Drug Targets of the Heartworm, *Dirofilaria immitis*

Inauguraldissertation

zur
Erlangung des Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von

Christelle Godel

aus La Sagne (NE) und Domdidier (FR)

Schweiz

Avenches, 2012

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Jürg Utzinger
Prof. Dr. Pascal Mäser
P.D. Dr. Ronald Kaminsky
Prof. Dr. Georg von Samson-Himmelstjerna

Basel, den 26th of June 2012

Prof. Dr. M. Spiess

Dekan

To my husband and my daughter
With all my love.

Table of Content

Acknowledgements 5

Summary 8

Introduction 11

Dirofilaria immitis 12

Phylogeny and morphology 12

Repartition and ecology 14

Life cycle 16

Diagnosis, treatment and resistance 18

Old and new sequencing strategies 21

DNA sequencing 21

Premises - Sanger, Maxam and Gilbert 21

Dye-labeling terminator sequencing 23

Automated sequencing and high-throughput sequencing development 23

454-Pyrosequencing 24

Illumina sequencing 25

Third generation of high-throughput sequencing 27

Summary and further analysis 27

Assembly methods 29

Example of de novo genome sequencing 30

Anthelmintics and targets 31

Nicotinic acetylcholine receptors and cys-loop ligand-gated ion channels 33

Aims and objectives 36

Chapter 1. Understanding the magic bullet: molecular opportunities for antiparasitic drug selectivity 38

Chapter 2. Efficacy testing of AADs against *D. immitis* (unpublished) 61

Chapter 3. The genome of the heartworm, *Dirofilaria immitis* 66

Genome-wide survey for ligand gated ion channels in *D. immitis* 104

Chapter 4. Loss of DEG-3-subfamily acetylcholine receptors and lack of stereoselective monepantel sensitivity in *Dirofilaria immitis* 106

Deeper investigations through the DEG-3-subfamily....117
In silico novel potential drug targets of *D. immitis*....118
Chapter 5. Drug sensitivity tests (unpublished)....121
RNA-dependant RNA polymerase....122
Chitin synthase....123
Potential novel receptors of *Wolbachia*, endosymbiont of *D. immitis*....124
Discussion....126
1. Nucleic acid synthesis and repair receptors....127
RNA-dependant RNA polymerase....127
2. Glycosylation and sugar metabolism....128
Beta-1,4-mannosyltransferase....128
UDP-galactopyranose mutase....130
Chitin synthase....131
3. Parallel study performed on *Wolbachia* of *D. immitis*....132
Conclusions....134
References....136
Appendices....146
List of abbreviations....147
List of figures....152
List of tables....153
Curriculum vitae....155

Acknowledgements

I would like to thank all the people who contributed to the success of this doctoral thesis:

Pascal Mäser for his precious support and supervision throughout my PhD duration; for his help in innovation and for bringing relevant new input. And I would like to specially thank him for being present to answer important questions and to bring judicious advices.

Christian Epe for the internal support and supervision within Novartis Animal Health and good advices, trust and friendship. The CAP, companion animal parasiticides, team for their help in laboratory work and parasites maintenance.

Ronald Kaminsky to lead the project in house and help me with the internal questions, to guide me through the meanders of Novartis guidelines; to believe in me through my entire PhD and in my results.

Georg von Samson-Himmelstjerna for his invaluable advice and comments as my co-referee.

Jacques Bouvier for his support and help concerning screening tests and precious advices and information about chemical compounds as well as for his valuable new scientific input and ideas; for our open-minded interesting discussions.

Lucien Rufener for his precious advices in molecular biology; especially to help to come across difficulties in the laboratory work; for brainstorming situations and interesting scientific discussions as well as for friendship and sharing the same office.

Sandra Schorderet-Weber for her help and precious advices concerning the in vivo tests with the rodents. The Rodents Model team for the laboratory work and the parasites maintenance.

