Divergence of an introduced population of the Swimbladder-nematode *Anguillicola crassus* - a transcriptomic perspective

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)
Fakultät fr Chemie und Biowissenschaften

Karlsruher Institut für Technologie (KIT) - Universitätsbereich

vorgelege
Dissertation
von

Emanuel Heitlinger

geboren in
Schwäbisch Gmünd

Dekan:

Referent: Prof. Dr. Horst Taraschewski

Korreferent: Prof. Mark Blaxter

Tag der mndlichen Prfung:

Abstract

The difference of the immune attack on *A. crassus* in the two diffent hosts provides an opportunity to investigate the parasite's response to different "immune environments" on a transcriptomic basis.

To my grandmother Ruth my brother Roman and my wife Silvia

Acknowledgements

I would like to acknowledge the thousands of individuals who have coded for free software and open source projects. It is due to their efforts that code is shared, tested, challenged and improved. Sharing their intellecual property as a general good, they serve progress in science and technology.

Contents

List of Figures

LIST OF FIGURES

List of Tables

Glossary

DNA Desoxy Ribonucleic Acid; a chemical molecule bearing the heritable genetic information in all life on earth

dpi Days post infection; In infection experiments, a point in time given in days after an individual has been infected

ORF Open Reading Frame; a region in a DNA-sequence begining with a start-codon and not containing a stop-condon. For example a region within a processed mRNA transcript being transcribed into a protein

SNP Single Nucleotide Polymorphism; variation occurring in a single nucleotide between two closely related homlogous sequences. Leading to for example to allelic differences within a population or even the homologous chromosomes in an individual

1

Introduction

1.1 The study organism: *Anguillicola crassus*

1.1.1 Ecological significance

Anguillicola crassus Kuwahara, Niimi and Ithakagi 1974 (1, 2) is a swimbladder nematode naturally parasitizing the Japanese eel (*Anguilla japonica*) indigenous to East-Asia. In the last 30 years anthropogenic expansions of its geographic- and host-range to new continents and host-species attracted interest of limnologists and ecologists. The newly accquired hosts are, like the native host, freshwater eels of the genus *Anguilla* (and the use of the final host seems to be limited to this genus (3)), but the nematode displayed a high versatility and plasticity in most other aspects of it's life to successfully invade new continents (4).

First *A. crassus* colonized Europe in the eraly 1980ies and spread through almost all populations of the European eel (*Anguilla anguilla*) during the following decades (reviewed in (5)). This spread includes populations of the European eel in North Africa(6, 7). At the present day *A. crassus* is found in all but the northernmost population of the European eel in Iceland (8). It has to be noted however, that low water temperature (9) and salinity (10) limit the dispersal of *A. crassus* larvae and high epidemiological prameters are rather expected in freshwater and in southern latitudes.

Wielgoss et al. (11) studied the population structure of *A. crassus* using microsattelite markers and inferred details about the colonization process and history. These details are in very good agreement with previous knowledge about the introduction

history, and lead to the situation that the process of introduction and spread can be considered very well described:

A. crssus was first recorded in 1882 in North-West Germany, and this record was published in a German fishery magazine in 1985 (12). The import of Japanese Eels from Taiwan to the habor of Bremerhaven in 1980, was soon identified as most likely introduction event (13). Taiwan as the most likely geographical source of the introduction was in turn also inferred from population structure by Wielgoss et al. Furthermore, from the fact that genetic diversity is highest in northern regions of Germany and gradually declines to the south, they concluded a single introduction event to Germany as source for all populations of *A. crassus* in the comprehensive set of investigated populations of the European eel. This signal was persistent together with a strong signal for anthropogenic mixing of eel and parasite populations due to restocking (14).

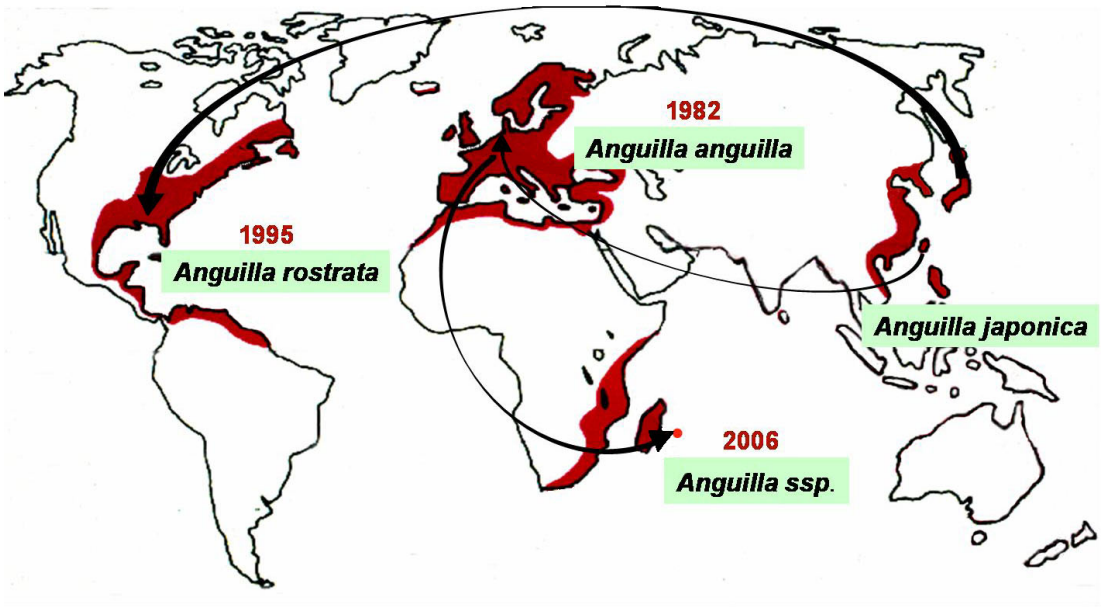

Figure 1.1: Transcontinental dispersal of *A. crassus:* - Invasions of different continents by different source-populationa are illsutrated using arrows. Red color indicates the range of the eel species targeted by the invasion. Modified form (15), based on data reviewed in (5) and newer findings in (11) and (16)

A little less is known about a second colonization of *A. crassus*, which succeeded in North-America. Since the 1990s populations of the American eel (*Anguilla rostrata*) have been invaded as novel hosts (17, 18, 19). Wielgoss et al. identified Japan as the

most likely source of this American population of *A. crasssus*.

Finally *A. crassus* has been detected in three indigenous species of freshwater eels on the island of Reunion near Madagascar (16).

Copepods and ostracods serve as intermediate hosts of *A. crassus* in Asia, as well as in the introduced ranges (20). In these L2 larvae develop to L3 larvae, infective to the final host. Once ingested by an eel they migrate through the intestinal wall and via the body cavity into the swimbladder wall (21), i.a. using a trypsin-like proteinase(22). In the swimbladder wall L3 larvae hatch to L4 larvae. After a final moult from L4 to adults (via a short preadult stage) the parasites inhabit the lumen of the swimbladder, where they eventually mate. Eggs containing L2 larvae are released via the ductus pneumaticus into the eels gut and finally into the water(23).

One of the possible differences between Asian and European population of *A. crassus* is the widespread use of paratenic hosts in European waters (24, 25). Such a use of paratenic hosts has not been from the Asian range of the parasite yet and there are some speculation that the use and availability of paratenic hosts could be a factor explaining the success of invasion or even the higher epidemiological parameters in Europe (25). However the lack of evidence for the use of paratenic host in Asia could as well be a a result of the lack of appropriate studies in Asian water systems.

The impact of *A. crassus* on the European eel has been a major focus of research during the past decades. Pathogenic effects on the eels such as a thinkening (26) and infammation (27) of the swimbladder wall, can lead to mortality of eels, when combined with co-stressors (28). Especially the changes in the tissue of the swimbladder wall have been shown to influence swimming behavior and it has been speculated that eel may fail to complete theier spawning migration (29). Anguillicolosis (the condition caused by *A. crassus*) has therefore been speculated to be a cofactor in the decline of European eel stocks (30) caused by overfishing of glass-eels (31).

High prevalences of the parasite of above 70% (e.g. (32)), as well as high intesities of infections were reported, throughout the newly colonized area (33).

These differences in abundance of *A. crassus* in East Asia compared to Europe are commonly attributed to the different host-parasite relations in the final eel host permitting a differential survival of the larval and the adult parasites (34). Recently, data from experimental infections of European eels with *A. crassus* have been published

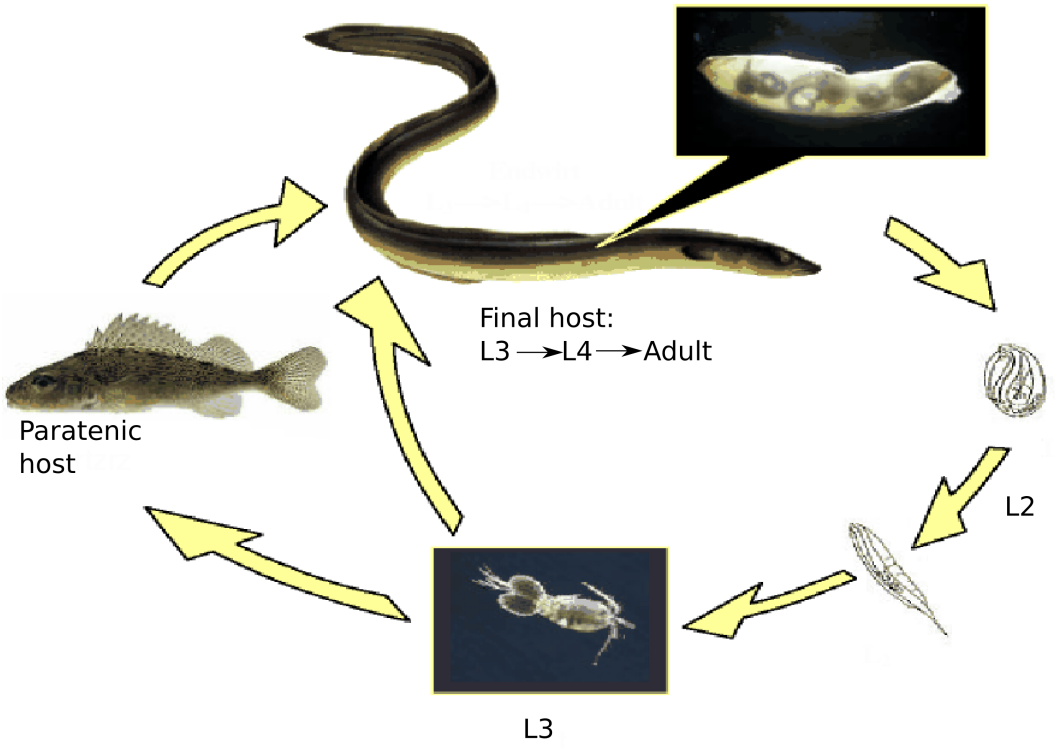

Figure 1.2: Life-cycle of *A. crassus* - Adult females deposit already hatched L2 in the lumen of the swimbladder. Larvae migrate through the *ductus pneumaticus* and the intestine into the open water.Copepodes serve as intermediate host where infectiv L3-larvea develop. These can be transported and accumulated in paratenic hosts or directly ingested by an eel. They migrate through the eel's intestinal wall into the swimbladder wall. After the final molt to adults worms arrive in the lumen of the swimbladder, feed on blood and reproduce. Modified from (15).

(35). They show that the parasite undergoes (under experimental conditions) a density-dependent regulation keeping the number of worms within a certain range. As in the natural host in Asia prevalences and intesities are lower (36), high epidemiological parameters were attributed to the inadequate immune-response of the European Eel (37). Interestingly the differences in the two host also affect the size and life-history of the worm: In European eels the nematodes are bigger and develop and reproduce faster (34). While the Japanese eel is capable of killing larvae of the parasite after vaccination (38) or under high infection pressure (39).

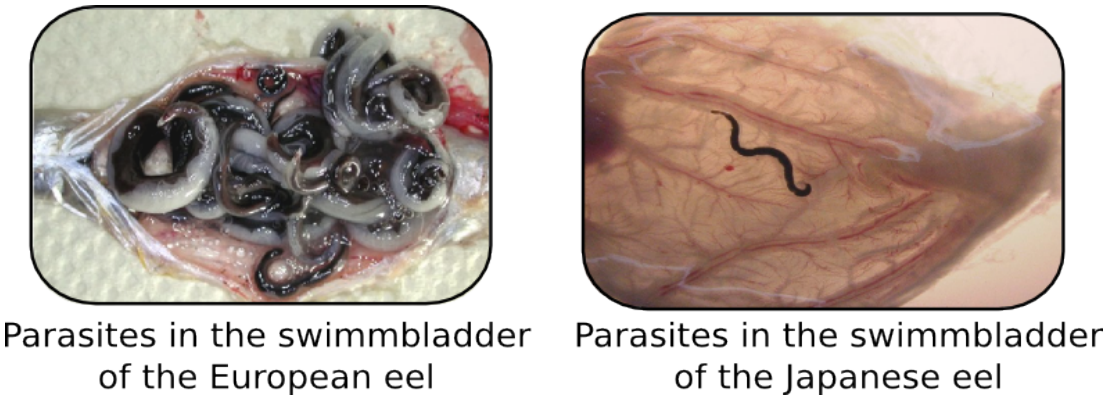

Figure 1.3: Difference between worms in the swimbladder of the European eel and the Japanese eel - Note the bigger size and higher number of worm in a typically infected European eel. In comparison in the Japanese eel worms are smaller and intensities of infection are much lower. The dark brown matter is ingested eel-blood visible through the transparent nematode body- and intestinal wall, the white matter are developing eggs and larvae in ovaries of female *A. crassus*.

