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 für 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 mündlichen PrÃijfung:

Abstract

The ability to expand into new environments and niches, despite being highly adapted for survival in their angestammten environment, is a fascinating feat of organisms. The propensity of *Anguillicola crassus* to capture new hosts can serve as a model for an extreme case of this, in which parasites accquire new hosts. Selection in such new environments leading to adaptation is considered a driving force of divergence and thus for the origin of species and biotic diversity.

Gene regulatory networks, as a bridge between genotype and phenotype, are thought to play a central role both in the response to stress (e.g. from sofar unexperienced environmental stressors) and in the divergence and eventually establisment of reproductive barriers between populations.

In the present project the differences in gene-expression in *A. crassus* populations should be illuminated. The parasite was introduced to Europe 30 years ago, spread successfully in a new host and established stable populations.

Zusammenfassung

Die FÃď higkeit sich in neuen Umgebungen und Nieschen auszubreiten, obwohl sie hÃűchst angepasst an ihren angestammten Lebensraum sind, stellt eine faszinierende FÃď higkeit von Lebenwesen dar. Der Wechsel der Wirtsart durch *Anguillicola crassus* kann als Modell fÃijr einen Extremfall dieses Vorganges gesehen werden, bei dem Parasiten neue Wirte besiedeln. Selektion in solch einer neuen Umgebung, die zu einer Anpassung fÃijhrt gilt als eien treibende Kraft fÃijr Divergenz und so zum Entstehen neuer Arten und biologischer Vielfalt. Gen-regulatorische Netzwerke, als eine BrÃijcke zwischen Genotyp and Phenotyp, haben eine zentrale Rolle sowohl in der Antwort auf Stress (etwa durch eine verÃď nderte Umwelt) als auch in der Entwicklung von Barrieren fÃijr die Fortpflanzung.

Im hier vorgestellen Projekt sollen die Unterschiede im Transkriptom zweier Populationen von *A. crassus* untersucht werden. Der Parasit wurde vor 30 Jahren nach Europa eingeschleppt, wo er sich erfolgreich in einer neuen Wirtsart ausbreitet und etablierte.

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 dfinitive host seems to be limited to this genus (3). However the nematode displayed a high versatility and plasticity in most other aspects of it's life, and this has been proposed as one of the reasons for its success invading new continents (4).

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 thus 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 history of introduction and dispersal. Therefore the process of introduction and spread can be

considered very well illuminated:

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 source of introduction (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). However a recent study of Laetsch et all !!CITE found additional haplotypes for Cytochrome C oxidase subunit II in Turkey, and a second introduction to the Eastern Meditereanean seems possible. These Turkish haplotypes cluster with Taiwanese haplotypes and the introduction source would be similar to the main itroduction.

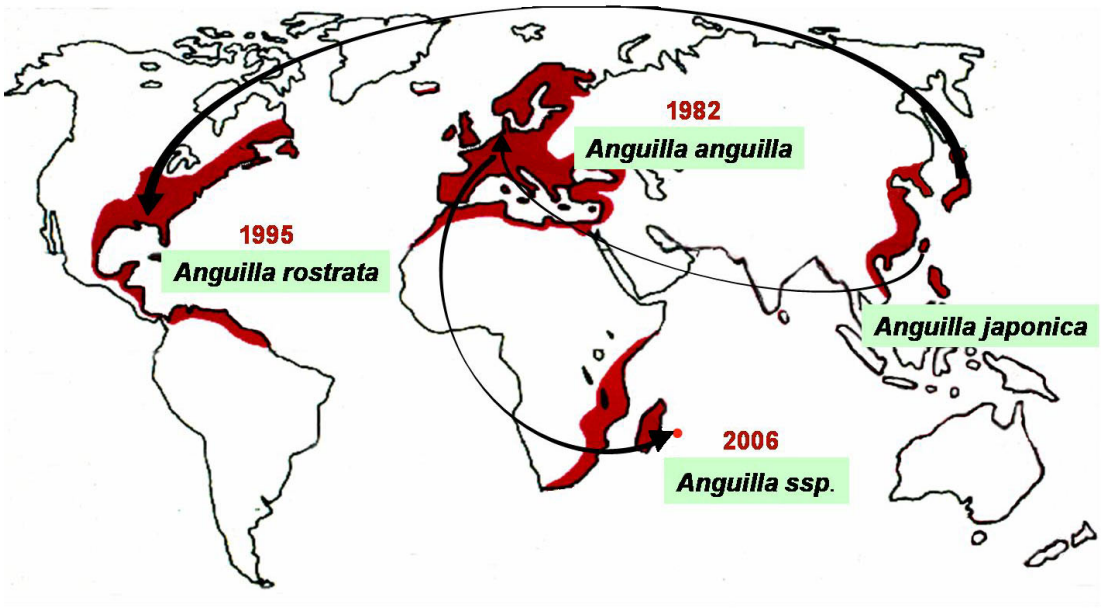

Figure 1.1: Transcontinental dispersal of *A. crassus:* - Invasions of different continents by different source-populationa are illsutrated using arrows. Red color indicates the range of the eel species targeted by the invasion. Modified form (15), based on data reviewed in (5) and newer findings in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

A second colonization of *A. crassus*, 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*. Laetsch et al. CITE!! showed that all source populations for different intoductions are from one of two clearly seperated clades of *A. crassus*.

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 hosts L2 larvae develop to L3 larvae infective for 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 the L4 stage 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 eel's *ductus pneumaticus* into it's intestine 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 reported from the Asian range of the parasite 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 thickening (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 their spawning migration (29). Anguillicolosis (the condition caused by *Anguillicola*) 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). In the natural host in Asia prevalences and intesities are lower (34).

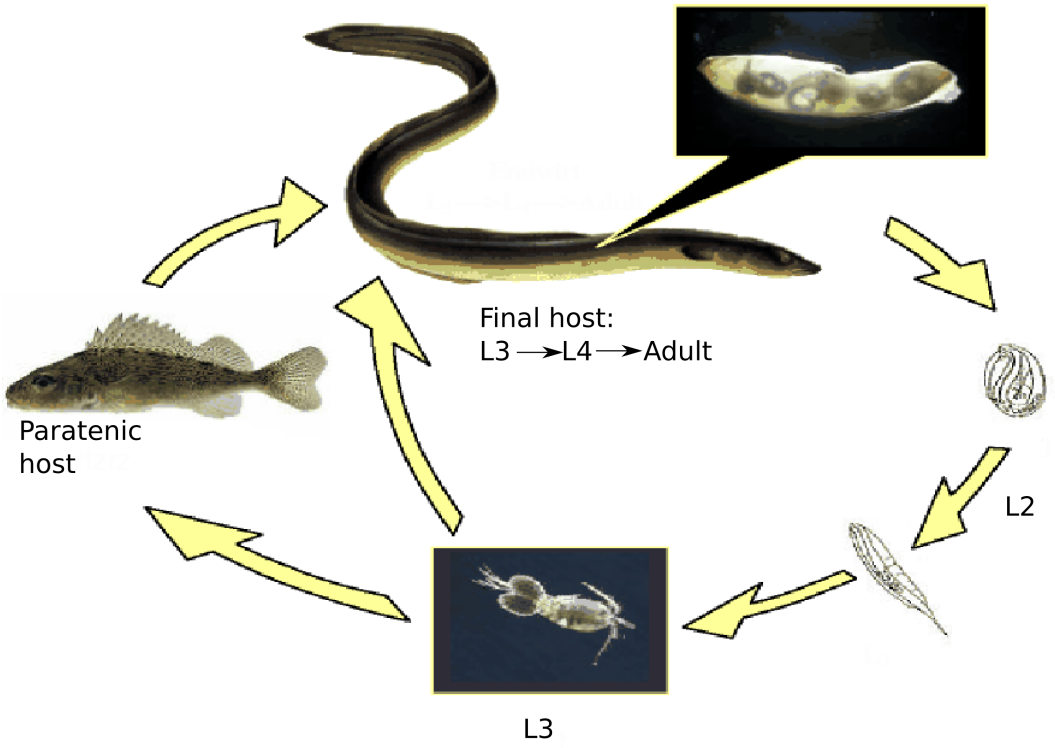

Figure 1.2: Life-cycle of *A. crassus* - Adult females deposit already hatched L2 in the lumen of the swimbladder. Larvae migrate through the *ductus pneumaticus* and the intestine into the open water.Copepodes serve as intermediate host where infectiv L3-larvea develop. These can be transported and 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).