Daniel Nilsson from the Karolinska Institutet in Sweden for his great help and knowledge in informatics, computer science and bioinformatic studies.

Sujai Kumar, **Georgios Koutsovoulos** and **Mark Blaxter** from the University of Edinburgh for their help to share their knowledge and success in bioinformatics as well as their huge education concerning nematode parasites to set a good publication.

Frederik Bringaud to share his educative information concerning transposons and apply it to the heartworm.

Philipp Ludin and **Christoph Schmid** from the Swiss Tropical and Public Health Institute for their collaboration and work in the genome publication.

Novartis Animal Health St-Aubin FR and the **Swiss Tropical and Public Health Institute** to allow me to perform my PhD and share the results and interesting data with relevant persons and for the fellowship grant and the internal support.

Special thanks go to **Timothy Geary** and **Roger Prichard** for their adequate and quick advices as well as for their interesting scientific opinions and for their support to my studies.

Finally, I would like to thank my husband, **Sylvain**, and my daughter, **Mélie**, who supported me during the whole PhD and were always behind me as well as my family and my friends.

Summary

The filarial parasite, *Dirofilaria immitis,* infects dogs and canids in warm and tropical areas of the globe. Located in the right pulmonary arteries, the heartworm infection is a severe and potentially fatal disease; caused mainly by the adult worm stages. The zoonotic roundworm is transmitted by various mosquitoes, *Culicidae*.

Currently, control measures are based on monthly prophylaxis with macrocyclic lactones (MLs). The only adulticide treatment is melarsomine dihydrochloride injected intramuscularly which can cause complications. Resistance is emerging against the MLs, increasing the necessity to find novel anthelmintics.

The recently discovered class of anthelmintics, amino acetonitriles derivatives (AADs), with a new mode of action brought hope in the battle against helminth parasites. The aim of this thesis was to find candidate targets of AADs and other potential anthelmintics in *D. immitis* genome that could be used as therapeutic targets.

The first chapter discusses the possible mechanisms of antiparasitic drug action and selectivity.

The second chapter tests and confirms the efficacy of AADs against *D. immitis* in vitro. This suggests the presence of receptor(s) to AADs in the heartworm, a hypothesis which can be tested by genome sequencing.

The third chapter narrates the story of the genome project of the heartworm; de novo sequenced with Illumina HiSeq and assembled with Abyss and Velvet. This did not indicate that there is an AAD receptor (DEG-3-subfamily), to *Caenorhabditis elegans* and *Haemonchus contortus* in *D. immitis* genome. In fact, *D. immitis* does not appear to possess a single gene of the DEG-3 subfamily of acetylcholine receptors, the targets of monepantel. However, in-depth studies of the genome through exclusion criteria established a list of novel potential drug targets.

The fourth chapter further investigates the presumed loss of the DEG-3-subfamily in *D. immitis* and relates this finding to the sensitivity against both monepantel enantiomers in vitro and in vivo.

The fifth chapter picks up some of the newly identified candidate drug targets. It tests the predicted targets against the heartworm with known inhibitors.

This thesis presents a genomic approach to identify potential targets and new anthelmintics. I hope that, the genomic approach will support the development of drugs against the heartworm and related parasites.

Introduction

Dirofilaria immitis

Phylogeny and morphology

Nematodes are Ecdysozoans (as are Arthropods). This grouping is supported by morphological characters and includes all animals that shed their exoskeleton. Their major characteristic is the three-layered cuticle composed of organic material, which is periodically molted as the animal grows. This process is typically known as ecdysis. Furthermore, the nematodes are characterized by a complete open digestive system with mouth, intestine and anus (1).

The order Spirurida is characterized by ventral and caudoventral papillae and possesses an esophagus divided into muscular and posterior glandular parts. The nematodes of this order have intermediate hosts as part of their lifecycle. The superfamily Filarioidea, representing the filarial nematodes, is defined by intermediate hosts that are biting flies such as mosquitoes, blackflies or others. The infection of the definitive host occurs by inoculation of L_3 from these flies while feeding (2).