1.1.2 Evolutionary significance

A decline of *A. crassus* populations has been hypothised looking at populations over the last two decades (33). However it should be noted, that such a decline has not been confirmed in a explicit meta-analysis and even if present could be explained rather by a lower population density of the host, than by an evolution of *A. crassus* towards lower infectivity.

With a view on the potential co-evolution (i.e. adaptation), of the eel-host to *A. crassus*, that European and American eels are considered panmictic (40): Signals for population structure, interpreted as evidence against panmixia first (41), have been

shown to be an artifact of temporal variation between cohorts of juvinile eels (42, 43, 44). Such panmixia would severly reduce the effectiveness of selection, when unifected populations are participating in reproduction. While

1.1.2.1 Divergence of *A. crassus* populations

Common-garden experiments (also termed "transplant expreiments") are a method to identify genetic components of phenotypic differences between potentally diverged population of a species, used for almost as long as scientists investigate evolution (45, 46). In the reciprocal version of these experiments, representatives of each population intented to be studied are raised in the other population's natural environment. A modification of this would be to raise each population under experimental conditions resembling the environment of the other populaton.

When applied to parasites infecting two different hosts such an experiment can be best described as "cross-inoculation experiment under common garden conditions" (47). In a recent study using this method both European and Japanese eels were infected under laboratory conditions with worms from three geographic origins; Southern Germany, Poland and Taiwan.

In these experimets differences between the two European populations and and the Taiwanese population of worms manifested. Differences were especially (but not solely) visible in the early stages of the life-cycle.

In the European eel:

- The number of **L3 larvae** for Taiwanese population was higher than for European populations.
- In the Taiwanese population there were less **L4 larvae** at 25 dpi and the levels of this larval stage were stable during the infection, the numbers of L4 for the European populations decreased with the time.
- Up to 50 dpi there were less living **adults** observerd for Taiwanese worm population.
- Fewer **dead adult** worms were recorded for the Taiwanese population beginning from 50 dpi.

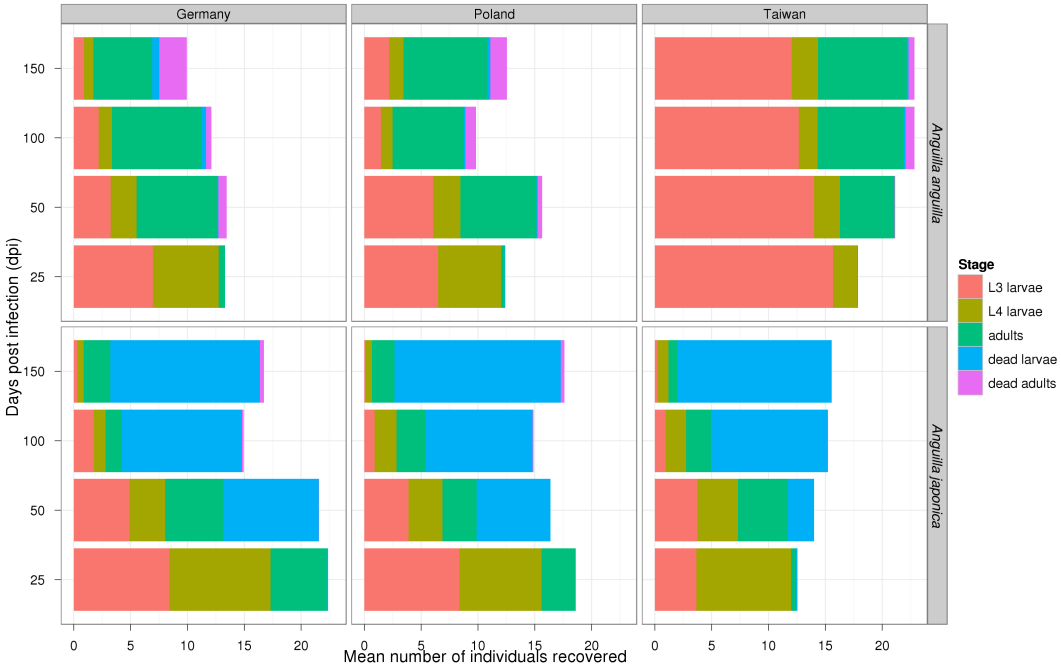

Figure 1.4: Differences in developmental speed - data courtesy of Urszula Weclawski

In the Japanese eel ther were for the Taiwanese population compared to the European population of worms:

- Fewer **L3 larvae** at 25dpi
- More **L4 larvae** at this point in time
- Fewer living **adults** at 25 and 150 dpi
- Fewer **dead adults** beginning from 50 dpi

These findings taken can be consolidated to the interpretation, that an increase in the speed of development can be observed in the European population. Although such experiments have their problems because environmental factors can interact with

1.1.2.2 Interest in *A. crassus* based on its phylogeny

The genus *Anguillicola* comprises five morphospecies (48): In East Asia in additon to *A. crassus*, *A. globiceps* Yamaguti, 1935 (49) parasitises *Anguilla japonica*. *A. novaezelandiae* is endemic to New Zealand and South-Eastern Australia in *Anguilla australis*

and *A. australiensis* Johnston et Mawson, 1940 (50) parasitizes the long-fin eel *Anguilla reinhardtii* in North-Eastern Australia. Finally *A. papernai* is known from the African longfin eel *Anguilla mossambica* in Southern Afrika and Madagascar.

In 2006 Moravec promoted the the former subgenus *Anguillicoloides* comprising all species but *A. globiceps* to the rank of a genus (2). This subdivision of the Anguillicolidae in two genera was revised based on the notion that monophyly of *Anguillicoloides* had to be rejected, *Anguillicolides crassus* was restored to *Anguillicola crassus*. In the same study on the phylogeny of the *Anguillicolidae A. crassus* was identified as the basal species in the genus, analysing nuclear genes SSU and LSU (see figure 1.5) or as forming a clade with the oceanic species with *A. globiceps* and *A. papernai* as a sister clade (see figure 1.6).

Neiter of these phylogenetic hypotheses is consistent with the phylogeny of the eel-hosts without host-switching: Assuming the establishment of *Anguillicola* in an ancestral Indo- pacific host at least three host-switch events are needed, even to explain classical (non-recent, non-anthropogenic) host-parasite associations, some of these host-switch events must have spanned the major splits in the eel phylogeny (51): Between the clade comprising *An. reinhardtii* and *An. japonica* to the clade in which *An. australis* is found and to the most basal species of frehwater eels *An. mossambica*.

Finally the recent anthropogenic host-switch of *A. crassus* from *An. japonica* to *An. anguilla* and *An. rostrata* constitues such an acquisition of phylogenetically well separated hosts and may be a evolutionary relict of this affinity for switching.

The to date most likely phylogenetic hypothesis places the genus *Anguillicola* (the only genus in the family Anguillicolidae) at a basal position in the Spirurina (clade III *sensu* (52)), one of 5 major clades of nematodes (53, 54). The Spirurina exclusively exhibit a animal-parasitic lifestyle and comprise improtant human pathogens as well as prominent parasites of livestock (e.g. the Filaroidea and Ascarididae). A finer subdivision of the Spirurina into Spirurina A, and the Sister clades Spriurnina B and C can be seen in figure 1.7.

Within the Spirurina B an enormous phylogentic diversity of the definitive hosts can be observed ranging from fresh-water fish as hosts for the Anguillicolidae to cartilaginous fish for Echinocephalus, mammals parasitized by Gnathostoma and Linstowinema to reptiles as hosts for Tanqua.

A complex life-cycle ...

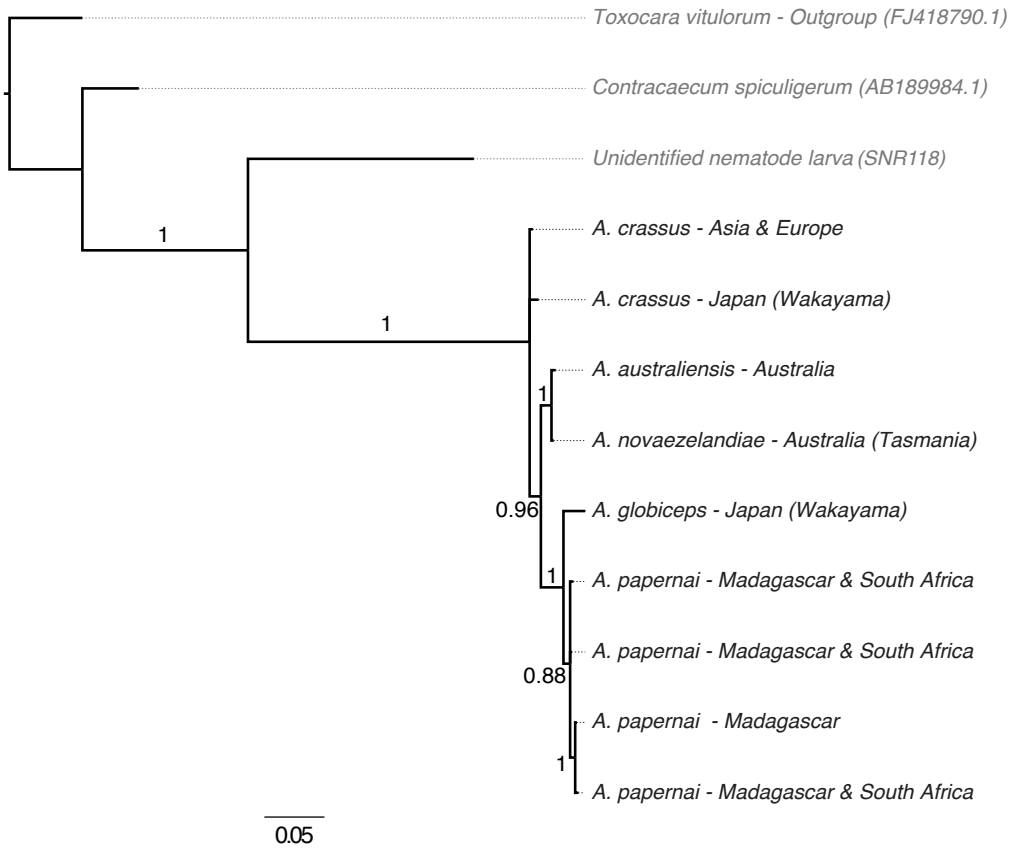

Figure 1.5: Phylogeny of the genus *Anguillicola* based on nLSU - Phylogram infered from large ribosomal subinit of *Anguillicola* and outgroups using Bayesian Inference. Lables on internal branches indicate Bayesian posterior probabilities. From

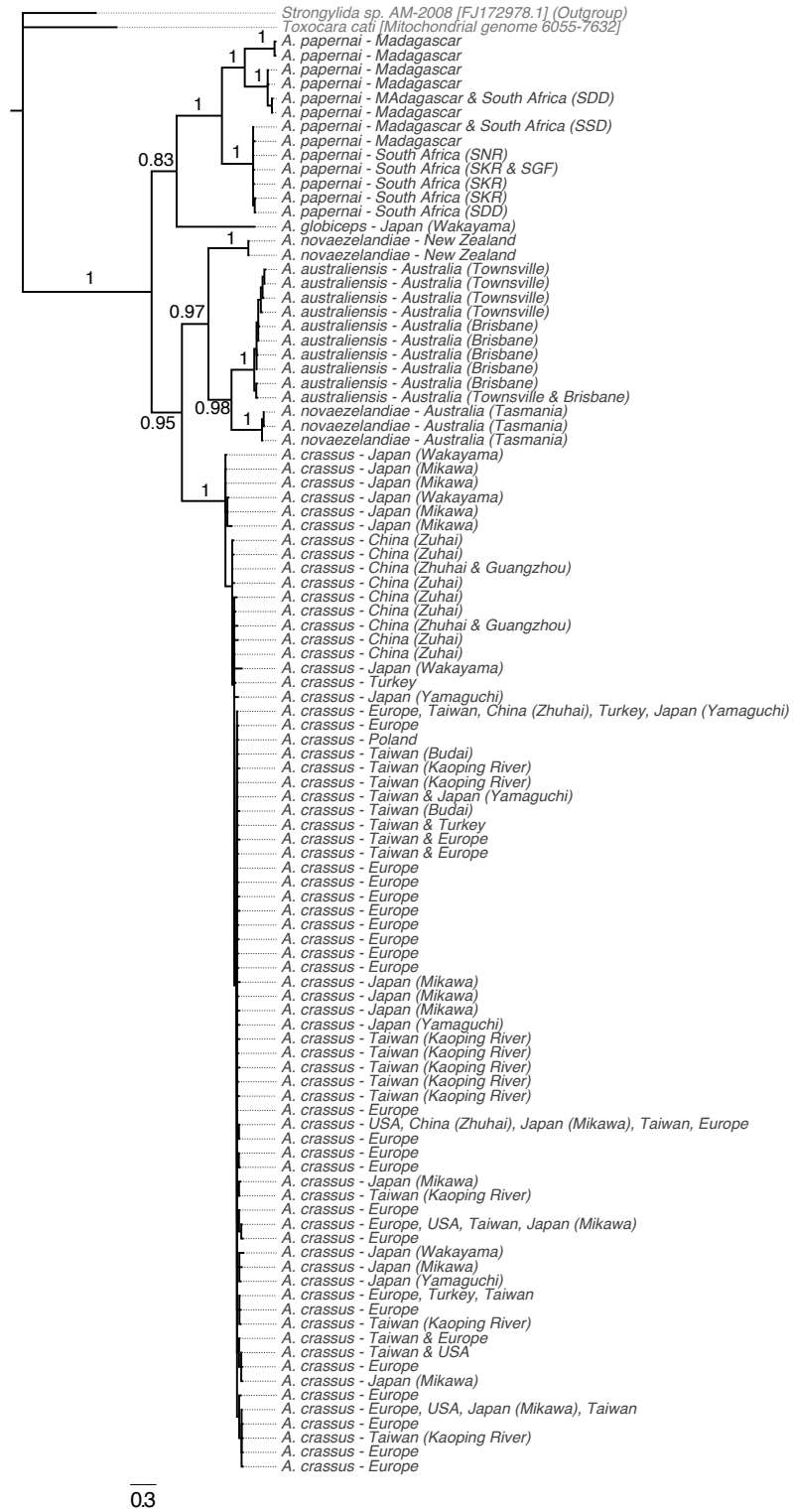

Figure 1.6: Phylogeny of the genus *Anguillicola* based on COXI - Phylogram infered for *Anguillicola* based on mitochondrial Cytochrome C oxidase subunit I and outgroups using Bayesian Inference. Lables on internal branches indicate Bayesian posterior probabilities. From