These differences in abundance and intensity of *A. crassus* infections 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 (35). High epidemiological parameters are attributed to the inadequate immune-response of the European Eel (36). While the Japanese eel is capable of killing larvae of the parasite after vaccination (37) or under high infection pressure (38), responses in *An. anguilla* have hallmarks of pathology. Recently, data from experimental infections of European eels with *A. crassus* have been published (39) that show that in this host the parasite undergoes (under experimental conditions) a density-dependent regulation keeping the number of worms within a certain range.

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 (35).

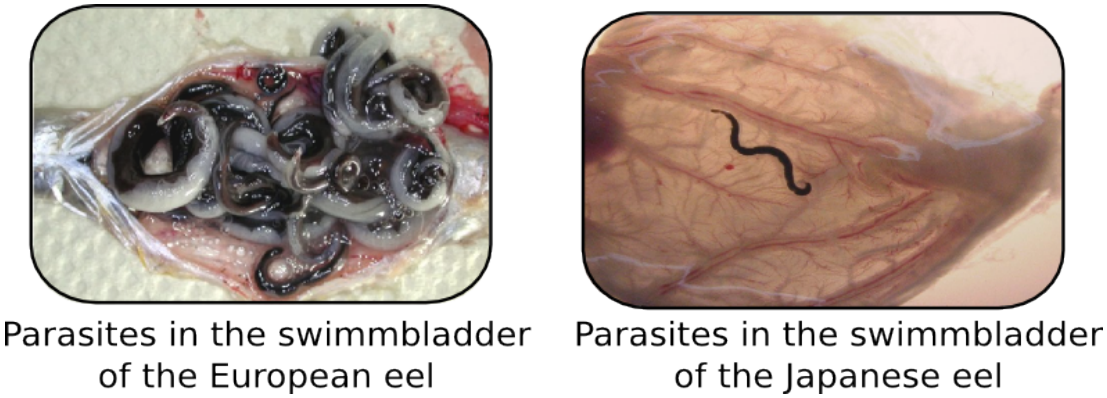

Figure 1.3: Difference between worms in the swimbladder of the European eel and the Japanese eel - Note the bigger size and higher number of worm in a typically infected European eel. In comparison in the Japanese eel worms are smaller and intensities of infection are much lower. The dark brown matter 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

With a view on the potential co-evolution (i.e. adaptation), of the eel-hosts to *A. crassus* the katadromous reproduction of freshwater eels might play an important role. Individuals of both species *An. anguilla* and *An. japonica* migrate thouthands of kilometers

to reproduce in the area of the Sargasso sea CITE!!. The Japanese eel in its endemic area migrates to XXX CITE!!!. Eel larvae then migrate to their freshwater habitates with the help of oceanic currents. While hybrids between the two Atlantic eel species have only been reported from Iceland !!CITE, European eels as a species 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 reduce the effectiveness of selection, when unifected populations are participating in reproduction.

A decline of epidemiological parameters for European populations of *A. crassus* has been hypothised based on data published over two decades (33). However this decline has not been confirmed in an explicit meta-analysis. If it would be present, possible expanations would include lower population density of the eel, an evolution of the eel host towards better resitance, and an evolution of *A. crassus* towards lower virulence.

!!! Fit here: Memory componetnt of the vertebrate immune system has been thought to be a driving positive selection on antigenes of microorganisms (45). The immune systems of teleost and of eels especially differs in many details from the mammalian immune system (i.e. it lacks all but the M-class of antibodies, response to macroparasites is carried out mainly by neutrophile rather than eosinophile granulocytes !!! CITE).

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

The genus *Anguillicola* comprises five morphospecies (46): In East Asia in additon to *A. crassus*, *A. globiceps* Yamaguti, 1935 (47) parasitises *Anguilla japonica*. *A. novaezelandiae* is endemic to New Zealand and South-Eastern Australia in *Anguilla australis* and *A. australiensis* Johnston et Mawson, 1940 (48) 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 CITE!!. 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 1.4) or as forming a clade with the oceanic species with *A. globiceps* and *A. papernai* in a sister clade (see figure 1.5).

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. Two of these host-capture events must have spanned the major splits in the eel phylogeny (49): Oceanic *Anguillicola* must have captured hosts transitioning between the clade of *An. reinhardtii* and *An. japonica* to the clade in which *An. australis* is found. Finally the the most basal have species of frehwater eels *An. mossambica* must have been captured.

The recent anthropogenic host-switch of *A. crassus* from *An. japonica* to *An. anguilla* and *An. rostrata* constitues additional acquisition of phylogenetically well separated hosts. This affinity for host-switching may be an evolutionary relict found only in one clade of *A. crassus* !!CITE Laetsch.

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* (50)), one of 5 major clades of nematodes (51, 52). 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). The finer subdivision of the Spirurina into Spirurina A, and the Sister clades Spriurnina B and C from Laetsch et al. can be seen in figure 1.6.

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. In addition to this diversity, a common characteristic of Spirurina A and C is a complex life-cycle involving freshwater or marine intermediate hosts. The observation of these complex traits render the assumption of evolution of the parasitic Spirurina from a free-living ancestor less parsimonous.

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.

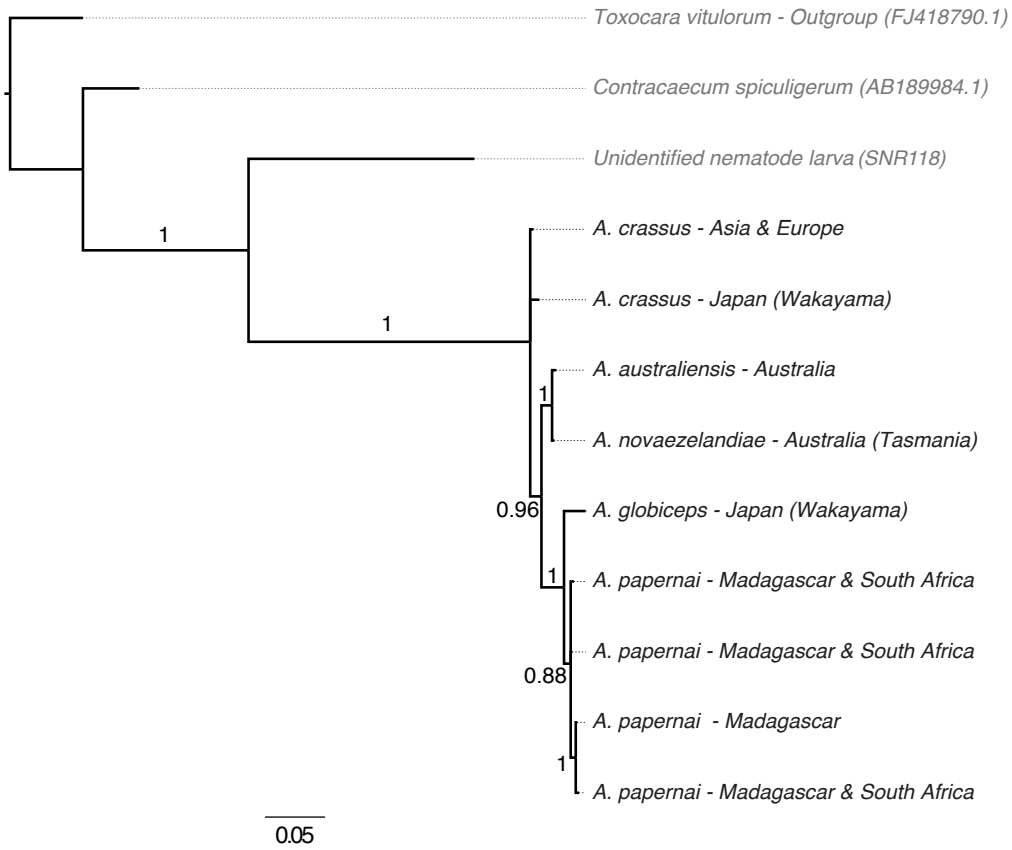

Figure 1.4: Phylogeny of the genus *Anguillicola* based on nLSU - Phylogram infered from large ribosomal subinit of *Anguillicola* and outgroups using Bayesian Inference. Lables on internal branches indicate Bayesian posterior probabilities. From Laetsch et al. CITE!!

