.
- Thomas, K. (1993). *The life cycle of the eel parasite Anguillicola crassus (Nematode: Dracunculoidea)*. PhD Thesis, Katholieke Universiteit Leuven.
- Torres, P., M. I. Jercic, J. C. Weitz, E. K. Dobrew and R. A. Mercado (2007). Human Pseudoterranocosis, an Emerging Infection in Chile. *Journal of Parasitology* 93, 440–443.
- Tsukamoto, K. (2006). Oceanic biology: spawning of eel near a seamount. *Nature* 439, 929.

- Tsukamoto, K., J. Aoyama and M. J. Miller (2002). Migration, speciation, and the evolution of diadromy in anguillid eels. *Canadian Journal of Fisheries and Aquatic Sciences* 59, 1989–1998.
- Tsukamoto, K. and T. Arai (2001). Facultative catadromy of the eel *Anguilla japonica* between freshwater and seawater habitats. *Marine Ecology Progress Series* 220, 265–276.
- Tzeng, W. N. (1997). Short and long-term fluctuations in catches of elvers of the Japanese Eel, *Anguilla japonica*, in Taiwan. *Developing and sustaining world fisheries resources. Proceedings of 2nd World Fisheries Congress*. CSIRO press, 85–89.
- Valentini, A., F. Pompanon and P. Taberlet (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution* 24, 110–117.
- Van Megen, H., S. Van Den Elsen, M. Holterman, G. Karssen, P. Mooyman, T. Bongers, O. Holovachov, J. Bakker and J. Helder (2009). A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* 11, 927–950.
- Vanfleteren, J. R. and A. Vierstraete (1999). Insertional RNA editing in metazoan mitochondria: The cytochrome *b* gene in the nematode *Teratocephalus lirellus*. *RNA* 5, 622–624.
- Vercruyse, J. and E. Claerebout (2001). Treatment vs non-treatment of helminth infections in cattle: defining the threshold. *Veterinary Parasitology* 98, 195–214.
- Watson, J. D. and F. H. Crick (1953). Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171, 737–738.
- Weir, B. S. (1996). *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Associates, Inc. (Massachusetts).
- Weir, B. S. and C. C. Cockerham (1984). Estimating F-statistics for the analysis of population structure. *Evolution* 38, 1358–1370.
- WHO World Health Report (2002). Annex table 2: deaths by cause, sex, and mortality stratum in WHO regions and annex table 3: burden of disease in DALYs by cause, sex and mortality stratum in WHO regions. Geneva: World Health Organization.
- WHO World Health Report (2004). Prevention and control of schistosomiasis and soil-transmitted helminthiasis. Geneva: World Health Organization.
- Wielgoss, S., H. Taraschewski, A. Meyer and T. Wirth (2008). Population structure of the parasitic nematode *Anguillicoloides crassus*, an invader of declining North Atlantic eel stocks. *Molecular Ecology* 17, 3478–3495.

Wijová, M., F. Moravec, A. Horák and J. Lukeš (2006). Evolutionary relationships of Spirurina (Nematoda: Chromadorea: Rhabditida) with special emphasis on dracunculoid nematodes inferred from SSU rRNA gene sequences. *International Journal for Parasitology* 36, 1067–1075.

Windsor, D. A. (1998). Most of the species on Earth are parasites. *International Journal for Parasitology* 28, 241–270.

Wright, S. (1978). *Evolution and Genetics of Populations*. University of Chicago Press (Chicago), 540.

Würtz, J., H. Taraschewski and B. Pelster (1996). Changes in gas composition in the swimbladder of the European eel (*Anguilla anguilla*) infected with *Anguillicola crassus* (Nematoda). *Parasitology* 112, 233–238.

Yamaguti, S. (1933). Studies on the helminth fauna of Japan, part 9. Nematodes of fishes. *Japanese Journal of Zoology* 6.



## E. Colophon

The **original source** for this diploma thesis was created and compiled in TeXShop 2.18 (GPL public license, © 2001-2008, Richard Koch, [http.texshop.org](http://texshop.org)) on a MacBook 5.1 under Mac OS X 10.5.8 (TM and © 1983-2009 Apple Inc., USA). **Images** were processed using ImageJ 1.43 (Public domain, National Institutes of Health, USA, <http://rsb.info.nih.gov/ij>), ImageMagick 6.2.8 (© 1999 - 2006 ImageMagick Studio LLC <http://www.imagemagick.org>) and yEd (© 2000 - 2010 yWorks GmbH, <http://www.yworks.com>). **Geographical information** was obtained using GoogleEarth Version 5.1.3 (© 2010 Google Inc., USA, <http://earth.google.com/intl/index.html/>). **Literature searches** were conducted using the GoogleScholar web front-end (© 2010 Google Inc., USA, <http://scholar.google.com/>). **Thesis data** was calculated using the LaTeX word count web frontend (<http://folk.uio.no/einarro/Services/texcount.html>).

This thesis was printed in the A4 (210 × 297 mm) format, using 120 g paper, duplex printing, Neo-colour, and laminated using thermal binding, at Repro-Dannenmaier GmbH, Klosterweg 10A, 76131 Karlsruhe (Germany) (<http://www.repro-dannenmaier.com/>)

**Thesis data:** 121 pages, 20,213 words in text, 26 (+6) figures and 18 tables

---

## Acknowledgment

*Wenn wir [...] die Menschen nur nehmen, wie sie sind, so machen wir sie schlechter. Wenn wir sie behandeln, als wären sie, was sie sein sollten, so bringen wir sie dahin, wohin sie zu bringen sind.*

- Johann Wolfgang von Goethe (1749 - 1832), "Wilhelm Meisters Lehrjahre" VIII, 4

The author of this thesis gratefully thanks the following persons for the provision of unimaginable support (in order of appearance in his life):

His parents, **Mirica Hanel Laetsch** and **Wolfgang E. Laetsch**. His grandparents, **Anna Kucic Hanel** and **Franz Hanel**. His confederates **Christian D. Reyes-Knoche** and **Daniel S. Weickgenannt**, and his girlfriend, **Elena Andres Prado**.

Furthermore, the author acknowledges the generous assistance, the knowledge and the seemingly unlimited patience provided by the following persons (in alphabetical order):

**Mark Blaxter, Charles Franz, Kerstin Geiß, Emanuel Heitlinger, Martin Jones, Jenna Mann, Steven Nadler, Peter Nick, Trevor Petney, Andrew Rambaut, Graham Stone, Horst Taraschewski, Sébastien Wielgoss** and the **Gene-Pool Team**.

---

English translation: If we [...] take men only as they are, we make them worse. If we treat them as for what they should be, we bring them to where they can be brought.