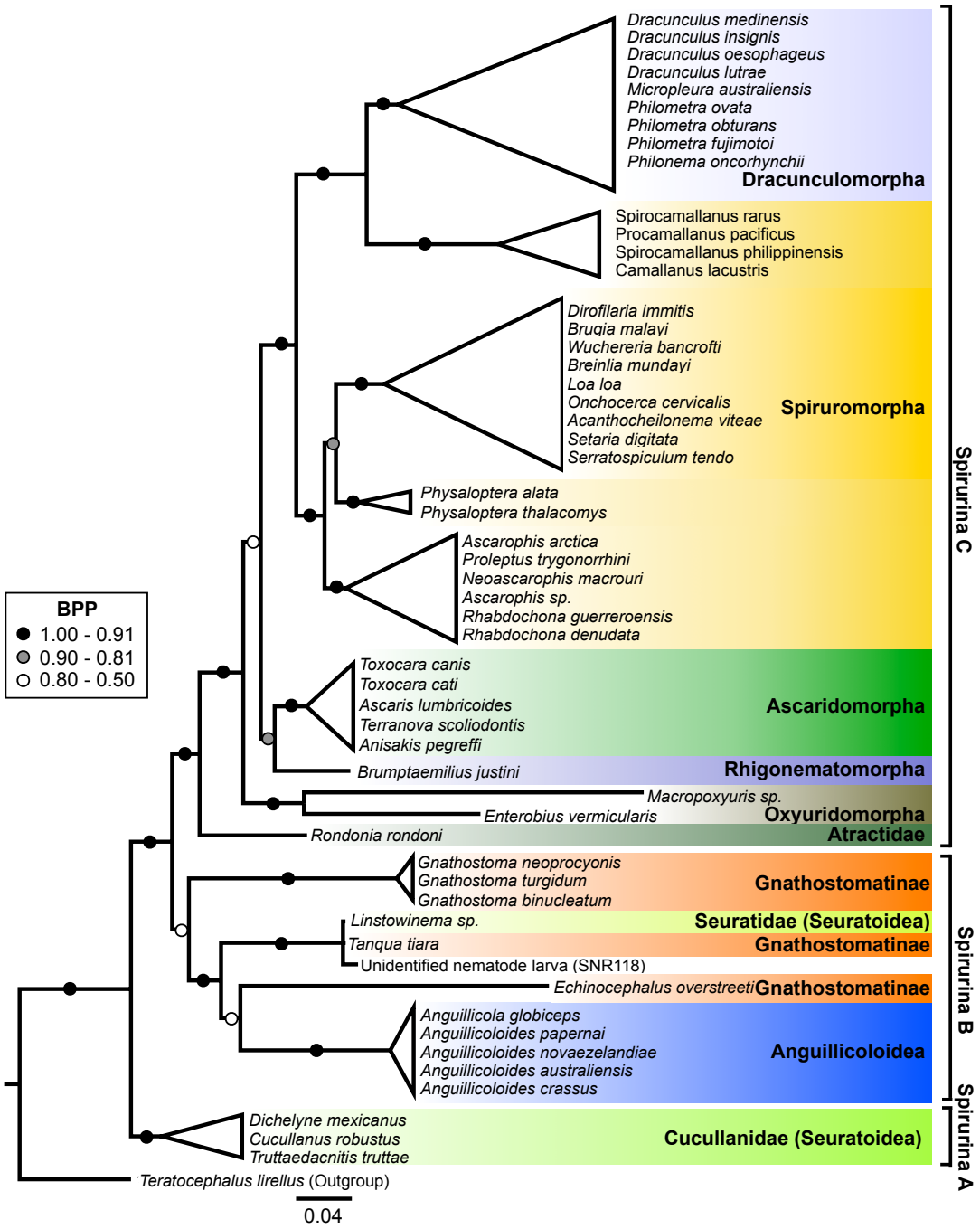

Figure 1.7: Phylogeny of nematode clade III based on nSSU - Phylogram inferred from nuclear small ribosomal subunit for Spirurina using Bayesian Inference. Branches are collapsed to highlight major groups. Lables on internal branches indicate Bayesian posterior probabilities. From

This phylogenetic position makes the Anguillicoloidae an interesting system in the endeavour to understand the emergence of parasitism in Spirurina and as an "outgroup" for functional studies of parasitism in this clade.

Some functionally interesting genes in this respect are thought to be under diversifying selection in an arms-race between host and parasite (55).

That positive or diversifying selection on parasite proteins from the host-parasite interface can lead to a overabundance of non-synonymous changes (altering the protein sequence) over synonymous polymorphisms e.g. in *Plasmodium* (56).

Memory componetnt of the vertebrate immune system has been thought to be a driving positive selection on antigenes of microorganisms (57). The immune systems of teleost has a immune system with interstin implications for the eels response to parastites.

1.2 Functional insights from other nematodes used to formulate hypotheses for *A.crassus*

In 1998 *Caenorhabditis elegans* became the first multicellular organism with a sequenced genome (58). Soon it was noted, that in addition to it's use as a general model system for the metazoa, knowledge gained in this species has the potential to be even more valuable in the phylum Nematoda (59). The breadth of genomic information available for *C. elegens* to date is illustrated by a recent publication of the Gerstein et al. (60): detailed annotation of the diverse functional genomic elements and their interactions by the modENCODE consortium.

The complete genome sequence of the nematode *Caenorhabditis elegans* (58) and *Caenorhabditis briggsae* (61), as well as the draft genome of *Brugia malayi* (62) provide useful sources for mining databases for homologous sequences.

Emerging genomes form *Trichinella spiralis* (63), *Meloidogyne incognita* (64), *Meloidogyne hapla* (65) *Pristionchus pacificus* the pinewood nematode *Bursaphelenchus xylophilus* (66)

The lack of genomic information in many species of nematodes promoted use of ESTs as a tool for gene-discovery and partial genomes *sensu* (67) were successfully interrogated for a large array of genes interesting for different scientific communities.

In nematode parasites of vertebrates, pathogenic factors were described as potential vaccine candidates (68).

Cystein-proteinase inhibitors (cystatins), Serin protenase inhibitors (serpins) were identifided

Homologues of mammalian cytokines were identified, which are believed to interact with receptors of mammalian

The abundant larval transcripts of *B. malay* (Bm-ALT) have been identified in the transcriptome-projects first CITE! as a gene family (69)

Bm-VAL-1

(70, 71, 72, 73))

In some of these studies secreted proteins were in the center of interest. They could potentially be excreted by the nematode to interact with the host's immune system. The detection of signal-peptides for secretion using *in silico* analysis of ESTs has been used to highlight candidate genes for example in *Nippostrongylus brasiliensis* (74), and across all nematode ESTs (75).

Proteomic analysis in *Brugia malayi* (76, 77), *Heligmosomoides polygyrus* (78) and *Haemonchus contortus* (79) was able to find evidence for excretion for some of the protein-products and to highlight additional candidate genes.

1.3 Advances in sequencing technology enabeling this study

For almost tree decades the method developed by Frederick Sanger (80) was the only practical choice for determining the sequence of nucleic acid. Although modern machines use the chain-termitaion method combined with capillary gel electorphoresis (81) in a highly paralized way, costs and labour constrained sequencing to a well established laboratory-model organisms (the bacterium *Escherichia coli*, 1997 (82); the baker's yeast *Saccharomyces cerevisiae*, 1996 (83); the nematode *Caenorhabditis elegans* 1998 (58), the fruit fly *Drosophila melanogaster*, 2000 (84); the mouse *Mus musculus*, 2002 (85); to name a few together with the year of publication their genome sequence).

This "first generation of genomics" culminated in the publication of the human genome in 2001 (86).

In this context a common characteristic of all DNA-sequencing methods has to be emphasized: Read-length is usually shorter than the length of the target molecule to

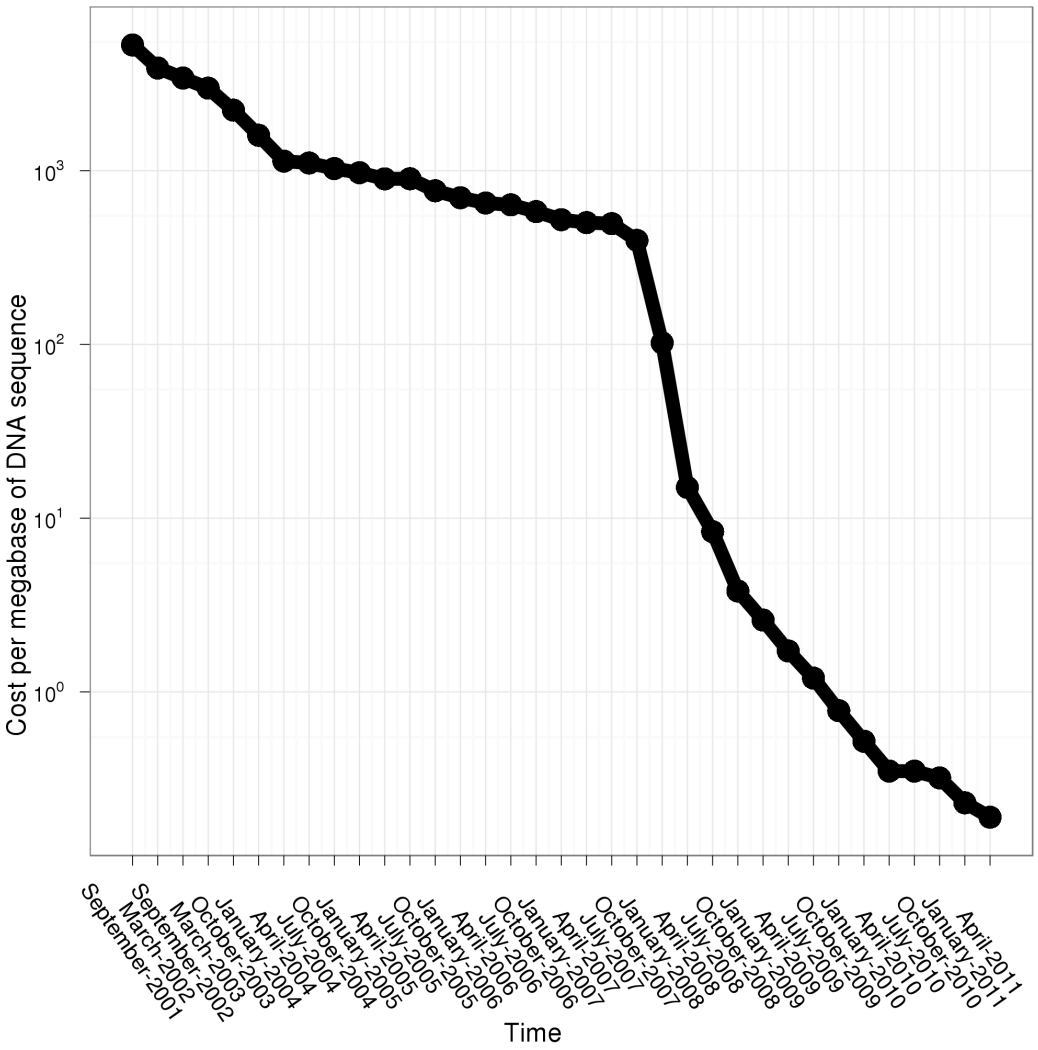

Figure 1.8: Falling seqeuncing costs - Sequencing costs falling due to advances in Solexa-sequencing: Due to improved read-length and data-volume on this plattform per base sequencing-prices for many applications thumble into free fall. Data provided by National Human Genome Research Institute, NHGRI.

be sequenced. This potential problem is solved by oversampling the target molecule, producing overlapping sequence. The amount of redundancy of the overlap is termed coverage (e.g. 10-fold coverage means a base is sequenced 10 times redundantly) the method as such is referred to as shotgun-sequencing and has - shortly after sequencing chemestry - been described by Sanger (87). Soon copmuter programs were necessary to align sequences, to compute overlaps and consensus sequences (88) and the process of computationally reconstructing the target molecule was termed sequence-assembly (89).

Advances in sequencing technology (often termed "Next Generation Sequencing"; NGS), provide the opprotunity for rapid and cost-effective generation of genome-scale data. The technologies described as NGS have in common that they use radically new chemistry comapared to the Sanger-method, up to date all these methods produce much more, but shorter reads than classical sanger sequencing. This fostered use and development of new methods to assemble large-scale shotgun sequences, as higher coverage but shorter read-length (and also lower accuracy) are increasing the computational complexity of the assembly-problem (reviewed in (90)).

NGS technologies are are increasinly used in studies on organisms with ecological and evolutionary significance. Such ecological and evolutionary "model organisms" often lack reference genomes to guide the assembly-process.

"Genome-scale" sequencing in the broadest context can also mean sequenching comprehensive transcriptome datasets: Such transcriptomic datasetes are still less expensive than genomic data-sets in terms sequencing costs and analytical needs.