Filarial nematodes are members of the order Spirurida and can be assigned to a single family, the Onchocercidae, divided into two subfamilies, the Dirofilarinae and the Onchocercinae. The genus *Dirofilaria* is divided in two subgenera, *Nochtiella* and *Dirofilaria*. In the *Nochtiella* subgenus, most species are tissue-dwelling while in the *Dirofilaria* subgenus; *Dirofilaria immitis* and other species inhabit heart cavities, subcutaneous tissues and abdominal cavities (3; 4).

Nematodes of the genus *Dirofilaria* are elongated and thin with round anterior extremity and rudimentary buccal capsule without lips and small cephalic papillae (Figure 1). The esophagus is differentiated into muscular and glandular regions with no distinction between them. The caudal extremity of the female is round and the vulva aperture is located just behind the union of esophagus and intestine while the caudal extremity of the male is rolled in a spiral and more conical. Caudal alae are present as well as two spicules of different lengths. Generally, *D. immitis* nematodes are long without striation on the cuticle and the female caudal papillae are slightly asymmetrical while the male greater spicule is not clearly pointed. The other species of *Dirofilaria* are smaller with striation on the cuticle and asymmetrical caudal alae, while the male greater spicule is clearly pointed (5).

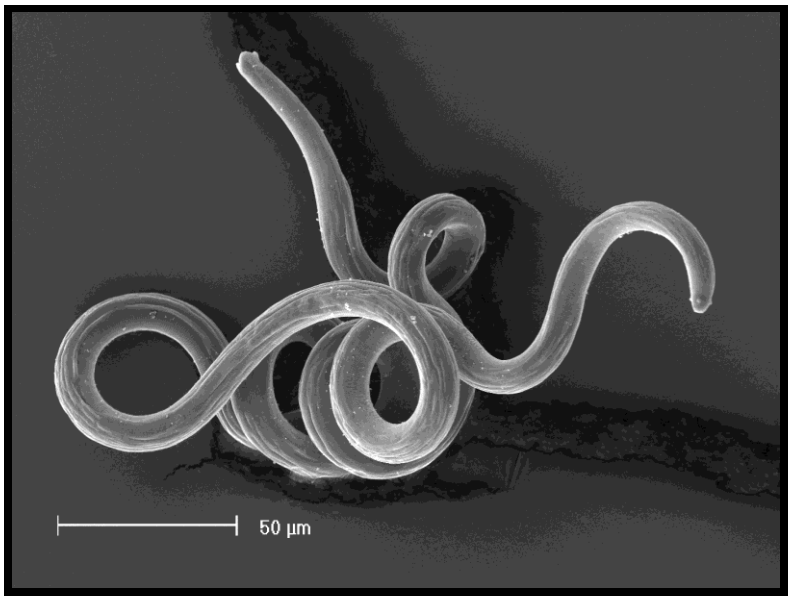


Figure 1. Third stage larvae of *D. immitis*. SEM picture showing the worm cuticle details.[1]

Figure 2 represents the phylogeny of nematodes based on 18S ribosomal RNA studies. It clearly shows the close relation between *D. immitis* and the *Onchocerca* genus (6; 7). The genus *Dirofilaria* is cosmopolitan in distribution; the species *D. immitis* is probably the most widespread species because of its association with domestic animals (8).