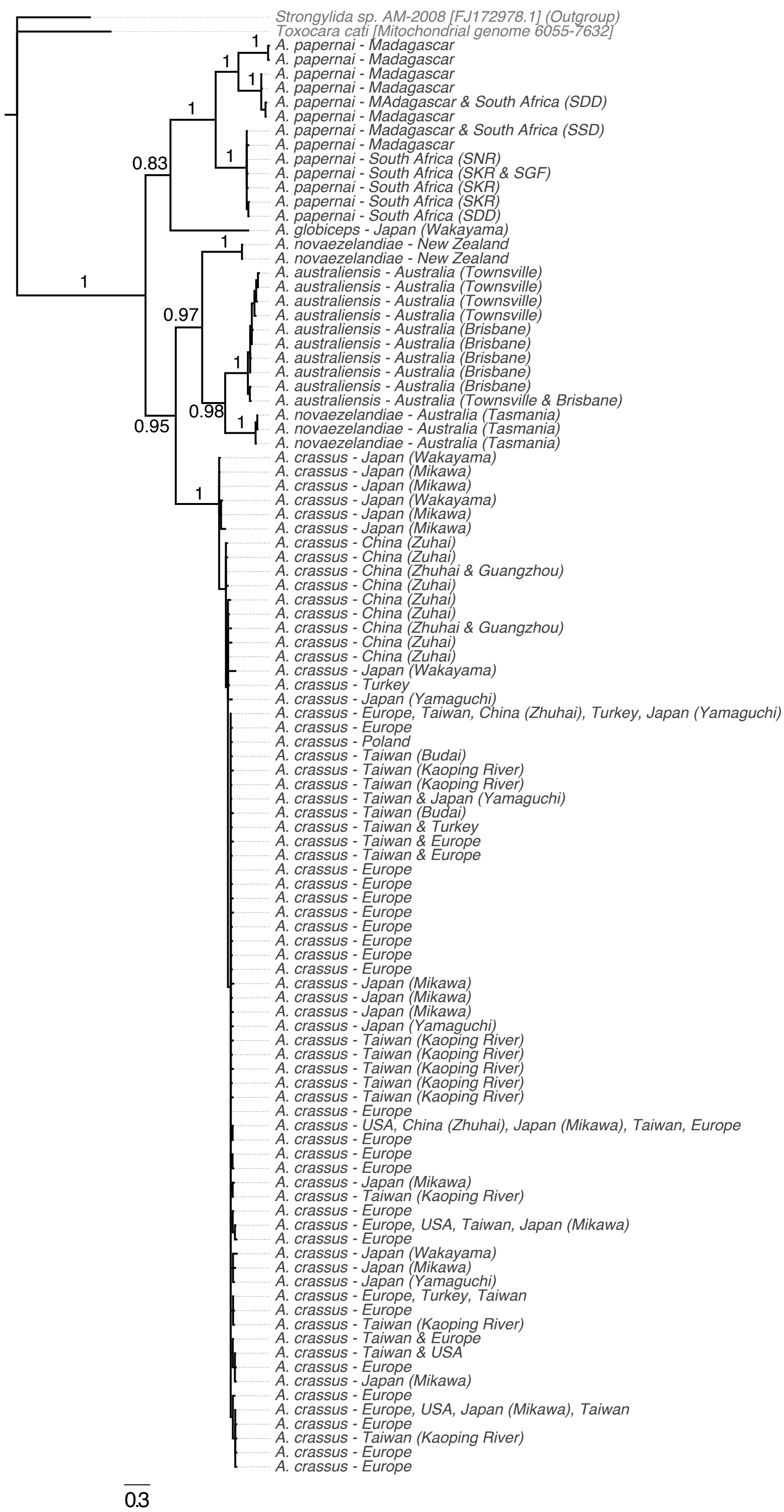

Figure 1.5: Phylogeny of the genus *Anguillicola* based on COXI - Phylogram infered for *Anguillicola* based on mitochondrial Cytochrome C oxidase subunit I and outgroups using Bayesian Inference. Lables on internal branches indicate Bayesian posterior probabilities. From Laetsch et al. CITE!!

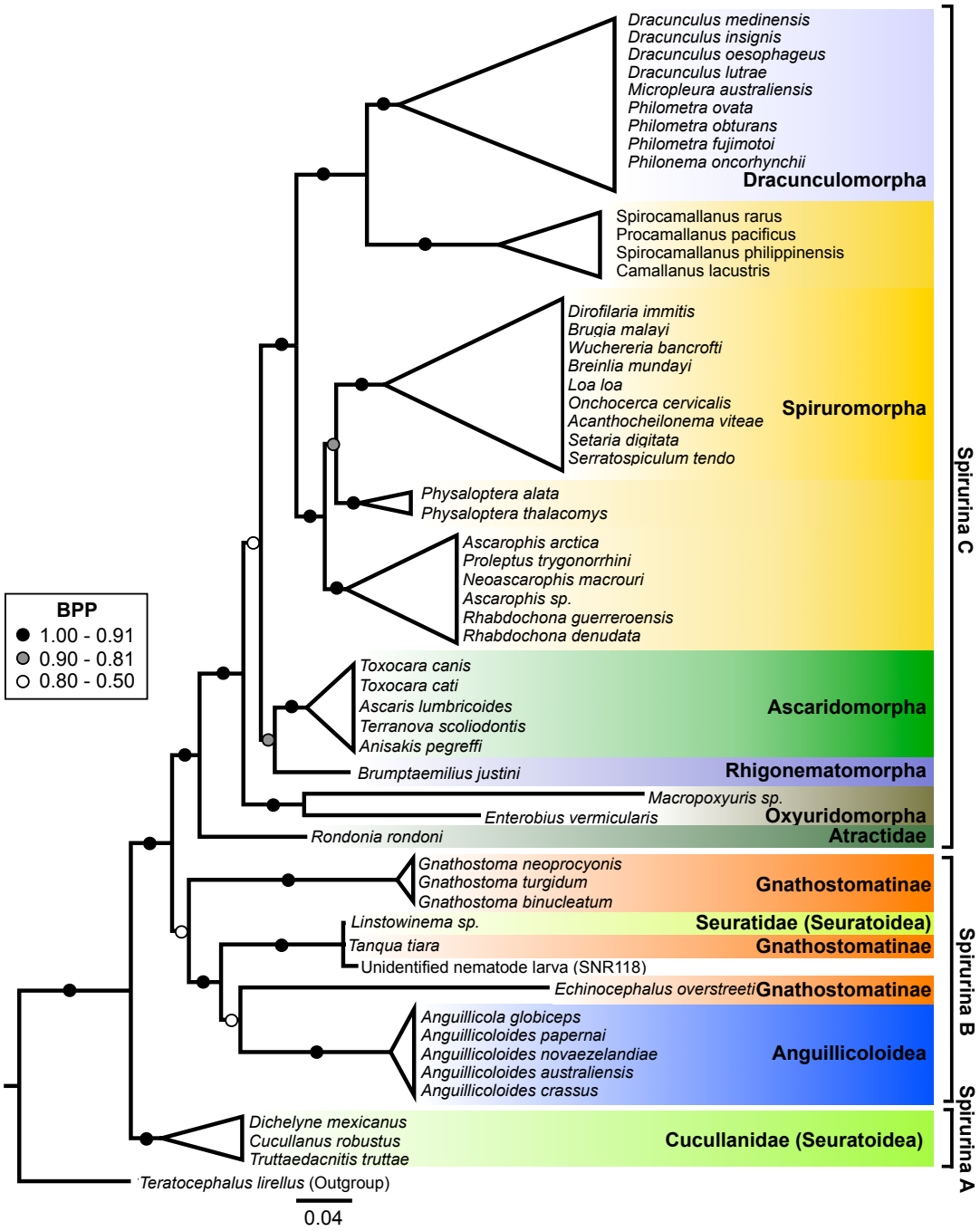

Figure 1.6: Phylogeny of nematode clade III based on nSSU - Phylogram inferred from nuclear small ribosomal subunit for Spirurina using Bayesian Inference. Branches are collapsed to highlight major groups. Lables on internal branches indicate Bayesian posterior probabilities. From Laetsch et al. CITE!!

1.1.2.2 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 (53, 54). 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 in an experimental setup under 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" (**?**). 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 from the Taiwanese population of worms was higher than from European worms. From the Taiwanese population less L4 larvae were observed at 25 dpi and the levels of this larval stage were stable during the infection, in contrast the numbers of L4 for the European populations decreased with the time. Additionally up to 50 dpi there were less living adults observerd for worm from the Taiwanese population and fewer dead adult worms were recorded for the Taiwanese population beginning from 50 dpi.

In the Japanese eel fewer L3 larvae at 25dpi were observed from the Taiwanese population compared to the European population of worms. Additioally more L4 larvae at this point in time and fewer living adults bat 25 and 150 dpi, as well as fewer adults beginning from 50 dpi from wroms of Taiwanese origin compared to worms of European origin.

These findings taken can be consolidated to the interpretation that an increase in the speed of development was observed in the European populations of *A. crassus* compared to the Taiwanese source population.