1.3.1 Pyro-sequencing

see also

The longer read length of 454-sequencing (91) compared to other NGS technologies, allows *de novo* assembly of Expressed Sequence Tags (ESTs) in organisms lacking previouse genomic or transcriptomic data (for a comprehensive list of studies using this approach before Oct 2010 see (92)).

A study on trout in Lake Superior (93) used an approach similar to the appoach in the work presented here: Fish show two different phenotypes were raised in a common environment, demonstrating the genetic fixation of the phenotypic trait. 454 sequencing was then used to measure the gene expression levels and to indentify 40 genes from two

pathways being differently expressed and therefore showing divergent evolution of gene-expression.

1.3.2 Illumina-Solexa sequencing

As shorter read-length but higher throughput of the Illumina-Solexa platform provides superior means for gene expression analyis (94):

RNA-seq (95)

Expression-tags (SuperSAGE (96)) provide the benefit of classical SAGE-analysis (97) with those of the ulta hight throughput of Illumina-Solexa sequencing.

1.4 Gene-expression and evolutionary divergence

Today, both theoretical arguments as well as field and laboratory data suggest that evolution, including divergence of populations, can occur very rapidly given the right selective pressure. Such situations provide us with the opportunity of examining how divergence and even speciation work at the molecular genetic level (98) .

Divergence in gene-expression is thought to be a factor for the establishment of reproductive barriers through hybrid

In *Drosophila* the effect of cis- and trans-regulatory differences (99)

In *Drosophila* hybrid sterility in hybrids between species (100)

2

Aims of the project

2.1 Final aim

The source of my interest in *A. crassus* and the ultimate goal of the study presented here is the identification of .

And followed my interest in

2.2 Preliminary aims

Establishment of transcriptomic and genomic resources,

3

Pilot sequencing (Sanger method)

In preparation of high-throughput transcriptome sequencing of the swimbladder nematode *A. crassus* expressed sequence tags (ESTs) were generated using traditional Sanger-technology. In total 945 reads from adult *A. crassus* (5 libraries from 4 cDNA preparations, including 541 sequences generated by students in a laboratory course) and 288 reads from liver-tissue of the host species *Anguilla japonica* (3 libraries from 3 cDNA preparations) were sequenced.

Initial quality screening

The initial quality screening of *A. crassus*-sequences revealed a high number of sequences that had to be discarded due to failed sequencing reactions (sequences beeing too short after quality trimming by `trace2seq`) in the library prepared by students. For sequences of *Anguilla japonica* and the other libraries from *A. crassus* failed sequencing reactions were less common.

In the next screening-step for *A. crassus* 125 (13.23%) and for *Anguilla japonica* 64 (22.22%) of the sequences were excluded because of homopolymer-runs considered artificial. This resulted in 452 of the nematode and 195 of the host reads regarded of sufficient quality for further processing after base-calling and quality screening.

rRNA screening

The further screening of sequences revealed a high abundance of rRNA (see Figure 3.1) ranging from 71.67% to 91.67% of obtained sequences. High abundances of rRNA were also found in the libraries from host liver tissue (see table 3.1), ranging from 71.67% to 77.42% . This contamination in libraries from both species was mainly responsible for a low ammount of sequences beeing of sufficient quality for submission to NCBI-dbEST. At this point for the *A. japonica*-dataset 36 sequences were submitted to NCBI-dbEST under the Library Name "*Anguilla japonica* liver" and were assigned the accession LIBEST_027503 by the curators.

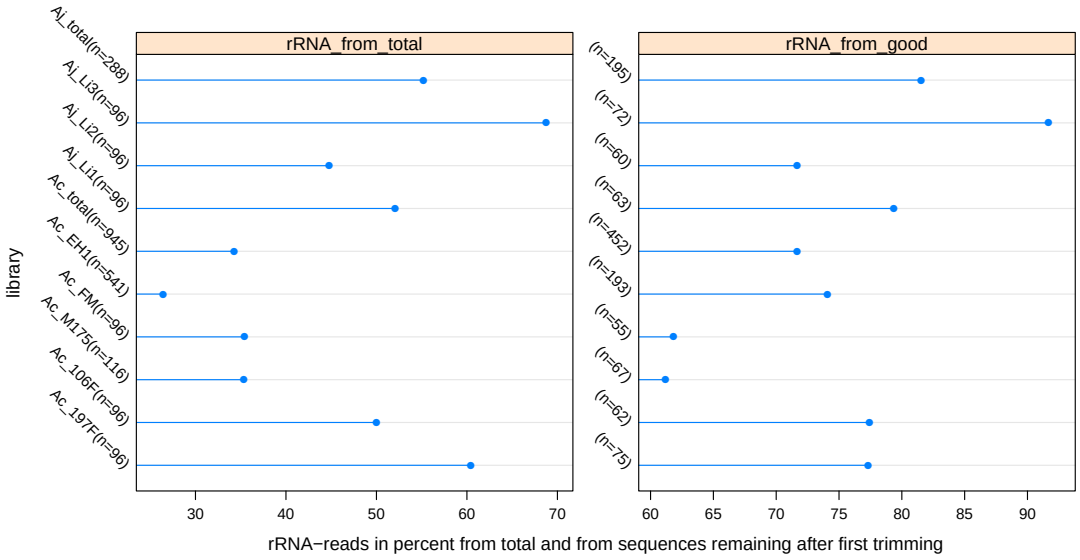

Figure 3.1: Proportion of rRNA in different libraries for *A. crassus* and *A. japonica* - rRNA abundance as proportion of the raw sequencing-reads (rRNA from total) and as proportion of the reads after quality screening (rRNA from good). Libraries starting with "Ac_" are from *A.crassus*, libraries starting with "Aj_" are from *A. japonica*.

Screening for host-contamination

For the *A. crassus*-dataset screening for host-sequences at this stage was regarded necessary based on the notion that a big proportion of the tissue prepared in RNA extraction consisted of eel-blood inside the gut of the worms (see also Figure 1.3). Additionally a bimodal distribution

	short	poly	rRNA	fishpep	good
Ac_197F(n=96)	4	17	58	1	16
Ac_106F(n=96)	25	9	48	0	14
Ac_M175(n=116)	30	19	41	3	23
Ac_FM(n=96)	12	29	34	1	20
Ac_EH1(n=541)	297	51	143	8	42
Ac_total(n=945)	368	125	324	13	115
Aj_Li1(n=96)	10	23	50		13
Aj_Li2(n=96)	10	26	43		17
Aj_Li3(n=96)	9	15	66		6
Aj_total(n=288)	29	64	159		36

Table 3.1: Screening statistics for pilot sequencing - Number of ESTs discarded at each screening-step for single libraries and totals for species. Short, sequence to short in `trace2seq`; poly, sequences with artificial homopolymer-runs from poly-A tails; rRNA, with hits to rRNA databases; fishpep with better hits to host-protein-databases than to nematode protein databases; good, sequences regarded "valid" after all screening steps. Note that the 13 sequences in the *A. crassus*-dataset, for which fish-origin was inferred, were still submintted to NCBI-dbEST.

of GC-content in the *A. crassus*-dataset was observed with one of the modes consistent with the mean GC-content of the ESTs from the Japanese eel.

Comparison of `Blast`- results for these sequences versus nempep4 and a fishprotein-database (derived from NCBI non-redundant), showed that 13 sequences were more likely to originate from host contamination than from *A. crassus*. These 13 sequences in the *A. crassus* data-set were submitted to NCBI-dbEST with a comment, that host origin had been inferred. This reduced the dataset esentially to 115 ESTs. However it has to be noted that these 13 ESTs are still accessible through the same library name "Adult *Anguillicola crassus*" and library-identifier LIBEST_027505 and are taxonomically associated to *A. crassus* on NCBI-dbEST.

After screening of host-sequences the GC-content of *A. crassus* ESTs had a unimodal distribution (see Figure 3.2). *A. crassus* had a lower mean GC-content (37.32 ± 8.36 mean $\pm$ sd) than *Anguilla japonica* (45.79 ± 8.36 mean $\pm$ sd; two-sided t-test $p < 0.001$). The distribution of the GC-contents for sequences, for which host-origin was inferred was in agreement with the GC-distribution for host sequences.

`Blast`-annotations obtained (by similarity searches against NCBI-nr, bit-score threshold of 55) for the sequences of putative host origin were also largely in agreement with the expectations for eel-blood: One sequence could be identified being higly similar to "Hemoglobin anodic subunit" from the European eel. Others were annotated with best hits to highly exressed

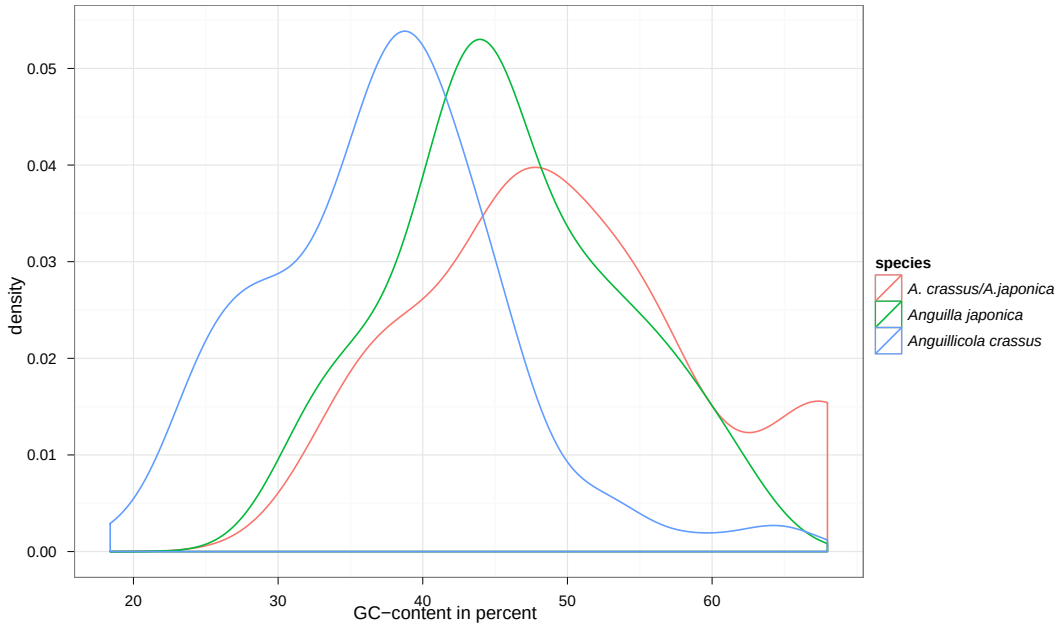

Figure 3.2: GC-content of sequences from *A. japonica* and *A. crassus* - The Japanese eel has a slighly higher GC-content than the parasite: This sequence characteristic is useful for seperation of sequences from the host-parasite interface, note the higer GC-content of the sequences from *A.crassus*, for which host origin was inferred from similarity searches (red line labeled *A.crassus/A.japonica*).

houskeeping genes from fish or vertebrates (see table 3.2). Two sequences in the set had lower similarities only to proteins predicted from genome-sequences of Chordates, and one sequence of the 13 lacked any similarity to NCBI-nr above the treshold of 55 bits.

115 of the sumbitted sequences for "Adult *Anguillicola crassus*" (LIBEST_027505) were regarded "valid" i.e. not clearly host origin.

However it should be noted, that two ESTs (Ac_EH1f_01D10 and Ac_EH1r_01D10; forward and reverse read of the same clone) were annotated with "ref|ZP_05032178.1|; Exopolysaccharide synthesis, ExoD superfamily" from *Brevundimonas* sp. BAL3. The family Caulobacteraceae, comprises bacteria living in oligothroph freshwater and sequences are probably derived from a commensal, symbiont or pathogen of eels or swimbladder-nematodes. These off-target data was left in the submission file.

For 66 (58.4%) of the remaining 113 ESTs annotations were obtained from orthologous sequences. All these orthologous sequences were from other species in the phylum nematoda.

sequence	hit identifier	hit description	species	bit-score	e-value
Ac_EH1f_005B07	gb|AAQ97992.1|	cyclin G1	*Danio rerio*	67.0	9e-10
Ac_EH1f_01A02	gb|ACO10003.1|	Nicotinamide riboside kinase 2	*Osmerus mordax*	333	1e-89
Ac_EH1f_01C10	gb|ADF80517.1|	ferritin M subunit	*Sciaenops ocellatus*	328	5e-88
Ac_EH1r_004A04	ref|XP_003340320.1|	cytoplasmic 1-like actin	*Monodelphis domestica*	102	3e-20
Ac_EH1r_005B07	gb|ABN80454.1|	cyclin G1	*Poecilia reticulata*	90.5	8e-17
Ac_EH1r_009C03	ref|NP_001122208.1|	THAP domain containing protein 4	*Danio rerio*	176	1e-42
Ac_EH1r_01A07	sp|P80946.1|	Hemoglobin anodic subunit beta	*Anguilla anguilla*	283	1e-74
Ac_FMf_08F03	ref|XP_003226802.1|	cohesin subunit SA-2-like isoform 2	*Anolis carolinensis*	219	8e-56
Ac_M175_01H02	emb|CAQ87569.1|	NKEF-B protein	*Plecoglossus altivelis*	365	3e-99
Ac_197Ff_01E04	ref|XP_002121150.1|	CUB and sushi domain-containing protein 3	*Ciona intestinalis*	80.5	2e-13
Ac_EH1f_01D07	ref|XP_002606965.1|	hypothetical protein	*Branchiostoma floridae*	82.8	3e-14
Ac_M175_01B06	ref|XP_422710.2|	hypothetical protein	*Gallus gallus*	123	1e-26

Table 3.2: Annotaion of putative host-derived sequences in the *A. crassus*-dataset - Sequences excluded because of inferred host-origin comparing similarity to nematode- and fish-proteins. The annotation obtained against NCBI-nr are in agreement with this inference of host origin, as only best hits to vertebrate proteins are found.