[1] C. Perret (Godel) Master Thesis (2008)

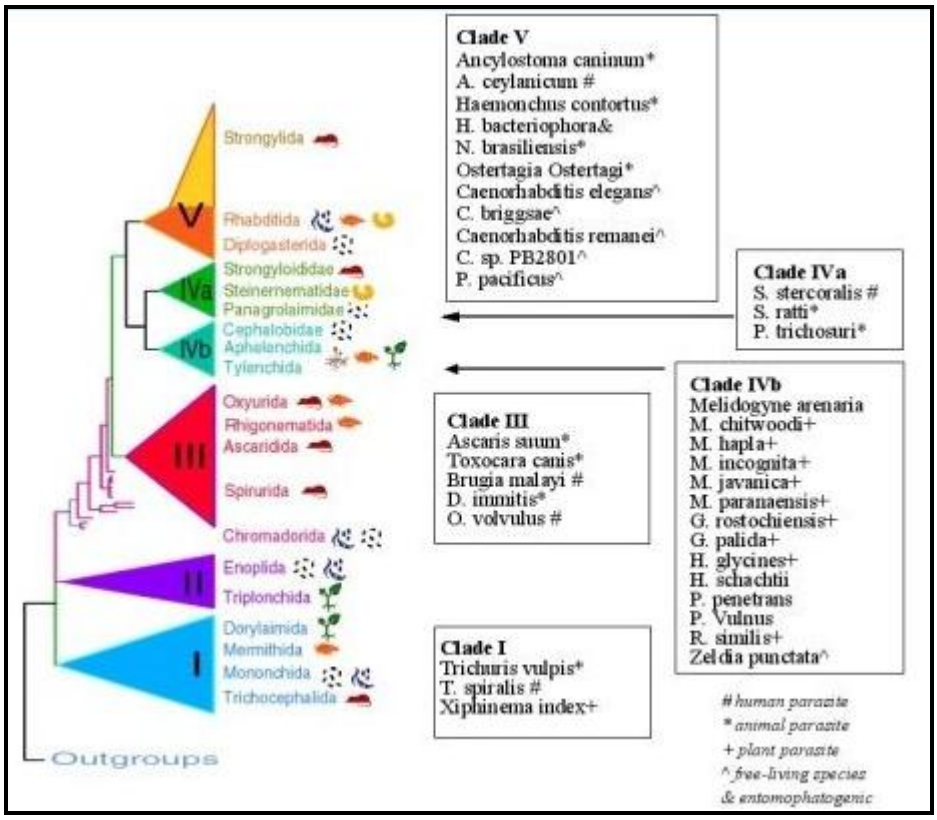


Figure 2. Phylogeny of nematodes based on 18S ribosomal RNA (1).[2]

Repartition and ecology

Dirofilaria immitis is commonly known as the heartworm. It is found in over 30 mammalian species including foxes, cats, wolves, coyotes and other wild carnivores, but mainly in dogs (9). Dogs are the definitive host and serve as the main reservoir of infection, with the heaviest worm burden. The parasite is named heartworm due to the location of the adults in the arteries of the lungs and occasionally in the right ventricle of the heart. Very common in numerous warm countries and particularly in tropical areas, it is widespread throughout the Far West, Equatorial Africa and in the Pacific and also occurs in North (Figure 3) and South America, Australia and North Africa (8), and South and East Europe (Figure 4).

[2] Blaxter et al. (1998) A molecular evolutionary framework for the phylum Nematoda *Nature* 392; 71-75