Interpretation of morphological characters in these studies proved difficult. Size of the worms seems to be mainly determined by the uptake of host-blood and is thus

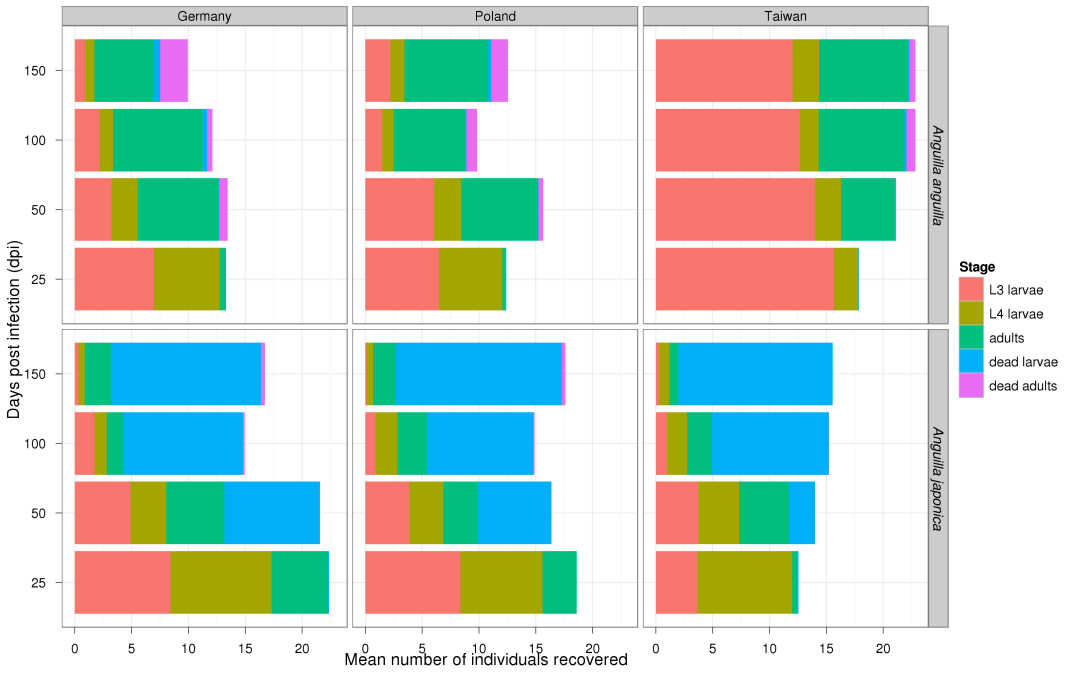

Figure 1.7: Differences in developmental speed - three populations of *A. crassus* (rows) were raised in two different hosts (column). Bars represent means of recovered individuals from three life-clycle stages. Differences are pointed out in the main text. Data courtesy of Urszula Weclawski.

largely object to phenotypic modification, with a genetic component hard to detect.

1.2 DNA sequencing

1.2.1 A very short history of DNA-sequencing

For almost tree decades the method developed by Frederick Sanger (55) was the only practical choice for determining the sequence of nucleic acid. The method uses labeled (first radioactive lables were used later fluorescent) chain termination nucleotides. If such a molecule is incorporated into a strand of DNA, synthesis stops and the lenth of the partial DNA-sequence can be determined on a single-base resolution agarose gel along with the corresponding base at that position. However cloning, Cloning!!! PCR!!! Although modern machines use the chain-termitaion method combined with capillary gel electorphoresis (56) in a highly paralized way, costs and labour constrained sequencing to a well established laboratory-model organisms. After phages !!!CITE and small bacteria !!!CITE in the first years of DNA sequencing. The bacterium *Escherichia coli* in 1997 (57), the baker's yeast *Saccharomyces cerevisiae* in 1996 (58), the nematode *Caenorhabditis elegans* in 1998 (59), the fruit fly *Drosophila melanogaster* in 2000 (60) and the mouse *Mus musculus* in 2002 (61) were the model orgaisms, for which multi-national consortia sequenced genomes in multi-million dollar projects. This "first generation of genomics" culminated in the publication of the human genome in 2001 (62).

1.2.2 Computational methods in DNA-sequence analysis

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 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 (63). Soon copmuter programs were necessary to align sequences, to compute overlaps and consensus sequences (64) and the process of computationally reconstructing the target molecule was termed sequence-assembly (65).

The first step in this overlap-consensus approach is to detect overlapping sequence in a series of pairwise alignments. Two classical approaches exist, the first being local "smith-waterman" alignment CITTE!!! the second "needleman-wunsch" global alignment.

These methods have usages outside of sequence assembly in general sequence comparison. After a seeding step, in which small regions of similarity are found, `blast` for example uses smith-waterman alignments to extend regions of similarity to form high-scoring segment pairs (HSPs). A `blast`-hit can contain multiple such such HSPs if sequence from one seed can't be extended to another.

Whith the advent of next generation sequencing (see 1.3) a new kind of alignment methods was needed to handle data volume, error structure and short read-length.

Ssaha2 XXX buffows-wheeler transform XXX graph alignment methods XXX.

1.2.3 DNA-sequencing in Nematodes

In 1998 *Caenorhabditis elegans* had become the first multicellular organism with a sequenced genome (59). Soon it was noted, that in addition to it's use as a general model system for the metazoa and bejond, knowledge gained in this species has the potential to be even more valuable in the phylum nematoda (66). The breadth and detail of genomic information available for *C. elegens* to date is illustrated by a recent publication of the Gerstein et al. (67) providing detailed annotation of the diverse functional genomic elements at single base resolution and their interactions.

The genome sequence of *Caenorhabditis elegans* was soon complemented by the gneome of *Caenorhabditis briggsae* (68), a second nematode from the genus *Caenorhabditis* sequenced a satellite system for comparative genomics instide this genus. As a second satellite model in clade V the necroXXXX *Pristionchus pacificus* have published draft genomes.

The first published genome of a parasitic nematode in the Spirurina was the draft genome of *Brugia malayi* (69). As a second gneome in the Spirurina recently the genome of *Ascaris suum* !!! CITE was publised.

Also in the remaining clades genome sequencingf folorished in the nematoda: For the animal-prasite *Trichinella spiralis* from clade I (70), the plant parasites *Meloidogyne*

incognita (71) and *Meloidogyne hapla* (72) as well as the the pinewood nematode *Bursaphelenchus xylophilus* (73) (a plant parasite using a beetle as an vector) from clade IV have recently genome sequences have been pulished.

The current revolution in sequencing methodology (see 1.3) brings into sight many more sequenced nematode genomes (including that of *A. crassus*). The 959 nematode genomes initiative promotes such sequencing of nematode genomes and makes workin-drafts of genome-assemblies available for analytical purposes in a `blast`-server CITE!!

Before the advent of NGS the lack of genomic information in many species of nematodes promoted the use of ESTs as a tool for gene-discovery. Partial genomes *sensu* (74) were successfully interrogated for a large array of genes interesting for various scientific communities. In nematode parasites of vertebrates, pathogenic factors were described as potential vaccine candidates (75).

Cystein-proteinase inhibitors (cystatins) and serin protenase inhibitors (serpins) are thought to interact with the antigen presentation in vertebrat hosts (75). Homologues of mammalian cytokines were identified, which are believed to interact with mammalian cytocine receptors to divert the immune response to a TH2-XXX-type response (76).

Further molecules involved in host-parasite interaction, which have been identified in the transcriptome-projects include abundant larval transcripts of *B. malay* (Bm-ALT) (77) and venom like allergens (Bm-VLA) CITE!!!

In some of these studies secreted proteins were in the center of interest. They could potentially be excreted by the nematode to allow movement and food-uptake bot also 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* (78), and across all nematode ESTs (79).

Proteomic analysis in *Brugia malayi* (80, 81), *Heligmosomoides polygyrus* (82) and *Haemonchus contortus* (83) was able to find evidence for excretion for some of the protein-products and to highlight additional candidate genes.

!!! FIT: 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* (84).

1.3 Advances in sequencing technology

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 (85)).

1.3.1 Pyro-sequencing

454- or pyro-sequencing uses emulsion PCR to apmlify single DNA molecules attached to beads. This covers each bead with multiple copies of one target molecule. These beads are then distributed over the 600,000 wells of a so called picolitre plate, a single bead in a well is covered with polymerase emitting light when phosporous is released during base-incorporation. Nucleotides are flushed over the plate one at a time and a high resolution camera records the emission of light, the intensity of emmited light is proprotional to the number of nucleotides incorporated. The ability to distinguish length of homopolymer runs in sequence decreases with the length of such homopolymer runs. Current "Titanium chemistry" is producing read of 400 bases length, "FLX chemistry" (used up to 2009) was able to produce reads of 250 bases length !!! CITE (also see ref results).