4

Pyrosequencing of the *A. crassus* transcriptome

454

5

NlaIII-tag sequencing (Super-SAGE)

5.1 Comparison with pyrosequencing-data

6

Transcriptomic divergence inferred from expression differences in common garden experiments

6. TRANSCRIPTOMIC DIVERGENCE INFERRED FROM EXPRESSION DIFFERENCES IN COMMON GARDEN EXPERIMENTS

7

Discussion

7.1 Sanger-method pilot-sequencing

One reason to sequence the libraries from the eels host was to elucidate whether this contamination was nematode or species-typical (e.g caused by poly-dT primers binding to A-rich rRNA regions), or caused by shortcomings in the preparation.

Nevertheless the stringent quality trimming and processing of raw reads, as summarized in the present document, make the remaining ESTs a valuable resource for comparison with future 454-sequencing-data..

7.2 454-pyrosequencing

8

Materials & methods

8.1 Sampling of worms from wild eels

8.1.1 Sampling in Taiwan

Cultured eels were acquired from an aquaculture directly adjacent to Kaoping river (22.6418N; 120.4440E) 15km stream upwards from it's estuary, on the 29th of April 2008. On the same day wild eels were picked up at Tunkang Biotechnology Research Centre Fisheries Research institute in Tunkang, Pintung, Taiwan, where they had been sheltered since the time of purchase during the 2nd two weeks of April 2008 from a fisherman, fishing in the estuary of Kao-Ping river (22.5074N; 120.4220E). All eels were transported to the Institute of Fisheries Science at the National Taiwan University in Taipei in aerated plastic bags, where they were sheltered until dissection.

Dissection of eels was carried out during May 2008. Eels were decapitated, length (to the nearest 1.0mm) and weight (to the nearest 0.1g) were measured, and sex was determined by visual inspection of the gonads. The swimbladder was opened, adult worms were removed from the lumen with a forceps, their sex was determined, and they were counted. All adult *A. crassus* were preserved in RNAlater(Quiagen, Hilden, Germany) in individual plstic tubes.

8.1.2 Sampling of European worms

Worms from the European eel were sampled in Sniardwy Lake, Poland (53.751959N ,21.730957E) by Urszula Weclawski and from the Linkenheimer Altrhein, Germany (49.0262N; 8.310556E), following a procedure similar to the one described above for worms from Taiwan.

8.2 RNA-extraction and cDNA synthesis for Sanger- and 454-sequencing

Total RNA was extracted from single, whole worms using the RNeasy kit (Quiagen, Hilden, Germany), following the manufacturers protocol. Alternatively parts of the liver of the host species *Anguilla japonica*, which also had been preserved in RNAlater were used for RNA extraction, following the same protocol.

The Evrogen MINT cDNA synthesis kit (Evrogen, Moscow, Russia) was then used to amplify mRNA transcripts according to the manufacturers protocol. It uses an adapter sequence at 3' the end of a a poly dT-primer for first strand synthesis and adds a second adapter complementary to the bases at the 5' end of the transcripts by terminal transferase activity and template switching. Using these adapters it is possible to specifically amplify mRNA enriched for full-length transcripts.

8.3 Cloning and Sanger-sequencing

The obtained cDNA preparations were undirectionally cloned into TOPO2PCR-vectors (Invitrogen, Carlsbad, USA) and TOP10 chemically competent cells (Invitrogen, Carlsbad, USA) were transformed with this construct. The cells were plated on LB-medium-agarose containing Kanamycin (5mg/ml), xGal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and IPTG (Isopropyl-β-D-1-thiogalactopyranosid). After 24h of incubation at 36 °C cells were picked into 96-well micro-liter-plates containing liquid LB-medium and Kanamycin (5mg/ml) and incubated for another 24h. Subsequently 2ml of the cells were used as template for amplification of the insert by PCR using the primers

Forward M13F(GTAAAACGACGGCCAGT) and

Reverse M13R(GGCAGGAAACAGCTATGACC)

in a concentration of 10μM. The protocol for PCR cycling is shown

Inital denaturation	94 °C	5min	
Denaturation	94 °C	30s	
Annealing	54 °C	45s	35 cycles
Elongation	72 °C	2min	
Filnal Elongation	72 °C	10min	

Table 8.1: PCR protocol for insert amplification

Amplification products were controlled on gel and cleaned using SAP (Shrimp Alkaline Phosphatase) and ExoI (Exonuclease I). Sequencing reactions were performed using the BigDye-Terminator kit and PCR-primers (forward or reverse) in a concentration of 3.5μM and sequenced

on an ABI 3730 DNA Analyzer (Applied Biosystems, Foster City, California, USA). For *A. crassus* the following libraries were prepared:

Ac_197F: Female from Taiwanese aquaculture

Ac_106F: Female from Taiwanese aquaculture

Ac_M175: Male from Taiwanese aquaculture

Ac_FM: Female from Taiwanese aquaculture

Ac_EH1: Same cDNA preparation as Ac_FM, but sequenced by students in a practical

For *Anguilla japonica* the following three libraries:

Aj_Li1: liver of an eel from aquaculture

Aj_Li2: liver of an eel from aquaculture

Aj_Li3: liver of an eel from aquaculture

8.4 Bioinformatic analysis of pilot Sanger-sequencing

The original sequencing-chromatographs ("trace-files") were renamed according to the NERC environmental genomics scheme. "Ac" was used as project-identifier for *Anguillicoloides crassus*, "Aj" for *Anguilla japonica*. In *Anguillicoloides* sequences information on the sequencing primer (forward or reverse PCR primer *Anguilla japonica* sequences were all sequenced using the forward PCR primer) was stored in the middle "library"-field, resulting in names of the following form:

```
   Ac_[\d|\w]{2,4}(f|r)_\d\d\w\d\d
Aj_[\d|\w]{2,4}_\d\d\w\d\d
```

The last field indicates the plate number (two digits), the row (one letter) and the column (two digits) of the corresponding clone. For first quality trimming trace2seq, a tool derived from trace2dbEST (both part of PartiGene (67)) was used, briefly it performs quality trimming using phred(101) and trimming of vector sequences using cross-match(102). The adapters used by the MINT kit were trimmed by supplying them in the vector-file used for trimming along with the TOPO2PCR-vector. After processing with trace2seq additional quality trimming was performed on the produced sequence-files using a custom script. This trimming was intended to remove artificial sequences produced when the sequencing reaction starts at the 3' end of the transcript at the poly-A tail. These sequences typically consist of numerous homo-polymer-runs throughout their length caused by "slippage" of the reaction. The basic perl regular expression used for this was:

```
/(.*A{5,}|T{5,}|G{5,}|C{5,}.*){$lengthfac,}/g
```

Where `$lengthfac` was set to the length of the sequence devided by 70 and rounded to the next integer. So only one homo-polymer-run of more then 5 bases was allowed per 105 bases.

Sequences were screened for host contamination by a comparison of `BLAST` searches against the version of nempep4 and a fish protein database. Sequences producing better bit scores againt fish proteins than nematode proteins were labeled as host-contamination.

Only the trace-files corresponding to the sequences still regarded as good after this step were processed with trace2dbEST. Additionally to the processing of traces already included in trace2seq sequences were preliminary annotated using `BLAST` versus the NCBI-NR non-redundant protein database and EST-submission-files were produced.

8.5 Bioinformatic analysis of 454-pyro-sequencing

Trimming, quality control and assembly

Raw sequences were extracted in fasta format (with the corresponding qualities files) using sffinfo (Roche/454) and screened for adapter sequences of the MINT-amplification-kit using cross-match (102) (with parameters -minscore 20 and -minmatch 10). Seqclean (103) was used to screen poly-A-tails, low quality, repetitive and short (¡100 bases) sequences. In addition all reads were `blast`ed (1e-5 -F F) against the following databases:

- a combined eel-mRNA database consisting of an assembly of sequences from the liver of the Japanese eel sequenced for this purpose (as described above), a sequence assembly of unpublished (sanger-) ESTs (made available to us by Gordon Cramb; University of St Andrews) and from EeelBase (104) a publically availble transcriptome database for the European eel.
- a eel-rRNA database from a rRNA screening of the above and assembly together with publically available rRNA-sequences.
- an *A.crassus* rRNA-database from screening of our dataset against nematode-rRNA, and assembly of these rRNA reads. This database notably also contained xenobiotic rRNA sequences.

Reads mapping to one of the databases with more than 80% of their length and 95% identity were removed from the dataset. Screenig and trimming information was written back into sff-format using `sfffile` (Roch/454).

We used an approach proposed by Kumar and Blaxter(92), combining assemblies from the `mira` (105) and `newbler` (91). Briefly the two assemblies are combined into one using Cap3(106) and only contigs supported by both assemblers are regarded good quaility. For further details see the supplementary methods.

Post assembly classification and taxonomic assesment

After assembly contigs were assesed a second time for host-contamination and other xenobiotics:

The contigs were `blast`ed (with a cut-off 1e-5) against the same databases used prior to assembly (Eel-mRNA, Eel-rRNA, *A.crassus*-rRNA and additionally against the nucleotide version of nempep4 (107, 108), determining the best hit across databases. These best hits across databases were screened and only such hits involving more then 50% of the

Additionally `blast` (`blastn` e-value cut-off 1e-5) against NCBI-nt and (`blastx` e-value cut-off 1e-5) against NCBI-nt was used to determine taxon-membership of the top hit at the family, phylum and kingdom rank.

Protein prediction and annotation

Proteins were predicted using the `Prot4EST` (version 3.0b) (109): First `blast` searches against a rRNA-database, a mitochondrial database and against uniref100(110) were preformed. Then results were used to predict proteins directly (joining single high scoring pairs, and thereby intorducing gaps and ambiguous bases if needed). Secondly using the codon-usage from `blast`-predictions a simulated transcriptome was generated, reverse translating the *B. malayi* proteom, as training-data-set for `ESTscan`s(111) hidden Markov models. If both `blast`-based prediction and `ESTscan` failed, simply the longest ORF is inferred.

`Blast`-based annotations were inferred using Annot8r (version 1.1.1) (112): Searches were performed against all sequences in uniref100 being annotated with GO-terms, EC-numbers and KEGG-parthways. Up to 10 (possibly contradictory) annotations based on a bitscore cut-off of 55 were obtained for each annotated database.

`SignalP V3.0` (113) was used to predict signal peptide cleavage sites and signal anchor signatures.

SNP analysis

As protein-prediction inferres gaps (e.g from sequencing errors) to predict the most likely protein, not only start- and end-coordinates of open reading frames (ORFs) had to be extracted from the output of `Prot4EST`. We did this in a custom `perl`-script using a `blast`-search with the nucleotide equivalent of the protein as query and the raw sequence as subject. We obtain the hit-coordinates as ORF-coordinates and imputed the `blast`-query as corrected ORF-sequencs.

We mapped the raw reads against the the complete unigene set, with the imputed sequences for those contigs with proteins predicted, using ssaha2 (with parameters -kmer 13 -skip 3 -seeds 6 -score 100 -cmatch 10 -ckmer 6 -output sam -best 1).

`pileup`-files were produced using `samtools` (114), discarding sequences mapping to multiple regions with the best hit. VarScan (115) (pileup2snp) was used with default parameters on `pileup`-files. This output was further screened as described in the results part of the manuscript.

Gene-expression analysis

For NlaIII-tag-sequencing total RNA was prepared as described above from a worm from the Polish sampling site. A sequence-tag libray was created following the protocol supplied by Illmina for this method. Briefly after synthesis of cDNA on oligo(dt)-beads, this cDNA is digested with the enzyme NlaIII (restriction site "CATG"). After ligation of an adaptor containing its restriction site the enzyme MmeI cuts 17 bases downstream of its binding site generating a sequence tag of in total 21 bases.

For 454 reads, read counts were obtained from the mapping to imputed sequence descrebed above. Tag-sequences were mapped using `BWA`(116). And read counts extracted using `Samtools`.

The R-package DESeq(117) was used to normalize for library-size and analyse statistical significance of differential expression.