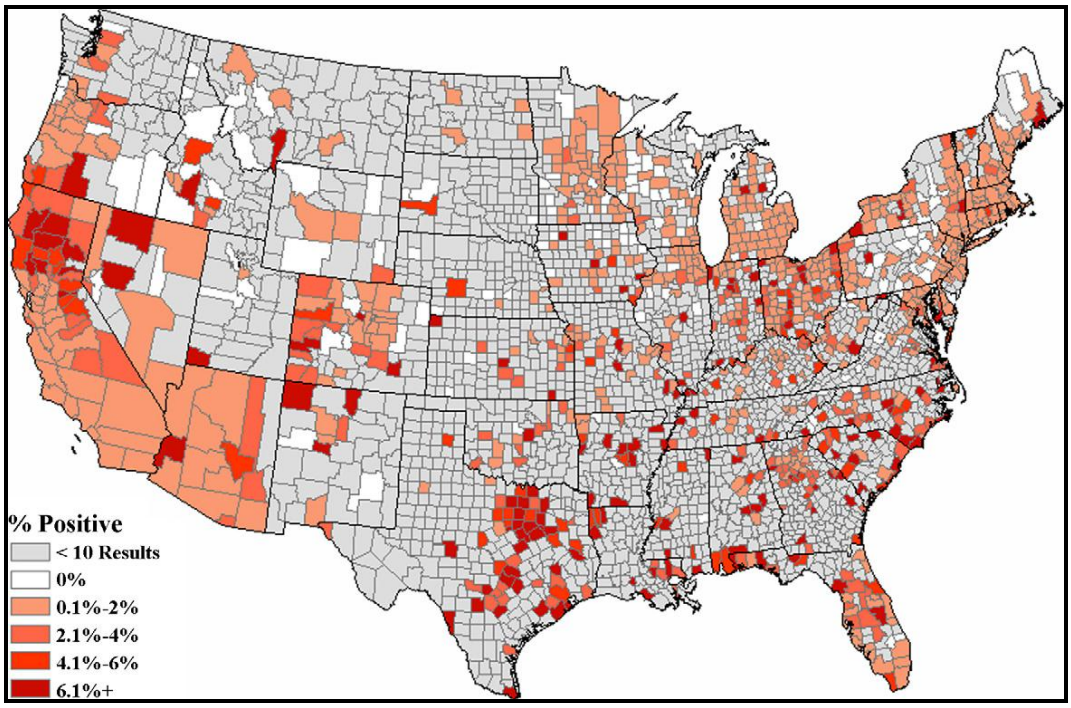


Figure 3. US prevalence of antigen positive tests to *D. immitis* in dogs per county. Counties in grey did not reported results. In white, no dog reported as positive (0%). Remaining counties were coded as follows: 0.1–2.0% (taupe), 2.1–4.0% (salmon), 4.1–6.0% (red), 6.1–20.0% (brick red).[3]

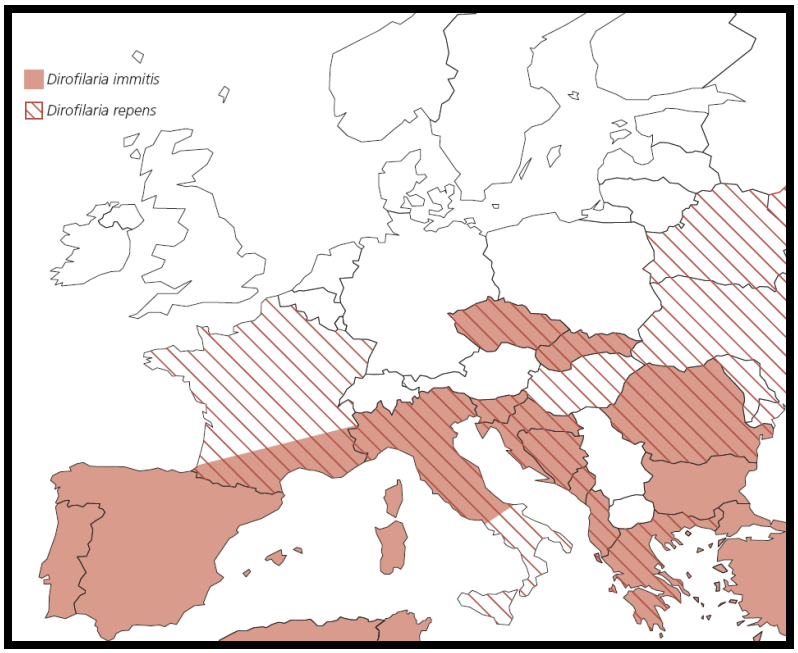


Figure 4. Distribution of *D. immitis* in Europe.[4]