This longer read length of 454-sequencing (86) compared to other NGS technologies (see 1.3.2), 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 (87)).

1.3.2 Illumina-Solexa sequencing

The Illumina-Solexa platform uses bridge amplification to produce copies of single DNA molecules in clusters on a glass slide. These clusters are then sequenced usign a sequencing by synthesis technique: "Removable termitator" nucleatodes emitting a base specific fluorescence are flushed over the class slide transient incorproation is detected using a

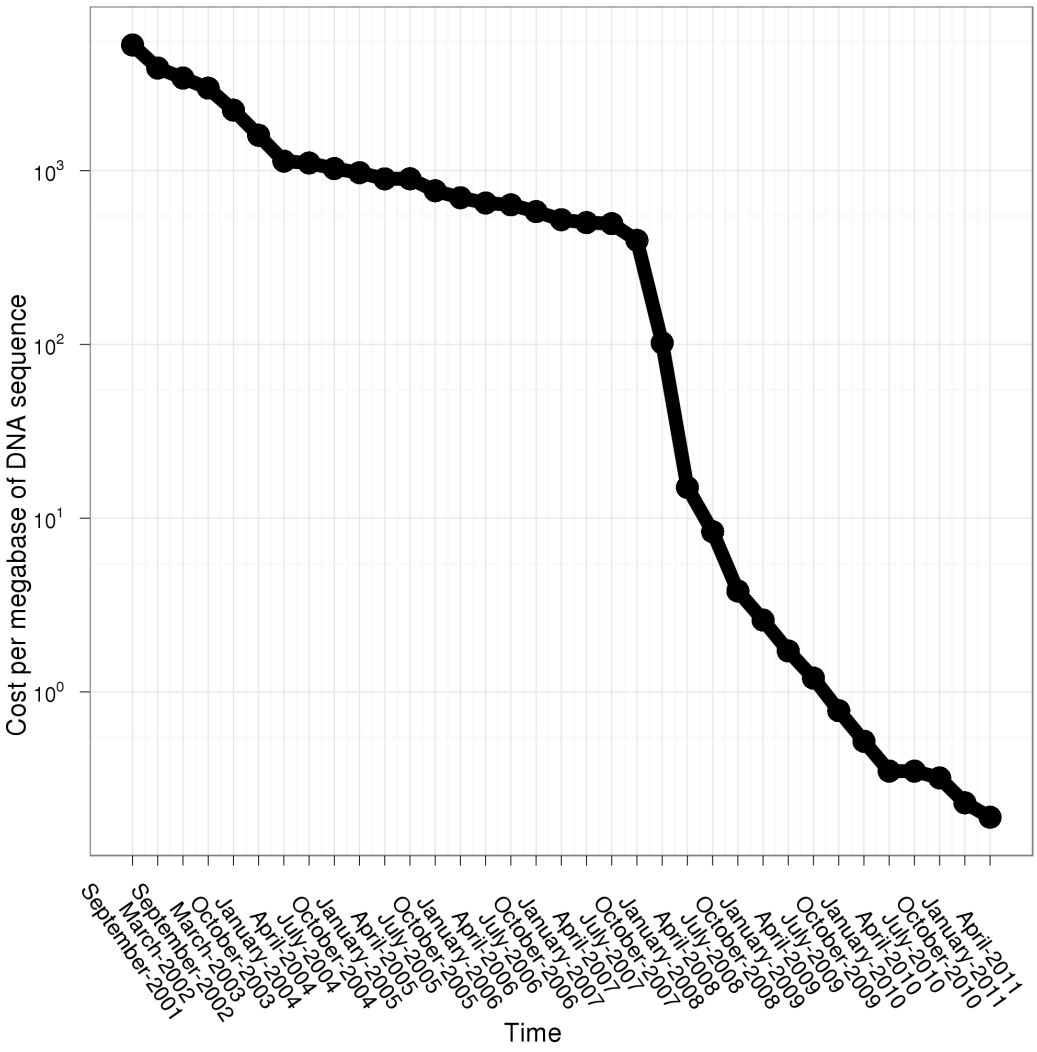

Figure 1.8: Falling seqeuncing costs - Sequencing costs falling due to advances in Solexa-sequencing: Due to improved read-length and data-volume on this plattform per base sequencing-prices for many applications thumble into free fall. Data provided by National Human Genome Research Institute, NHGRI.

high resolution camera. This leads to an error model different from 454 sequencing: Homoploymer runs are non-problematic, but due to the decreasing propensity of terminators for removal, sequencing quality decrases in from 5' to 3' direction.

Solexa illumina technology is to date (Dec. 2011) the most competitive sequencing platforms enabeling a broad spectrum of applications.

Recent increases in read length (from 35 bases in 2008 !!!CITE to over 100 bases in 2011 !!!CITE) are beginning to allow *de novo* sequencing of large genomes !!! CITE (panda) and transcriptomes !!!CITE. In the same periode throughput also increased from 6,000,000 reads in 2008 to 20.000.000 reads in 2011 per lane of the instrument.

The high throughput of the Illumina-Solexa platform makes it also first choice for gene expression analyis (88):

RNA-seq (89)

SuperSAGE (90) using expression-tags provides the benefit of classical SAGE-analysis (91) with those of the ulta hight throughput of Illumina-Solexa sequencing. normalisations

1.3.3 Applications of NGS in ecology and evolution

A study on trout in Lake Superior (92) used an approach similar to the appoach in the work presented here: Fish, which 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 successfully indentied 40 genes from two biochemical pathways being differently expressed. However, in addition to showing divergent evolution of gene-expression, this sutdy highlighted the limitations of 454 sequencing for gene-expression analysis.

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.

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 (93) .

In *Drosophila* variation of gene-expression within a single species can be attributed more to trans-regulatory elements, while expression divergent between species is dominated by cis-regulatory differences (94). Furthermore sterility of hybrid between species of this genus has been shown to result from incompatibilities in gene-regulatory networks (95).

2

Aims of the project

2.1 Preliminary aims

In order to investigate transcriptomic response to environmental stimuli, the responding unit, the transcripts have to be established first. As extremely short reads providing ultra high thoroughput are hard to assemble *de-novo*, a reference was created first using 454 pyrosequencing technology providing longer read-length.

2.2 Final aim

In a common garden environment

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.

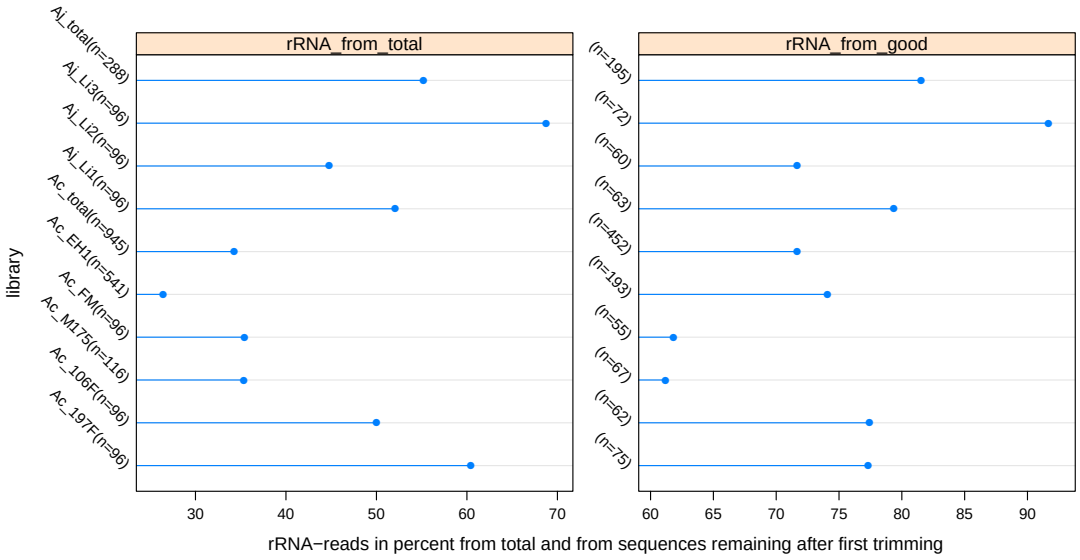

Figure 3.1: Proportion of rRNA in different libraries for *A. crassus* and *A. japonica* - rRNA abundance as proportion of the raw sequencing-reads (rRNA from total) and as proportion of the reads after quality screening (rRNA from good). Libraries starting with "Ac_" are from *A.crassus*, libraries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

	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.