References

[1] A Kuwahara, H Niimi, and H Itagaki. **Studies on a nematode parasitic in the air bladder of the eel I. Descriptions of Anguillicola crassa sp. n. (Philometridea, Anguillicolidae)**. *Japanese Journal for Parasitology*, **23**(5):275–279, 1974. 1

[2] Frantisek Moravec. *Dracunculoid and anguillicoloid nematodes parasitic in vertebrates*. Academia, 2006. 1, 8

[3] B Sures, K Knopf, and H Taraschewski. **Development of Anguillicola crassus (Dracunculoidea, Anguillicolidae) in experimentally infected Balearic congers Ariosoma balearicum (Anguilloidea, Congridae)**. *Diseases of Aquatic Organisms*, **39**(1):75–8, December 1999. 1

[4] H. Taraschewski. **Hosts and Parasites as Aliens**. *Journal of Helminthology*, **80**(02):99–128, 2007. 1

[5] R. S. Kirk. **The impact of Anguillicola crassus on European eels**. *Fisheries Management & Ecology*, **10**(6):385–394, 2003. 1, 2

[6] Lamia Gargouri Ben Abdallah and Fadhila Maamouri. **Spatio-temporal dynamics of the nematode Anguillicola crassus in Northeast Tunisian lagoons**. *Comptes Rendus Biologies*, **329**(10):785–789, October 2006. 1

[7] Abdechahid Loukili and Driss Belghyti. **The dynamics of the nematode Anguillicola crassus, Kuvahara 1974 in eel Anguilla anguilla (L. 1758) in the Sebou estuary (Morocco)**. *Parasitology Research*, **100**(4):683–686, March 2007. 1

[8] A. Kristmundsson and S. Helgason. **Parasite communities of ecls Anguilla anguilla in freshwater and marine habitats in Iceland in comparison with other parasite communities of eels in Europe**. *Folia Parasitologica*, **54**(2):141, 2007. 1

[9] K. Knopf, J. Wuertz, B. Sures, and H. Taraschewski. **Impact of low water temperature on the development of Anguillicola crassus in the final host Anguilla anguilla**. *Diseases of Aquatic Organisms*, **33**:143–149, 1998. 1

[10] R S Kirk, C R Kennedy, and J W Lewis. **Effect of salinity on hatching, survival and infectivity of Anguillicola crassus (Nematoda: Dracunculoidea) larvae**. *Diseases of Aquatic Organisms*, **40**(3):211–8, April 2000. 1

[11] Sébastien Wielgoss, Horst Taraschewski, Axel Meyer, and Thierry Wirth. **Population structure of the parasitic nematode Anguillicola crassus, an invader of declining North Atlantic eel stocks**. *Molecular Ecology*, **17**(15):3478–95, August 2008. 1, 2

[12] W Neumann. **Schwimblasenparasit Anguillicola bei Aalen**. *Fischer und Teichwirt*, page 322, 1985. 2

[13] H. Koops and F. Hartmann. **Anguillicola-infestations in Germany and in German eel imports**. *Journal of Applied Ichthyology*, **5**(1):41–45, 1989. 2

[14] S. Wielgoss, F. Hollandt, T. Wirth, and A. Meyer. **Genetic signatures in an invasive parasite of Anguilla anguilla correlate with differential stock management**. *J. Fish Biol.*, **77**:191–210, Jul 2010. 2

[15] Münderle. *Ökologische, morphometrische und genetische Untersuchungen an Populationen des invasiven Schwimmblasen-Nematoden Anguillicola crassus aus Europa und Taiwan*. PhD thesis, University of Karlsruhe, 2005. 2, 4

[16] Pierre Sasal, Horst Taraschewski, Pierre Valade, Henri Grondin, Sébastien Wielgoss, and František Moravec. **Parasite communities in eels of the Island of Reunion (Indian Ocean): a lesson in parasite introduction**. *Parasitology Research*, **102**(6):1343–1350, May 2008. 2, 3

[17] LT Fries, DJ Williams, and Sken Johnson. **Occurrence of Anguillicola crassus, an exotic parasitic swim bladder nematode of eels, in the Southeastern United States**. *Transactions of the American Fisheries Society*, **125**(5):794–797, 1996. 2

[18] A. M. Barse and D. H. Secor. **An exotic nematode parasite of the American eel**. *Fisheries*, **24**(2):6–10, 1999. 2

[19] Ann M. Barse, Scott A. McGuire, Melissa A. Vinores, Laura E. Eierman, and Julie A. Weeder. **The swimbladder nematode Anguillicola crassus in American eels (Anguilla rostrata) from middle and upper regions of Chesapeake bay**. *Journal of Parasitology*, **87**(6):1366–1370, December 2001. 2

[20] Frantisek Moravec, Kazuya Nagasawa, and Munenori Miyakawa. **First record of ostracods as natural intermediate hosts of Anguillicola crassus, a pathogenic swimbladder parasite of eels Anguilla spp**. *Diseases of Aquatic Organisms*, **66**(2):171–3, September 2005. 3

[21] O.L.M. Haenen, T.A.M. van Wijngaarden, M.H.T. van der Heijden, J. Höglund, J.B.J.W. Cornelissen, L.A.M.G. van Leengoed, F.H.M. Borgsteede, and W.B. van Muiswinkel. **Effects of experimental infections with different doses of Anguillicola crassus (Nematoda, Dracunculoidea) on European eel (Anguilla anguilla)**. *Aquaculture*, **141**(1-2):101–8, July 2006. PMID: 16956057. 3

[22] M Polzer and H Taraschewski. **Identification and characterization of the proteolytic enzymes in the developmental stages of the eel-pathogenic nematode Anguillicola crassus**. *Parasitology Research*, **79**(1):24–7, 1993. 3

REFERENCES

[23] D. De Charleroy, L. Grisez, K. Thomas, C. Belpaire, and F. Ollevier. **The life cycle of Anguillicola crassus**. *Diseases of Aquatic Organisms*, **8**(2):77–84, 1990. 3

[24] M. Pietrock and T. Meinelt. **Dynamics of Anguillicola Crassus Larval Infections in a Paratenic Host, the Ruffe (Gymnocephalus Cernuus) from the Oder River on the Border of Germany and Poland**. *Journal of Helminthology*, **76**(03):235–240, 2002. 3

[25] K. Thomas and F. Ollevier. **Paratenic hosts of the swimbladder nematode Anguillicola crassus**. *Diseases of Aquatic Organisms*, **13**:165–174, 1992. 3

[26] J Wrtz and H Taraschewski. **Histopathological changes in the swimbladder wall of the European eel Anguilla anguilla due to infections with Anguillicola crassus**. *Diseases of Aquatic Organisms*, **14**(39):121–134, 2000. 3

[27] A Beregi, K Molnár, L Békési, and C Székely. **Radiodiagnostic method for studying swimbladder inflammation caused by Anguillicola crassus (Nematoda: Dracunculoidea)**. *Diseases of Aquatic Organisms*, **34**(2):155–60, October 1998. 3

[28] Matthew J Gollock, Clive R Kennedy, and J Anne Brown. **Physiological responses to acute temperature increase in European eels Anguilla anguilla infected with Anguillicola crassus**. *Diseases of Aquatic Organisms*, **64**(3):223–8, May 2005. 3

[29] A.P. Palstra, D.F.M. Heppener, V.J.T. van Ginneken, C. Sz?kely, and G.E.E.J.M. van den Thillart. **Swimming performance of silver eels is severely impaired by the swim-bladder parasite Anguillicola crassus**. *Journal of Experimental Marine Biology and Ecology*, **352**(1):244–256, November 2007. 3

[30] B. Sures and K. Knopf. **Parasites as a threat to freshwater eels?** *Science*, **304**(5668):209–11, Apr 2004. 3

[31] T. Wirth and L. Bernatchez. **Decline of North Atlantic eels: a fatal synergy?** *Proc. Biol. Sci.*, **270**:681–688, Apr 2003. 3

[32] J Würtz, K Knopf, and H Taraschewski. **Distribution and prevalence of Anguillicola crassus (Nematoda) in eels Anguilla anguilla of the rivers Rhine and Naab, Germany**. *Diseases of Aquatic Organisms*, **32**(2):137–43, March 1998. 3

[33] F S Lefebvre and A J Crivelli. **Anguillicolosis: dynamics of the infection over two decades**. *Diseases of Aquatic Organisms*, **62**(3):227–32, December 2004. 3, 5

[34] K Knopf and M Mahnke. **Differences in susceptibility of the European eel (Anguilla anguilla) and the Japanese eel (Anguilla japonica) to the swimbladder nematode Anguillicola crassus**. *Parasitology*, **129**(Pt 4):491–6, October 2004. 3, 5

[35] G. Fazio, P. Sasal, C. Da Silva, B. Fumet, J. Boissier, R. Lecomte-Finiger, and H. Monè. **Regulation of Anguillicola crassus (Nematoda) infrapopulations in their definitive host, the European eel, Anguilla anguilla**. *Parasitology*, **135**(-1):1–10, 2008. 5

[36] M Münderle, H Taraschewski, B Klar, C W Chang, J C Shiao, K N Shen, J T He, S H Lin, and W N Tzeng. **Occurrence of Anguillicola crassus (Nematoda: Dracunculoidea) in Japanese eels Anguilla japonica from a river and an aquaculture unit in SW Taiwan**. *Diseases of Aquatic Organisms*, **71**(2):101–8, July 2006. 5

[37] K Knopf. **The swimbladder nematode Anguillicola crassus in the European eel Anguilla anguilla and the Japanese eel Anguilla japonica: differences in susceptibility and immunity between a recently colonized host and the original host**. *Journal of Helminthology*, **80**(2):129–36, June 2006. 5

[38] K Knopf and R Lucius. **Vaccination of eels (Anguilla japonica and Anguilla anguilla) against Anguillicola crassus with irradiated L3**. *Parasitology*, **135**(5):633–40, April 2008. 5

[39] Emanuel Heitlinger, Dominik Laetsch, Urszula Weclawski, Yu-San Han, and Horst Taraschewski. **Massive encapsulation of larval Anguillicoloides crassus in the intestinal wall of Japanese eels**. *Parasites and Vectors*, **2**(1):48, 2009. 5

[40] J. M. Pujolar, G. A. De Leo, E. Ciccotti, and L. Zane. **Genetic composition of Atlantic and Mediterranean recruits of European eel Anguilla anguilla based on EST-linked microsatellite loci**. *J. Fish Biol.*, **74**:2034–2046, Jun 2009. 5

[41] T. Wirth and L. Bernatchez. **Genetic evidence against panmixia in the European eel**. *Nature*, **409**:1037–1040, Feb 2001. 5

[42] S. Palm, J. Dannewitz, T. Prestegaard, and H. Wickstrom. **Panmixia in European eel revisited: no genetic difference between maturing adults from southern and northern Europe**. *Heredity*, **103**:82–89, Jul 2009. 6

[43] T. D. Als, M. M. Hansen, G. E. Maes, M. Castonguay, L. Riemann, K. Aarestrup, P. Munk, H. Sparholt, R. Hanel, and L. Bernatchez. **All roads lead to home: panmixia of European eel in the Sargasso Sea**. *Mol. Ecol.*, **20**:1333–1346, Apr 2011. 6

[44] J. Dannewitz, G. E. Maes, L. Johansson, H. Wickstrom, F. A. Volckaert, and T. Jarvi. **Panmixia in the European eel: a matter of time..** *Proc. Biol. Sci.*, **272**:1129–1137, Jun 2005. 6

[45] A Kerner. **The natural history of plants, their forms, growth, reproduction, and distribution**. *Translated by F. W. Oliver.*, 1895. 6

[46] G Bonnier. **Recherches exprimentales sur ladaptation des plants au climat alpin**. *Ann. Scie. Nat. (Bot.)*, **20**:217–358, 1895. 6

[47] O Kaltz and J. A. Shykoff. **Local adaptation in host-parasite systems**. *Heredity*, pages 361–370, May 1998. 6

[48] H. Taraschewski and F. Moravec. **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 Parasitol (Praha)*, **35**(2):125–146, 1988. 7

[49] **Studies on the helmith fauna of Japan, part 9. Nematodes of fishes.** *Japanese Journal of Zoology*, **6**, 1933. 7

[50] T.H. JOHNSTON and P.M. Mawson. **Some nematodes parasitic in Australian freshwater fish.** *Transactions of the Royal Society of South Australia*, **64**(2):340–352, 1940. 8

[51] Yuki Minegishi, Jun Aoyama, Jun G. Inoue, Masaki Miya, Mutsumi Nishida, and Katsumi Tsukamoto. **Molecular phylogeny and evolution of the freshwater eels genus Anguilla based on the whole mitochondrial genome sequences**. *Molecular Phylogenetics and Evolution*, **34**(1):134–146, 2005. 8

[52] Mark L. Blaxter, Paul De Ley, James R. Garey, Leo X. Liu, Patsy Scheldeman, Andy Vierstraete, Jacques R. Vanfleteren, Laura Y. Mackey, Mark Dorris, Linda M. Frisse, J. T. Vida, and W. Kelley Thomas. **A molecular evolutionary framework for the phylum Nematoda**. *Nature*, **392**(6671):71–75, March 1998. 8

[53] S.?A. NADLER, R.?A. CARRENO, H. MEJ?A-MADRID, J. ULLBERG, C. PAGAN, R. HOUSTON, and J.-P. HUGOT. **Molecular Phylogeny of Clade III Nematodes Reveals Multiple Origins of Tissue Parasitism**. *Parasitology*, **134**(10):1421–1442, 2007. 8

[54] Martina Wijová, Frantisek Moravec, Ales Horák, and Julius Lukes. **Evolutionary relationships of Spirurina (Nematoda: Chromadorea: Rhabditida) with special emphasis on dracunculoid nematodes inferred from SSU rRNA gene sequences**. *International Journal for Parasitology*, **36**(9):1067–75, August 2006. 8

[55] Xingxing Zang and Rick M. Maizels. **Serine proteinase inhibitors from nematodes and the arms race between host and pathogen**. *Trends in Biochemical Sciences*, **26**(3):191–197, March 2001. 12

[56] M. K. Hughes and A. L. Hughes. **Natural selection on Plasmodium surface proteins**. *Mol. Biochem. Parasitol.*, **71**:99–113, Apr 1995. 12

[57] DJ CONWAY and SD POLLEY. **Measuring immune selection**. *Parasitology(London. Print)*, **125**:3–16, 2002. 12

[58] The C. elegans Sequencing Consortium. **Genome sequence of the nematode C. elegans: a platform for investigating biology**. *Science*, **282**:2012–2018, Dec 1998. 12, 13

[59] Mark Blaxter. **Caenorhabditis elegans Is a Nematode**. *Science*, **282**(5396):2041–2046, December 1998. 12