[3] Bowman D. et al. (2009) Prevalence and geographic distribution of *Dirofilaria immitis*, *Borrelia burgdorferi*, *Ehrlichia canis*, and *Anaplasma phagocytophilum* in dogs in the United States: Results of a national clinic-based serologic survey *Veterinary Parasitology* 160; 138-148

[4] European map of *Dirofilaria immitis* and *D. repens* distribution, http://www.evsrl.it/vet.journal/archivio_immagini_grandi/2009/3532.jpg

The heartworm infection is spreading into areas previously considered free of disease and the transmission is also active in heavily urbanized areas. The disease is spread by the vector, the mosquito (Culicidae).

In the last decades, the prevalence of the heartworm increased among dogs and cats across the US and Europe (Figures 3 and 4). Recognizing the importance of the disease, the American Heartworm Society was established in 1974 (10). The society addresses various scientific aspects associated with this potentially fatal disease of dogs, which persists even as diagnostic methods advance, preventive therapies improve, and disease awareness increases among veterinary professionals and pet owners. Another task of the society is to establish protocols and guidelines for diagnosis, treatment and prevention in dogs and cats (10). The spread of the disease through North America is mainly due to a movement to the North, along rivers and streams, of dogs and mosquitoes. The factors favoring the spread of the parasite are related to the behavior of vectors and reservoir hosts and to human activities, travel of infected dogs. The prevalence rises also due to the urbanization of the wildlife-parasite relationship resulting in a higher prevalence in the urban zones (11). However, the main factor is the temperature. It influences the mosquito abundance in the environment and expands the mosquito activity season (12; 13).

The overall prevalence of the feline heartworm disease is low. Infection with immature adults results in heartworm associated respiratory disease (HARD). Usually, no clinical signs are observed. Cats infected experimentally showed lesions in the lungs as a consequence of the arterial disease and the intense interstitial pneumonia. In dogs, radiographs are used as a tool to assess infections with adult *D. immitis*. In cats, the severity of the lung disease cannot be evaluated by lung radiography. Moreover, the dog heartworm prevalence cannot be used to estimate the prevalence in cats (14). Feline show a native resistance to *D. immitis* and are less attractive to mosquitoes than dogs (15).

Life cycle

The infection of the dog occurs when an infected mosquito takes a blood meal and infective larvae are deposited on the skin of the dog in a drop of the haemolymph (Figure 5). L_3 enter the animal's body via the puncture wound made by the mosquito

and migrate through the subcutaneous tissues and thereafter through the muscular tissues.