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 attributed 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 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; for-

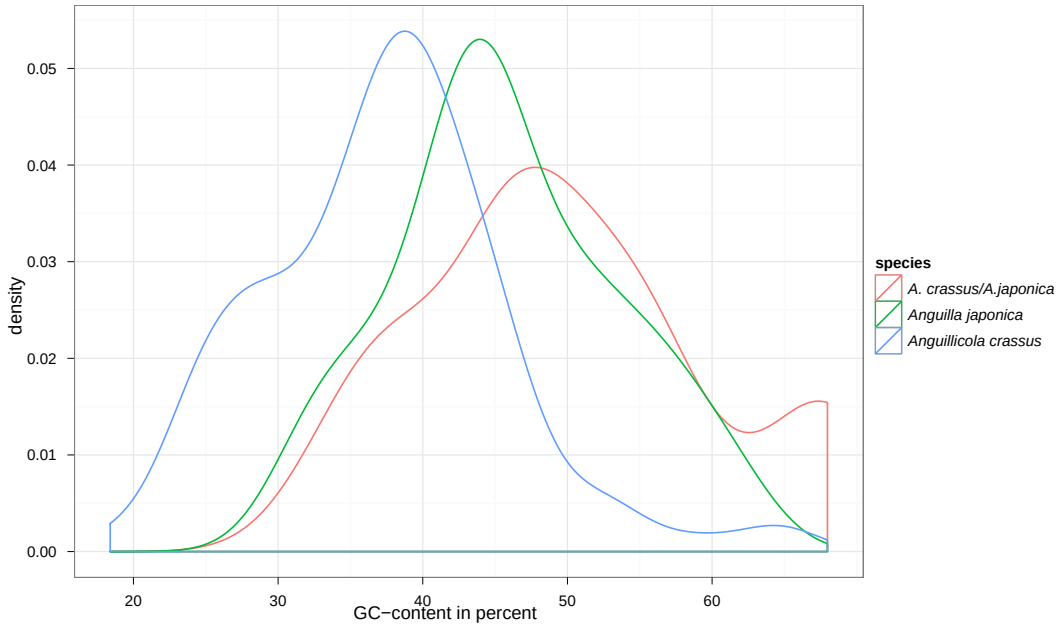

Figure 3.2: GC-content of sequences from *A. japonica* and *A. crassus* - The Japanese eel has a slighly higher GC-content than the parasite: This sequence characteristic is useful for seperation of sequences from the host-parasite interface, note the higer GC-content of the sequences from *A.crassus*,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

ward 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 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

We are providing transcriptome-data for the parasite *A. crassus*, enabling a broad spectrum of molecular research on this ecologically and economically important species.

We emphasize the importance of screening for xenobiotics. We consider this aspect important in any deep transcriptome project. First the depth of sequencing is leading to the generation of large amounts of off-target data from a "metatranscriptomic community" associated with a target organism. Second due to the abundance of laboratory contamination and the possibility of cross-contamination if libraries are sequenced only on a subset of a picotiterplate (i.e. without the use of barcodes distinguishing between samples (96)) non-biological contamination can be introduced. However, in the context of a parasite (or an infected host) the screening for off-target data and contamination becomes even more important: Correct inference of biological origin for a given contig constitutes a prerequisite for most downstream analysis or the interpretation of results.

Cross-contamination from different compartments of a picolitre-plate was ruled out by our sequence provider, using Multiplex Indexes (MID) for one library and similarity searches to neighboring lanes for the other libraries.

For the remaining off-target and contamination problem we archived separation of sequences

in two steps, one before assembly, one afterward. Both screening-steps had to rely solely on sequence comparison. The screening-step before assembly has to employ lower stringency as sequence comparisons on sequence as short as reads are less informative than on longer contig-sequence. In our case of *A. crassus*, neither of the two host species has genomic data available for use in similarity searches. A publicly available transcriptome-data-set for European eel (97) in addition to a unpublished data-set for the same species was augmented with a data-set generated from the Japanese eel sequenced for the purpose of screening *A. crassus*-sequences in the present project. The pre-assembly screening had the rationale of facilitating the assembly process reducing the amount of divergent sequence from two host-species and the amount of extensively covered rRNA sequence. In our sequencing we were not able to reproducibly alleviate the rRNA coverage. This has probably been due to the fact that extraction of total-RNA from worms filled with host blood resulted in low amounts of starting material, and amplification using standard kits did not allow to reproducibly alleviate rRNA abundance. As the same problems existed in preparation of liver tissue of the host species it seems likely that the blood of eels contains substances limiting the success of specific amplification protocols. In fact it is known that compounds like hemoglobin can inhibit PCR reactions (98) and reverse transcription (99).

Although raw reads with rRNA hits were screened out prior to assembly, it was still possible to gain insights from these off-target data, as we assembled and annotated screening databases. Some of the rRNA data especially from the L2 library showed high similarity to flagellate eukaryotes. It could be possibly derived from an unknown protist living in the swimbladder of eels (possibly as a commensal of *A. crassus*), from where the L2 larvae for RNA-preparation were washed out. This seems worth further investigation, especially as it has been controversial whether encapsulated objects in the swimbladder of eels could be attributed solely to *A. crassus* (38) or to opportunist coinfections.

We were able to demonstrate, that screening of SNPs in or adjacent to homopolymer regions "improved" overall measurements on SNP-quality:

First the ratio of transitions to transversions (ti/tv) increased. Such an increase is explainable by the removal of "noise" associated with common homopolymer-errors (100). Assuming that errors would be independent of transversion-transition bias erroneous SNPs would have a ti/tv of 0.5 and thereby lower the overall value. Other explanations for these observations are hard to find so it can be concluded that removing noise from homopolymer sequencing-error ti/tv increases. The value of XXX XXXX outside, XXX inside ORFs) is in good agreement with the overall ti/tv of humans (2.16) or *Drosophila* (2.07 (101)).

The ratio of non-synonymous SNPs per non-synonymous site to synonymous SNPs per synonymous site (dn/ds) decreased with removal of SNPs adjacent to homopolymer regions from XXX to XXXX after full screening. Similar to ti/tv it the most plausible explanation is the removal of error, as unbiased error would lead to a dn/ds of 1. While dn/ds is not unproblematic to interpret within populations (102), assuming negative (purifying) selection on most protein-coding genes lower values seem more plausible, also in comparison with other studies (see further text).

We used a threshold value for the minority allele of 7% for exclusion of SNPs, this corresponds to the ca. 10 "haploid equivalents" (5 individual worms plus an negligible amount of L2 larvae - in the L2 library and within the female adult worms - bearing possibly additional diversity). It is hard to explain, that ti/tv decreased in this filtering step, while dn/ds still further decreased.

The benefit of this screening was mainly a reduction of non-synonymous SNPs in high coverage contigs. When it was applied dn/ds did not scale with coverage. Working with an estimate of dn/ds independent of coverage, efforts to control for sampling a biased by sampling depth (i.e. coverage) like developed (103) and used (104) could be avoided.

7.3 Experimental infections

Such experiments have their problems because environmental factors, such as the general quality of the environment (i.e. water temperature) can interact with the host-environment (**?**).

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 for 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 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 (74)) was used, briefly it performs quality trimming using phred(105) and trimming of vector sequences using cross-match(106). 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.

8.5 454-pyro-sequencing

Nematode samples, RNA extraction, cDNA synthesis and Sequencing

A. crassus from *JAn. japonica* were sampled from Kao-Ping river and an adjacent aquaculture in Taiwan as described in (38). Worms from An. anguilla were sampled in Sniardwy Lake, Poland (53.751959N, 21.730957E) and from the Linkenheimer Altrhein, Germany (49.0262N, 8.310556E). After determination of the sex of adult nematodes, they were stored in RNA-later (Quiagen, Hilden, Germany) until extraction of RNA. RNA was extracted from individual adult male and female nematodes and from a population of L2 larvae (Table 1). RNA was reverse transcribed and amplified into cDNA using the MINT-cDNA synthesis kit (Evrogen, Moscow, Russia). For host contamination screening a liver-sample from an uninfected A. japonica was also processed. Emulsion PCR was performed for each cDNA library according to the manufacturerâĂŹs potocols (Roche/454 Life Sciences), and sequenced on a Roche 454 Genome Sequencer FLX. All samples were sequenced using the FLX Titanium chemistry, except for the taiwanese female sample T2, which was sequenced using FLX standard chemistry, to generate between 99,000 and 209,000 raw reads. For the L2 larval library, which had a larger number of non-A. crassus, non-An. anguilla reads, we confirmed that these data were not laboratory contaminants by screening Roche 454 data produced on the same run in independent sequencing lanes.