[60] M. B. Gerstein, Z. J. Lu, E. L. Van Nostrand, C. Cheng, B. I. Arshinoff, T. Liu, K. Y. Yip, R. Robilotto, A. Rechtsteiner, K. Ikegami, P. Alves, A. Chateigner, M. Perry, M. Morris, R. K. Auerbach, X. Feng, J. Leng, A. Vielle, W. Niu, K. Rhrissorrakrai, A. Agarwal, R. P. Alexander, G. Barber, C. M. Brdlik, J. Brennan, J. J. Brouillet, A. Carr, M. S. Cheung, H. Clawson, S. Contrino, L. O. Dannenberg, A. F. Dernburg, A. Desai, L. Dick, A. C. Dose, J. Du, T. Egelhofer, S. Ercan, G. Euskirchen, B. Ewing, E. A. Feingold, R. Gassmann, P. J. Good, P. Green, F. Gullier, M. Gutwein, M. S. Guyer, L. Habegger, T. Han, J. G. Henikoff, S. R. Henz, A. Hinrichs, H. Holster, T. Hyman, A. L. Iniguez, J. Janette, M. Jensen, M. Kato, W. J. Kent, E. Kephart, V. Khivansara, E. Khurana, J. K. Kim, P. Kolasinska-Zwierz, E. C. Lai, I. Latorre, A. Leahey, S. Lewis, P. Lloyd, L. Lochovsky, R. F. Lowdon, Y. Lubling, R. Lyne, M. MacCoss, S. D. Mackowiak, M. Mangone, S. McKay, D. Mecenas, G. Merrihew, D. M. Miller, A. Muroyama, J. I. Murray, S. L. Ooi, H. Pham, T. Phippen, E. A. Preston, N. Rajewsky, G. Ratsch, H. Rosenbaum, J. Rozowsky, K. Rutherford, P. Ruzanov, M. Sarov, R. Sasidharan, A. Sboner, P. Scheid, E. Segal, H. Shin, C. Shou, F. J. Slack, C. Slightam, R. Smith, W. C. Spencer, E. O. Stinson, S. Taing, T. Takasaki, D. Vafeados, K. Voronina, G. Wang, N. L. Washington, C. M. Whittle, B. Wu, K. K. Yan, G. Zeller, Z. Zha, M. Zhong, X. Zhou, J. Ahringer, S. Strome, K. C. Gunsalus, G. Micklem, X. S. Liu, V. Reinke, S. K. Kim, L. W. Hillier, S. Henikoff, F. Piano, M. Snyder, L. Stein, J. D. Lieb, and R. H. Waterston. **Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project**. *Science*, **330**:1775–1787, Dec 2010. 12

[61] Lincoln D. Stein, Zhirong Bao, Darin Blasiar, Thomas Blumenthal, Michael R. Brent, Nansheng Chen, Asif Chinwalla, Laura Clarke, Chris Clee, Avril Coghlan, Alan Coulson, Peter D'Eustachio, David H. A. Fitch, Lucinda A. Fulton, Robert E. Fulton, Sam Griffiths-Jones, Todd W. Harris, laDeana W. Hillier, Ravi Kamath, Patricia E. Kuwabara, Elaine R. Mardis, Marco A. Marra, Tracie L. Miner, Patrick Minx, James C. Mullikin, Robert W. Plumb, Jane Rogers, Jacqueline E. Schein, Marc Sohrmann, John Spieth, Jason E. Stajich, Chaochun Wei, David Willey, Richard K. Wilson, Richard Durbin, and Robert H. Waterston. **The Genome Sequence of Caenorhabditis briggsae: A Platform for Comparative Genomics**. *PLoS Biology*, **1**(2):e45 EP –, November 2003. 12

[62] Elodie Ghedin, Shiliang Wang, David Spiro, Elisabet Caler, Qi Zhao, Jonathan Crabtree, Jonathan E. Allen, Arthur L. Delcher, David B. Guiliano, Diego Miranda-Saavedra, Samuel V. Angiuoli, Todd Creasy, Paolo Amedeo, Brian Haas, Najib M. El-Sayed, Jennifer R. Wortman, Tamara Feldblyum, Luke Tallon, Michael Schatz, Martin Shumway, Hean Koo, Steven L. Salzberg, Seth Schobel, Mihaela Pertea, Mihai Pop, Owen White, Geoffrey J. Barton, Clotilde K. S. Carlow, Michael J. Crawford, Jennifer Daub, Matthew W. Dimmic, Chris F. Estes, Jeremy M. Foster, Mehul Ganatra, William F. Gregory, Nicholas M. Johnson, Jinming Jin, Richard Komuniecki, Ian Korf, Sanjay Kumar, Sandra Laney, Ben-Wen Li, Wen Li, Tim H. Lindblom, Sara Lustigman, Dong Ma, Claude V. Maina, David M. A. Martin, James P. McCarter, Larry McReynolds, Makedonka Mitreva, Thomas B. Nutman, John Parkinson, Jose M. Peregrin-Alvarez, Catherine Poole, Qinghu Ren, Lori Saunders, Ann E. Sluder, Katherine Smith, Mario Stanke, Thomas R. Unnasch, Jenna Ware, Aguan D. Wei, Gary Weil, Deryck J. Williams, Yinhua Zhang, Steven A. Williams, Claire Fraser-Liggett, Barton Slatko, Mark L. Blaxter, and Alan L. Scott. **Draft Genome of the Filarial Nematode Parasite Brugia malayi**. *Science*, **317**(5845):1756–1760, September 2007. 12

[63] M. Mitreva, D. P. Jasmer, D. S. Zarlenga, Z. Wang, S. Abubucker, J. Martin, C. M. Taylor, Y. Yin, L. Ful-

REFERENCES

ton, P. Minx, S. P. Yang, W. C. Warren, R. S. Fulton, V. Bhonagiri, X. Zhang, K. Hallsworth-Pepin, S. W. Clifton, J. P. McCarter, J. Appleton, E. R. Mardis, and R. K. Wilson. **The draft genome of the parasitic nematode Trichinella spiralis**. *Nat. Genet.*, **43**:228–235, Mar 2011. 12

[64] P. Abad, J. Gouzy, J. M. Aury, P. Castagnone-Sereno, E. G. Danchin, E. Deleury, L. Perfus-Barbeoch, V. Anthouard, F. Artiguenave, V. C. Blok, M. C. Caillaud, P. M. Coutinho, C. Dasilva, F. De Luca, F. Deau, M. Esquibet, T. Flutre, J. V. Goldstone, N. Hamamouch, T. Hewezi, O. Jaillon, C. Jubin, P. Leonetti, M. Magliano, T. R. Maier, G. V. Markov, P. McVeigh, G. Pesole, J. Poulain, M. Robinson-Rechavi, E. Sallet, B. Segurens, D. Steinbach, T. Tytgat, E. Ugarte, C. van Ghelder, P. Veronico, T. J. Baum, M. Blaxter, T. Bleve-Zacheo, E. L. Davis, J. J. Ewbank, B. Favery, E. Grenier, B. Henrissat, J. T. Jones, V. Laudet, A. G. Maule, H. Quesneville, M. N. Rosso, T. Schiex, G. Smant, J. Weissenbach, and P. Wincker. **Genome sequence of the metazoan plant-parasitic nematode Meloidogyne incognita**. *Nat. Biotechnol.*, **26**:909–915, Aug 2008. 12

[65] C. H. Opperman, D. M. Bird, V. M. Williamson, D. S. Rokhsar, M. Burke, J. Cohn, J. Cromer, S. Diener, J. Gajan, S. Graham, T. D. Houfek, Q. Liu, T. Mitros, J. Schaff, R. Schaffer, E. Scholl, B. R. Sosinski, V. P. Thomas, and E. Windham. **Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism**. *Proc. Natl. Acad. Sci. U.S.A.*, **105**:14802–14807, Sep 2008. 12

[66] T. Kikuchi, J. A. Cotton, J. J. Dalzell, K. Hasegawa, N. Kanzaki, P. McVeigh, T. Takanashi, I. J. Tsai, S. A. Assefa, P. J. Cock, T. D. Otto, M. Hunt, A. J. Reid, A. Sanchez-Flores, K. Tsuchihara, T. Yokoi, M. C. Larsson, J. Miwa, A. G. Maule, N. Sahashi, J. T. Jones, and M. Berriman. **Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen Bursaphelenchus xylophilus**. *PLoS Pathog.*, **7**:e1002219, Sep 2011. 12

[67] John Parkinson, Alasdair Anthony, James Wasmuth, Ralf Schmid, Ann Hedley, and Mark Blaxter. **PartiGene–constructing partial genomes**. *Bioinformatics*, **20**(9):1398–1404, June 2004. 12, 35

[68] R. M. Maizels, N. Gomez-Escobar, W. F. Gregory, J. Murray, and X. Zang. **Immune evasion genes from filarial nematodes**. *Int. J. Parasitol.*, **31**:889–898, Jul 2001. 13

[69] Natalia Gomez-Escobar, William F. Gregory, Collette Britton, Linda Murray, Craig Corton, Neil Hall, Jen Daub, Mark L. Blaxter, and Rick M Maizels. **Abundant larval transcript-1 and -2 genes from Brugia malayi: diversity of genomic environments but conservation of 5' promoter sequences functional in Caenorhabditis elegans**. *Molecular and Biochemical Parasitology*, **125**(1-2):59–71, 2002. 13

[70] Mark L. Blaxter, Nithyakalyani Raghavan, Inca Ghosh, David Guiliano, Wenhong Lu, Steven A. Williams, Barton Slatko, and Alan L. Scott. **Genes expressed in Brugia malayi infective third stage larvae**. *Molecular and Biochemical Parasitology*, **77**(1):77–93, April 1996. 13

[71] J. DAUB, A. LOUKAS, DI PRITCHARD, and M. BLAXTER. **A survey of genes expressed in adults of the human hookworm, Necator americanus**. *Parasitology*, **120**(02):171–184, 2000. 13

[72] John Parkinson, Makedonka Mitreva, Claire Whitton, Marian Thomson, Jennifer Daub, John Martin, Ralf Schmid, Neil Hall, Bart Barrell, Robert H Waterston, James P McCarter, and Mark L Blaxter. **A transcriptomic analysis of the phylum Nematoda**. *Nat Genet*, **36**(12):1259–1267, December 2004. 13

[73] Makedonka Mitreva, Douglas P. Jasmer, Judith Appleton, John Martin, Mike Dante, Todd Wylie, Sandra W. Clifton, Robert H. Waterston, and James P. McCarter. **Gene discovery in the adenophorean nematode Trichinella spiralis: an analysis of transcription from three life cycle stages**. *Molecular and Biochemical Parasitology*, **137**(2):277–291, October 2004. 13

[74] Yvonne Harcus, John Parkinson, Cecilia Fernandez, Jennifer Daub, Murray Selkirk, Mark Blaxter, and Rick Maizels. **Signal sequence analysis of expressed sequence tags from the nematode Nippostrongylus brasiliensis and the evolution of secreted proteins in parasites**. *Genome Biology*, **5**(6):R39, 2004. 13

[75] Shivashankar H. Nagaraj, Robin B. Gasser, and Shoba Ranganathan. **Needles in the EST Haystack: Large-Scale Identification and Analysis of Excretory-Secretory (ES) Proteins in Parasitic Nematodes Using Expressed Sequence Tags (ESTs)**. *PLoS Neglected Tropical Diseases*, **2**(9):e301, 2008. 13

[76] Y. Moreno and T. G. Geary. **Stage- and gender-specific proteomic analysis of Brugia malayi excretory-secretory products**. *PLoS Negl Trop Dis*, **2**:e326, 2008. 13

[77] S. Bennuru, R. Semnani, Z. Meng, J. M. Ribeiro, T. D. Veenstra, and T. B. Nutman. **Brugia malayi excreted/secreted proteins at the host/parasite interface: stage- and gender-specific proteomic profiling**. *PLoS Negl Trop Dis*, **3**:e410, 2009. 13

[78] J. P. Hewitson, Y. Harcus, J. Murray, M. van Agtmaal, K. J. Filbey, J. R. Grainger, S. Bridgett, M. L. Blaxter, P. D. Ashton, D. A. Ashford, R. S. Curwen, R. A. Wilson, A. A. Dowle, and R. M. Maizels. **Proteomic analysis of secretory products from the model gastrointestinal nematode Heligmosomoides polygyrus reveals dominance of Venom Allergen-Like (VAL) proteins**. *J Proteomics*, **74**:1573–1594, Aug 2011. 13

[79] A. P. Yatsuda, J. Krijgsveld, A. W. Cornelissen, A. J. Heck, and E. de Vries. **Comprehensive analysis of the secreted proteins of the parasite Haemonchus contortus reveals extensive sequence variation and differential immune recognition**. *J. Biol. Chem.*, **278**:16941–16951, May 2003. 13

[80] F. Sanger, S. Nicklen, and A. R. Coulson. **DNA sequencing with chain-terminating inhibitors**. *Proc. Natl. Acad. Sci. U.S.A.*, **74**:5463–5467, Dec 1977. 13

[81] H. Swerdlow and R. Gesteland. **Capillary gel electrophoresis for rapid, high resolution DNA sequencing**. *Nucleic Acids Res.*, **18**:1415–1419, Mar 1990. 13

[82] F. R. Blattner, G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. **The complete genome sequence of Escherichia coli K-12**. *Science*, **277**:1453–1462, Sep 1997. 13

[83] A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, E. J. Louis, H. W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, and S. G. Oliver. **Life with 6000 genes**. *Science*, **274**:563–567, Oct 1996. 13

[84] M.D. Adams, S.E. Celniker, R.A. Holt, C.A. Evans, J.D. Gocayne, P.G. Amanatides, S.E. Scherer, P.W. Li, R.A. Hoskins, R.F. Galle, et al. **The genome sequence of Drosophila melanogaster**. *Science*, **287**(5461):2185, 2000. 13