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 (106) (with parameters -minscore 20 and -minmatch 10). Seqclean (107) was used to identify and remove poly-A-tails, low quality, repetitive and short (<100 base) sequences. All reads were compared to a set of screening databases using BLAST (expect value cutoff E<1e-5, low complexity filtering turned off: -F F). The databases used were (a) a host sequence database comprising an assembly of the An. japonica Roche 454 data, an unpublished

assembly of An. anguilla Sanger dideoxy sequencesd expressed sequence tags (made available to us by Gordon Cramb, University of St Andrews) and transcripts from from EeelBase (97) a publically availble transcriptome database for the European eel; (b) a database of ribosomal RNA (rRNA) sequences from eel species derived from our Roche 454 data and EMBL-Bank; and (c) a database of rRNA sequences identified in our A. crassus data by comparing the reads to known nematode rRNAs from EMBL-Bank. This last database notably also contained xenobiont rRNA sequences. Reads with matches to one of these databases over more than 80% of their length and with greater than 95% identity were removed from the dataset. Screening and trimming information was written back into sff-format using sfffile (Roche 454). The filtered and trimmed data were assembled using the combined assembly approach (87), combining assemblies from the mira (108) and newbler (86). ****Give the details here and we will trim the text later **** . The two assemblies were combined into one using Cap3 (109) at default settings and contigs labeled by whether they derived from both assemblies or one assembly only.

Post-assembly classification and taxonomic assignment of contigs

After assembly contigs were assessed a second time for host and other contamination by comparing them (using BLAST) to the three databases defined above, and also to nembase4, a nematode transcriptome database derived from whole genome sequencing and EST assemblies (110, 111). For each contig, the highest-scoring match was recorded as long as it spanned more than 50% of the contig. We also compared the contigs to the NCBI non-redundant nucleotide (NCBI-nt) and protein (NCBI-nr) databases, recording the taxonomy of all best matches with expect values better than 1e-05.

Protein prediction and annotation

Protein translations were predicted from the contigs using prot4EST (version 3.0b) (112). Proteins were predicted either by joining single high scoring segment pairs (HSPs) from a BLAST search of uniref100 (113), or by ESTscan (114), using a training data the *Brugia malayi* complete proteome back-translated using a codon usage table derived from the BLAST HSPs, or, if the first two methods failed, simply the longest ORF in the contig. For contigs where the proein prediction required insertion or deletion of bases in the original sequence, we also imputed an edited sequence for each affected contig. Annotations with Gene Ontology (GO), Enzyme Commission (EC) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) terms were inferred for these proteins using Annot8r (version 1.1.1) (115), using the annotated sequences available in uniref100 (113). Up to 10 annotations based on a BLAST similarity bitscore cut-off of 55 were obtained for each annotation set. The complete *B. malayi* proteome (as present in uniref100) and the complete *C. elegans* proteome (as present in wormbase v.220) were also annotated in the same way. SignalP V4.0 (116) was used to predict signal peptide cleavage sites and signal anchor signatures.

Single nucleotide polymorphism analysis

We mapped the raw reads against the the complete set of contigs, replacing imputed sequences for originals where relevant, using ssaha2 (with parameters -kmer 13 -skip 3 -seeds 6 -score 100 -cmatch 10 -ckmer 6 -output sam -best 1). From the ssaha2 output, pileup-files were produced using samtools (117), discarding reads mapping to multiple regions. VarScan (118) (pileup2snp) was used with default parameters on pileup-files to output lists of single nucleotide polymorphisms (SNPs) and their locations.

Gene-expression analysis

For Roche 454 data, read counts for each transcript were obtained from the mapping to imputed sequence performed for SNP analyses. Tag-sequences were mapped using BWA (119). And read counts extracted using Samtools (117). For deepSAGE NlaIII-tag-sequencing, total RNA was prepared as described above from a female nematode from the Polish sampling site. A deepSAGE library was constructed following the protocol supplied by Illumina. Briefly after synthesis of cDNA on oligo(dT)-beads, cDNA was digested with the NlaIII (recognition site CATG), and the oligo(dT)-anchored 3' ends of mRNAs retained. After ligation of an adaptor containing an MmeI restriction site, the type II enzyme MmeI was used to cut 17 bases from the 3' end fragment, genrating a 21 base tag, expected to be unique for most mRNAs. The R-package DESeq (120) was used to normalize for library size and analyse statistical significance of differential expression of both Roche 454 and deepSAGE data. Spearman correlation coefficients were calculated for raw counts.

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, 6

[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 Františ ek 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. 3

[18] A. M. Barse and D. H. Secor. **An exotic nematode parasite of the American eel**. *Fisheries*, **24**(2):6–10, 1999. 3

[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. 3

[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 WÃijrtz 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, 6

[34] 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. 3

[35] 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. 5

[36] 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

[37] 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

[38] 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, 36, 42

[39] 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

[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. 6

[41] T. Wirth and L. Bernatchez. **Genetic evidence against panmixia in the European eel**. *Nature*, **409**:1037–1040, Feb 2001. 6

[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] DJ CONWAY and SD POLLEY. **Measuring immune selection**. *Parasitology(London. Print)*, **125**:3–16, 2002. 6

[46] 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. 6

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

[48] 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. 6

[49] 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. 7

[50] 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. 7

[51] 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. 7

[52] 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. 7

[53] A Kerner. **The natural history of plants, their forms, growth, reproduction, and distribution**. *Translated by F. W. Oliver.*, 1895. 11

[54] G Bonnier. **Recherches expÃŕrimentales sur lâĂŹadaptation des plants au climat alpin**. *Ann. Scie. Nat. (Bot.)*, **20**:217–358, 1895. 11

[55] 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

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

[57] 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

[58] 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

[59] The C. elegans Sequencing Consortium. **Genome sequence of the nematode C. elegans: a platform for investigating biology**. *Science*, **282**:2012–2018, Dec 1998. 13, 14

[60] 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

[61] 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

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

[63] 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]. 13

[64] R. Staden. **A strategy of DNA sequencing employing computer programs**. *Nucleic Acids Res.*, **6**:2601–2610, Jun 1979. 13

[65] T. R. Gingeras and R. J. Roberts. **Steps toward computer analysis of nucleotide sequences**. *Science*, **209**:1322–1328, Sep 1980. 13

[66] Mark Blaxter. **Caenorhabditis elegans Is a Nematode**. *Science*, **282**(5396):2041–2046, December 1998. 14

[67] 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. 14

[68] 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. 14

[69] 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. 14

[70] M. Mitreva, D. P. Jasmer, D. S. Zarlenga, Z. Wang, S. Abubucker, J. Martin, C. M. Taylor, Y. Yin, L. Fulton, 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. 14

[71] 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. 15

[72] 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. 15

[73] 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. 15

[74] John Parkinson, Alasdair Anthony, James Wasmuth, Ralf Schmid, Ann Hedley, and Mark Blaxter. **PartiGene–constructing partial genomes**. *Bioinformatics*, **20**(9):1398–1404, June 2004. 15, 41

[75] 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. 15

[76] Rick M. Maizels, Adam Balic, Natalia Gomez-Escobar, Meera Nair, Matt D. Taylor, and Judith E. Allen. **Helminth parasites; masters of regulation**. *Immunological Reviews*, **201**(1):89–116, 2004. 15

[77] 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. 15

[78] 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. 15

[79] 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. 15

[80] 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. 15

[81] 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. 15

REFERENCES

[82] 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. 15

[83] 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. 15

[84] M. K. Hughes and A. L. Hughes. **Natural selection on Plasmodium surface proteins**. *Mol. Biochem. Parasitol.*, **71**:99–113, Apr 1995. 15

[85] J. R. Miller, S. Koren, and G. Sutton. **Assembly algorithms for next-generation sequencing data**. *Genomics*, **95**:315–327, Jun 2010. 16

[86] 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. 16, 43

[87] S. Kumar and M. L. Blaxter. **Comparing de novo assemblers for 454 transcriptome data**. *BMC Genomics*, **11**:571, Oct 2010. 16, 43

[88] J. H. Malone and B. Oliver. **Microarrays, deep sequencing and the true measure of the transcriptome**. *BMC Biol.*, **9**:34, 2011. 18

[89] Z. Wang, M. Gerstein, and M. Snyder. **RNA-Seq: a revolutionary tool for transcriptomics**. *Nat. Rev. Genet.*, **10**:57–63, Jan 2009. 18

[90] 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. 18

[91] V. E. Velculescu, L. Zhang, B. Vogelstein, and K. W. Kinzler. **Serial analysis of gene expression**. *Science*, **270**:484–487, Oct 1995. 18

[92] 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. 18

[93] S. Via. **The Ecological Genetics of Speciation**. *The American Naturalist*, **159**(S3):1–7, 2002. 19

[94] 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. 19

[95] 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. 19

[96] N. J. Lennon, R. E. Lintner, S. Anderson, P. Alvarez, A. Barry, W. Brockman, R. Daza, R. L. Erlich, G. Giannoukos, L. Green, A. Hollinger, C. A. Hoover, D. B. Jaffe, F. Juhn, D. McCarthy, D. Perrin, K. Ponchner, T. L. Powers, K. Rizzolo, D. Robbins, E. Ryan, C. Russ, T. Sparrow, J. Stalker, S. Steelman, M. Weiand, A. Zimmer, M. R. Henn, C. Nusbaum, and R. Nicol. **A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454**. *Genome Biol.*, **11**:R15, 2010. 35

[97] 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, 43

[98] I. G. Wilson. **Inhibition and facilitation of nucleic acid amplification**. *Appl. Environ. Microbiol.*, **63**:3741–3751, Oct 1997. 36

[99] M. A. Valasek and J. J. Repa. **The power of real-time PCR**. *Adv Physiol Educ*, **29**:151–159, Sep 2005. 36

[100] S. Balzer, K. Malde, and I. Jonassen. **Systematic exploration of error sources in pyrosequencing flowgram data**. *Bioinformatics*, **27**:i304–309, Jul 2011. 36

[101] Andrew Adey, Hilary Morrison, X Asan, Xu Xun, Jacob Kitzman, Emily Turner, Bethany Stackhouse, Alexandra MacKenzie, Nicholas Caruccio, Xiuqing Zhang, Jacob Shendure, Emily Turner, Bethany Stackhouse, Alexandra MacKenzie, Nicholas Caruccio, Xiuqing Zhang, and Jay Shendure. **Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition**. *Genome Biol.*, **11**(12):R119, 2010. 36

[102] S. Kryazhimskiy and J. B. Plotkin. **The population genetics of dN/dS**. *PLoS Genet.*, **4**:e1000304, Dec 2008. 36

[103] E. Novaes, D. R. Drost, W. G. Farmerie, G. J. Pappas, D. Grattapaglia, R. R. Sederoff, and M. Kirst. **High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome**. *BMC Genomics*, **9**:312, 2008. 37

[104] S. T. O'Neil, J. D. Dzurisin, R. D. Carmichael, N. F. Lobo, S. J. Emrich, and J. J. Hellmann. **Population-level transcriptome sequencing of nonmodel organisms Erynnis propertius and Papilio zelicaon**. *BMC Genomics*, **11**:310, 2010. 37

[105] 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. 41

[106] Phil Green. *PHRAP documentation.*, 1994. 41, 42

[107] 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. 42

[108] 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. 43

[109] X. Huang and A. Madan. **CAP3: A DNA sequence assembly program**. *Genome Res.*, **9**:868–877, Sep 1999. 43

[110] 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. 43

[111] B. Elsworth, J. Wasmuth, and M. Blaxter. **NEMBASE4: The nematode transcriptome resource**. *Int. J. Parasitol.*, **41**:881–894, Jul 2011. 43

[112] James Wasmuth and Mark Blaxter. **prot4EST: Translating Expressed Sequence Tags from neglected genomes**. *BMC Bioinformatics*, **5**(1):187, 2004. 43

[113] 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. Fouger, 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. 43

[114] 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. 43

[115] Ralf Schmid and Mark L Blaxter. **annot8r: GO, EC and KEGG annotation of EST datasets**. *BMC Bioinformatics*, **9**:180, 2008. 43

[116] T. N. Petersen, S. Brunak, G. von Heijne, and H. Nielsen. **SignalP 4.0: discriminating signal peptides from transmembrane regions**. *Nat. Methods*, **8**:785–786, 2011. 43

[117] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, GonÃ¢alo R. Abecasis, and Richard Durbin. **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics*, **25**(16):2078–2079, 2009. 44

[118] 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. 44

[119] H. Li and R. Durbin. **Fast and accurate short read alignment with Burrows-Wheeler transform**. *Bioinformatics*, **25**:1754–1760, Jul 2009. 44

[120] S. Anders and W. Huber. **Differential expression analysis for sequence count data**. *Genome Biol.*, **11**:R106, 2010. 44

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.

Chapter 4 was in similar form submitted for publication to BMC Genomics, in the course of manuscript preparation Mark Blaxter edited the text.

Chapter 6 is in similar intended for publication in Plos biology, Mark Blaxter edited parts of the text.

The thesis work was conducted from May 2008 to December 2011 under the supervision of Prof. Dr. Horst Taraschewski at the Karlsruhe Institute of Technology and Prof. Mark Blaxter at the University of Edinburgh.

KARLSRUHE,

Emanuel G. Heitlinger

Redtenbacherstr. 9
76133 Karlsruhe
Germany

Phone: (+49) 721 292-5588
Email: emanuelheitligner@gmail.com

Born September, 12th in Schwäbisch Gmünd, Germany

Education

2008-2012 Doctoral studies, Karlsruhe Institute of Technology

Dissertation: Divergence of an intoduced parasite: a transcriptomic perspective on *Anguillicola crassus*

Supervisors: Prof. Dr. Horst Taraschewski and Prof. Mark Blaxter.

2007-2008 Work on diploma thesis, University of Karlsruhe, Zoological Institute, Department for Parasitology and Ecology

Thesis title: Vergleichende licht- und elektronenmikroskopische Untersuchungen am Intestinaltrakt des invasiven Schwimmblasennematoden *Anguillicola crassus* aus verschiedenen Aalarten.

2001-2007 Undergraduate studies in Biology, University of Karlsruhe.

Main subject: Zoology

Subsidiary subjects: Genetics, Botany

1991-2000 Secondary school, Privat-Gymnasium St.Paulusheim Bruchsal.

June 2000 Abitur (general qualification for university entrance).

1987-1991 Pirmary school, Kraichtal Oberöwisheim.

Employment

2008-2011 Research assistant, Karlsruhe Institute of Technology, Zoological Institute, Department for Parasitology and Ecology.

2000-2001 Alternative military service, youth centre Bruchsal.

Fields of Research Interest

Ecology and evolution of host-parastite interactions, Transcriptomics, Genomics

Research

Peer Reviewed Publications

Emanuel G Heitlinger, Dominik R Laetsch, Urszula Weclawski, Yu-San and Horst Taraschewski (2009) Massive encapsulation of larval *Anguillicoloides crassus* in the intestinal wall of Japanese eels. *Parasites & Vectors*, 2:48.

Dominik R Laetsch, Emanuel G Heitlinger, Horst Taraschewski, Steven A Nadler and Mark L Blaxter (2012) The phylogenetics of Anguillicolidae (Nematoda: Anguillicolidea), swimbladder parasites of eels. *under review BMC Evolutionary Biology.*

Conference Presentations

3rd Status Symposium, Volkswagen Foundation Funding Initiative Evolutionary Biology, November 7-11 2011, Sylt, Germany, Oral presentation: "Divergence of an intoduced parasite: a transcriptomic perspective on *Anguillicola crassus*".

2nd Status Symposium, Volkswagen Foundation Funding Initiative Evolutionary Biology, May 9-12 2010, Frauenchiemsee, Germany, Oral presentation: "The transcriptome of *Anguillicoloides crassus* sampled by pyrosequencing".

24th Biannual conference of the German society of parasitology (DGP), March 16-19 2010, Münster, Germany. 2 oral presentations: "The transcriptome of *Anguillicoloides crassus* sampled by pyrosequencing" and "Massive encapsulation of larval *Anguillicoloides crassus* in the intestinal wall of the Japanese eel".

Mind the gap: joining empirical and theoretical population genetics, October 2-3 2009, Freiburg, Germany. Oral Presentation: "Divergence between European and Asian populations of the swimbladder nematode *Anguillicoloides crassus*".

1st Status Symposium, Volkswagen Foundation Funding Initiative Evolutionary Biology, February 25-27 2009, Münster, Germany. Poster: "Divergence between East Asian and European populations of the swimbladder-nematode *Anguillicola crassus*".

Xth European Multicolloquium of Parasitology - EMOP 10, August 24-28, 2008, Paris, France. Oral Presentation: "Divergence between East Asian and European populations of the swimbladder-nematode *Anguillicola crassus*".

Honors, Awards, & Fellowships

2008 Volkswagen Stiftung PhD Fellowship, Funding Initiative Evolutionary Biology, full funding of research position and research material

Last updated: November 17, 2011