[85] R. H. Waterston, K. Lindblad-Toh, E. Birney, J. Rogers, J. F. Abril, P. Agarwal, R. Agarwala, R. Ainscough, M. Alexandersson, P. An, S. E. Antonarakis, J. Attwood, R. Baertsch, J. Bailey, K. Barlow, S. Beck, E. Berry, B. Birren, T. Bloom, P. Bork, M. Botcherby, N. Bray, M. R. Brent, D. G. Brown, S. D. Brown, C. Bult, J. Burton, J. Butler, R. D. Campbell, P. Carninci, S. Cawley, F. Chiaromonte, A. T. Chinwalla, D. M. Church, M. Clamp, C. Clee, F. S. Collins, L. L. Cook, R. R. Copley, A. Coulson, O. Couronne, J. Cuff, V. Curwen, T. Cutts, M. Daly, R. David, J. Davies, K. D. Delehaunty, J. Deri, E. T. Dermitzakis, C. Dewey, N. J. Dickens, M. Diekhans, S. Dodge, I. Dubchak, D. M. Dunn, S. R. Eddy, L. Elnitski, R. D. Emes, P. Eswara, E. Eyras, A. Felsenfeld, G. A. Fewell, P. Flicek, K. Foley, W. N. Frankel, L. A. Fulton, R. S. Fulton, T. S. Furey, D. Gage, R. A. Gibbs, G. Glusman, S. Gnerre, N. Goldman, L. Goodstadt, D. Grafham, T. A. Graves, E. D. Green, S. Gregory, R. Guigo, M. Guyer, R. C. Hardison, D. Haussler, Y. Hayashizaki, L. W. Hillier, A. Hinrichs, W. Hlavina, T. Holzer, F. Hsu, A. Hua, T. Hubbard, A. Hunt, I. Jackson, D. B. Jaffe, L. S. Johnson, M. Jones, T. A. Jones, A. Joy, M. Kamal, E. K. Karlsson, D. Karolchik, A. Kasprzyk, J. Kawai, E. Keibler, C. Kells, W. J. Kent, A. Kirby, D. L. Kolbe, I. Korf, R. S. Kucherlapati, E. J. Kulbokas, D. Kulp, T. Landers, J. P. Leger, S. Leonard, I. Letunic, R. Levine, J. Li, M. Li, C. Lloyd, S. Lucas, B. Ma, D. R. Maglott, E. R. Mardis, L. Matthews, E. Mauceli, J. H. Mayer, M. McCarthy, W. R. McCombie, S. McLaren, K. McLay, J. D. McPherson, J. Meldrim, B. Meredith, J. P. Mesirov, W. Miller, T. L. Miner, E. Mongin, K. T. Montgomery, M. Morgan, R. Mott, J. C. Mullikin, D. M. Muzny, W. E. Nash, J. O. Nelson, M. N. Nhan, R. Nicol, Z. Ning, C. Nusbaum, M. J. O'Connor, Y. Okazaki, K. Oliver, E. Overton-Larty, L. Pachter, G. Parra, K. H. Pepin, J. Peterson, P. Pevzner, R. Plumb, C. S. Pohl, A. Poliakov, T. C. Ponce, C. P. Ponting, S. Potter, M. Quail, A. Reymond, B. A. Roe, K. M. Roskin, E. M. Rubin, A. G. Rust, R. Santos, V. Sapojnikov, B. Schultz, J. Schultz, M. S. Schwartz, S. Schwartz, C. Scott, S. Seaman, S. Searle, T. Sharpe, A. Sheridan, R. Shownkeen, S. Sims, J. B. Singer, G. Slater, A. Smit, D. R. Smith, B. Spencer, A. Stabenau, N. Stange-Thomann, C. Sugnet, M. Suyama, G. Tesler, J. Thompson, D. Torrents, E. Trevaskis, J. Tromp, C. Ucla, A. Ureta-Vidal, J. P. Vinson, A. C. Von Niederhausern, C. M. Wade, M. Wall, R. J. Weber, R. B. Weiss, M. C. Wendl, A. P. West, K. Wetterstrand, R. Wheeler, S. Whelan, J. Wierzbowski, D. Willey, S. Williams, R. K. Wilson, E. Winter, K. C. Worley, D. Wyman, S. Yang, S. P. Yang, E. M. Zdobnov, M. C. Zody, and E. S. Lander. **Initial sequencing and comparative analysis of the mouse genome**. *Nature*, **420**:520–562, Dec 2002. 13

[86] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. **The sequence**

REFERENCES

of the human genome. *Science*, **291**:1304–1351, Feb 2001. 13

[87] F. Sanger, A. R. Coulson, B. G. Barrell, A. J. Smith, and B. A. Roe. **Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing**. *J. Mol. Biol.*, **143**:161–178, Oct 1980. [PubMed:6260957]. 15

[88] R. Staden. **A strategy of DNA sequencing employing computer programs**. *Nucleic Acids Res.*, **6**:2601–2610, Jun 1979. 15

[89] T. R. Gingeras and R. J. Roberts. **Steps toward computer analysis of nucleotide sequences**. *Science*, **209**:1322–1328, Sep 1980. 15

[90] J. R. Miller, S. Koren, and G. Sutton. **Assembly algorithms for next-generation sequencing data**. *Genomics*, **95**:315–327, Jun 2010. 15

[91] M. Margulies, M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y. J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. H. Ho, C. H. Ho, G. P. Irzyk, S. C. Jando, M. L. Alenquer, T. P. Jarvie, K. B. Jirage, J. B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg. **Genome sequencing in microfabricated high-density picolitre reactors**. *Nature*, **437**:376–380, Sep 2005. 15, 36

[92] S. Kumar and M. L. Blaxter. **Comparing de novo assemblers for 454 transcriptome data**. *BMC Genomics*, **11**:571, Oct 2010. 15, 36

[93] F. Goetz, D. Rosauer, S. Sitar, G. Goetz, C. Simchick, S. Roberts, R. Johnson, C. Murphy, C. R. Bronte, and S. Mackenzie. **A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (Salvelinus namaycush)**. *Mol. Ecol.*, **19 Suppl 1**:176–196, Mar 2010. 15

[94] J. H. Malone and B. Oliver. **Microarrays, deep sequencing and the true measure of the transcriptome**. *BMC Biol.*, **9**:34, 2011. 16

[95] Z. Wang, M. Gerstein, and M. Snyder. **RNA-Seq: a revolutionary tool for transcriptomics**. *Nat. Rev. Genet.*, **10**:57–63, Jan 2009. 16

[96] H. Matsumura, K. Yoshida, S. Luo, D. H. Kruger, G. Kahl, G. P. Schroth, and R. Terauchi. **High-throughput SuperSAGE**. *Methods Mol. Biol.*, **687**:135–146, 2011. 16

[97] V. E. Velculescu, L. Zhang, B. Vogelstein, and K. W. Kinzler. **Serial analysis of gene expression**. *Science*, **270**:484–487, Oct 1995. 16

[98] S. Via. **The Ecological Genetics of Speciation**. *The American Naturalist*, **159**(S3):1–7, 2002. 16

[99] C. J. McManus, J. D. Coolon, M. O. Duff, J. Eipper-Mains, B. R. Graveley, and P. J. Wittkopp. **Regulatory divergence in Drosophila revealed by mRNA-seq**. *Genome Res.*, **20**:816–825, Jun 2010. 16

[100] W. Haerty and R. S. Singh. **Gene regulation divergence is a major contributor to the evolution of Dobzhansky-Muller incompatibilities between species of Drosophila**. *Mol. Biol. Evol.*, **23**:1707–1714, Sep 2006. 16

[101] Brent Ewing, LaDeana Hillier, Michael C. Wendl, and Phil Green. **Base-Calling of automated sequencer traces using Phred. I. Accuracy Assessment**. *Genome Res.*, **8**(3):175–185, March 1998. 35

[102] Phil Green. *PHRAP documentation.*, 1994. 35, 36

[103] G. Pertea, X. Huang, F. Liang, V. Antonescu, R. Sultana, S. Karamycheva, Y. Lee, J. White, F. Cheung, B. Parvizi, J. Tsai, and J. Quackenbush. **TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets**. *Bioinformatics*, **19**:651–652, Mar 2003. 36

[104] A. Coppe, J. M. Pujolar, G. E. Maes, P. F. Larsen, M. M. Hansen, L. Bernatchez, L. Zane, and S. Bortoluzzi. **Sequencing, de novo annotation and analysis of the first Anguilla anguilla transcriptome: EeelBase opens new perspectives for the study of the critically endangered European eel**. *BMC Genomics*, **11**:635, 2010. 36

[105] B. Chevreux, T. Pfisterer, B. Drescher, A. J. Driesel, W. E. Muller, T. Wetter, and S. Suhai. **Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs**. *Genome Res.*, **14**:1147–1159, Jun 2004. 36

[106] X. Huang and A. Madan. **CAP3: A DNA sequence assembly program**. *Genome Res.*, **9**:868–877, Sep 1999. 36

[107] John Parkinson, Claire Whitton, Ralf Schmid, Marian Thomson, and Mark Blaxter. % bf NEMBASE: a resource for parasitic nematode ESTs. *Nucl. Acids Res.*, **32**(suppl_1):D427–430, 2004. 37

[108] B. Elsworth, J. Wasmuth, and M. Blaxter. **NEMBASE4: The nematode transcriptome resource**. *Int. J. Parasitol.*, **41**:881–894, Jul 2011. 37

[109] James Wasmuth and Mark Blaxter. **prot4EST: Translating Expressed Sequence Tags from neglected genomes**. *BMC Bioinformatics*, **5**(1):187, 2004. 37

[110] A. Bairoch, L. Bougueleret, S. Altairac, V. Amendolia, A. Auchincloss, G. Argoud-Puy, K. Axelsen, D. Baratin, M. C. Blatter, B. Boeckmann, J. Bolleman, L. Bollondi, E. Boutet, S. B. Quintaje, L. Breuza, A. Bridge, E. deCastro, L. Ciapina, D. Coral, E. Coudert, I. Cusin, G. Delbard, D. Dornevil, P. D. Roggli, S. Duvaud, A. Estreicher, L. Famiglietti, M. Feuermann, S. Gehant, N. Farriol-Mathis, S. Ferro, E. Gasteiger, A. Gateau, V. Gerritsen, A. Gos, N. Gruaz-Gumowski, U. Hinz, C. Hulo, N. Hulo, J. James, S. Jimenez, F. Jungo, V. Junker, T. Kappler, G. Keller, C. Lachaize, L. Lane-Guermonprez, P. Langendijk-Genevaux, V. Lara, P. Lemercier, V. Le Saux, D. Lieberherr, T. d. e. O. Lima, V. Mangold, X. Martin, P. Masson, K. Michoud,

M. Moinat, A. Morgat, A. Mottaz, S. Paesano, I. Pedruzzi, I. Phan, S. Pilbout, V. Pillet, S. Poux, M. Pozzato, N. Redaschi, S. Reynaud, C. Rivoire, B. Roechert, M. Schneider, C. Sigrist, K. Sonesson, S. Staehli, A. Stutz, S. Sundaram, M. Tognolli, L. Verbregue, A. L. Veuthey, L. Yip, L. Zuletta, R. Apweiler, Y. Alam-Faruque, R. Antunes, D. Barrell, D. Binns, L. Bower, P. Browne, W. M. Chan, E. Dimmer, R. Eberhardt, A. Fedotov, R. Foulger, J. Garavelli, R. Golin, A. Horne, R. Huntley, J. Jacobsen, M. Kleen, P. Kersey, K. Laiho, R. Leinonen, D. Legge, Q. Lin, M. Magrane, M. J. Martin, C. O'Donovan, S. Orchard, J. O'Rourke, S. Patient, M. Pruess, A. Sitnov, E. Stanley, M. Corbett, G. di Martino, M. Donnelly, J. Luo, P. van Rensburg, C. Wu, C. Arighi, L. Arminski, W. Barker, Y. Chen, Z. Z. Hu, H. K. Hua, H. Huang, R. Mazumder, P. McGarvey, D. A. Natale, A. Nikolskaya, N. Petrova, B. E. Suzek, S. Vasudevan, C. R. Vinayaka, L. S. Yeh, and J. Zhang. **The Universal Protein Resource (UniProt) 2009**. *Nucleic Acids Res.*, **37**:D169–174, Jan 2009. 37

[111] C Iseli, CV Jongeneel, and P Bucher. **ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences.** *Proc Int Conf Intell Syst Mol Biol*, pages 138–148, 1999. 37

[112] Ralf Schmid and Mark L Blaxter. **annot8r: GO, EC and KEGG annotation of EST datasets**. *BMC Bioinformatics*, **9**:180, 2008. 37

[113] O. Emanuelsson, S. Brunak, G. von Heijne, and H. Nielsen. **Locating proteins in the cell using TargetP, SignalP and related tools**. *Nat Protoc*, **2**:953–971, 2007. 37

[114] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Gonalo R. Abecasis, and Richard Durbin. **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics*, **25**(16):2078–2079, 2009. 37

[115] D. C. Koboldt, K. Chen, T. Wylie, D. E. Larson, M. D. McLellan, E. R. Mardis, G. M. Weinstock, R. K. Wilson, and L. Ding. **VarScan: variant detection in massively parallel sequencing of individual and pooled samples**. *Bioinformatics*, **25**:2283–2285, Sep 2009. 37

[116] H. Li and R. Durbin. **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics*, **25**:1754–1760, Jul 2009. 38

[117] S. Anders and W. Huber. **Differential expression analysis for sequence count data**. *Genome Biol.*, **11**:R106, 2010. 38

Declaration

I herewith declare that I have produced this paper without the prohibited assistance of third parties and without making use of aids other than those specified; notions taken over directly or indirectly from other sources have been identified as such. This paper has not previously been presented in identical or similar form to any other German or foreign examination board.

The thesis work was conducted from XXX to YYY under the supervision of PI at ZZZ.

CITY,