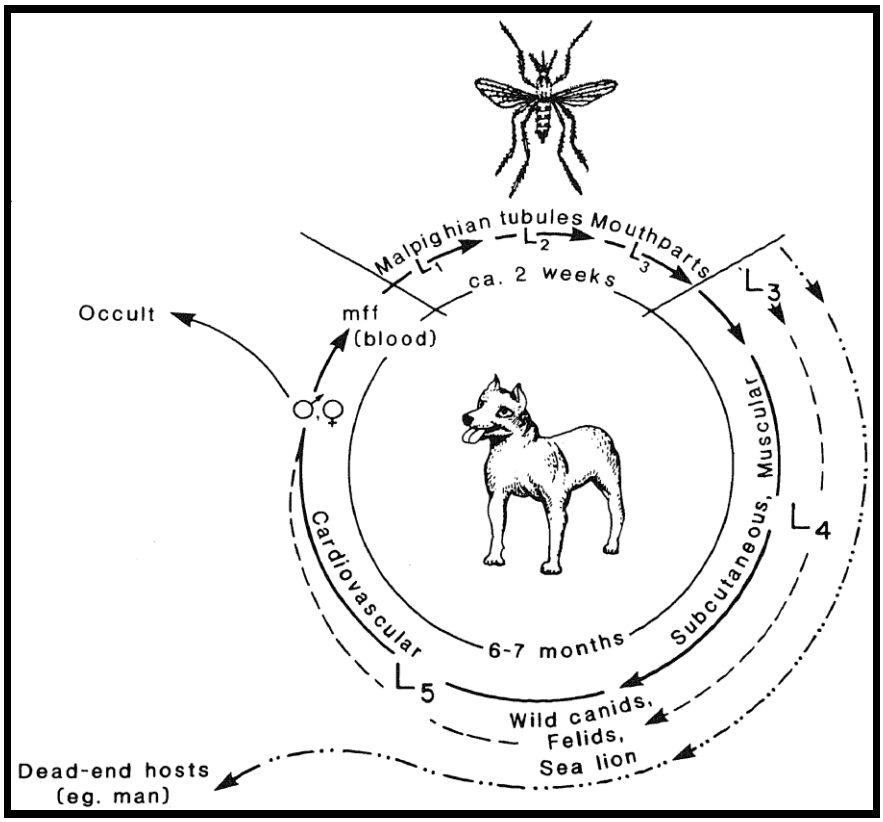


Figure 5. *D. immitis* life cycle.[5]

Presumably larvae are oriented towards the upper abdomen and thoracic cavity, traveling between muscle fibers and molting (16). Some 45 to 60 days after infection, a second molt occurs and fifth-stage larvae (immature adults) enter the bloodstream and migrate to the heart, where they lodge in the arteries of the lungs. Young adults continue to grow, become sexually mature around 120 days post-infection and mate. When juvenile heartworms first reach the heart and lungs, the pressure of venous blood forces them into the small pulmonary arteries. As they grow and increase in size, they progressively migrate upstream into larger arteries until the worms become fully mature. The eventual location of the mature adult worms appears to depend mainly on the size of the dog and the worm burden. Approximately 190 to 197 days after infection, microfilariae appear in the circulating blood (17). The lifetime of adult heartworms in dogs is about 5 years. The period between the initial infection when

[5] R.B. Grieve, “Epidemiology of canine heartworm infection”, 1983.

the dog is bitten by a mosquito and the maturation of worms into adults living in the heart takes 6 to 7 months, the prepatent period, before females begin producing microfilariae (18). Microfilariae are unsheathed and circulate in the vascular system. Their concentration in blood varies over a 24-hour period and seasonally. Microfilariae have been shown to develop to the infective stage in a great variety of mosquitoes of the genera *Aedes*, *Anopheles*, *Culex* and *Mansonia* (19; 20; 21). The pathophysiological response in the dog is mainly due to the presence of adults in pulmonary arteries, provoking a pulmonary hypertension that can progress to congestive heart failure. Microfilariae play a minor pathologic role, while larvae or adults can disturb the blood flow. The spectrum of pathologies related to chronic heartworm infection is broad.

Diagnosis, treatment and resistance

Diagnosis of dirofilariosis is mainly done by detecting the circulating microfilariae with a blood test or the adults with an antigen test (16). The ultimate goal in any heartworm treatment is the elimination of all adult heartworm with minimal post-treatment complications. To accomplish this aim a thorough understanding of the host-parasite interaction is necessary. A combination of doxycycline (10 mg/kg for 30 days) and ivermectin (6 µg/kg/15 days for 6 months), or another macrocyclic lactone, may be a valid alternative, for adulticide therapy in *D. immitis* naturally infected dogs (22).

The only adulticide available on the market is melarsomine dihydrochloride (Immiticide®). It is given through intramuscular injection into lumbar muscles. The complications with adulticide include thrombosis and clogging of pulmonary arteries due to dead worms (23). Currently, no adulticide treatment exists for cats. Figure 6 illustrates the timeline of *D. immitis* development, age of infection in relation to the susceptibility of treatment. The susceptibility gap can be eliminated by administering a macrocyclic lactone preventive for two to three months prior to administering melarsomine. The larvae less than two months old will be eliminated while the older larvae will have reached an age at which they will be susceptible to melarsomine (9).

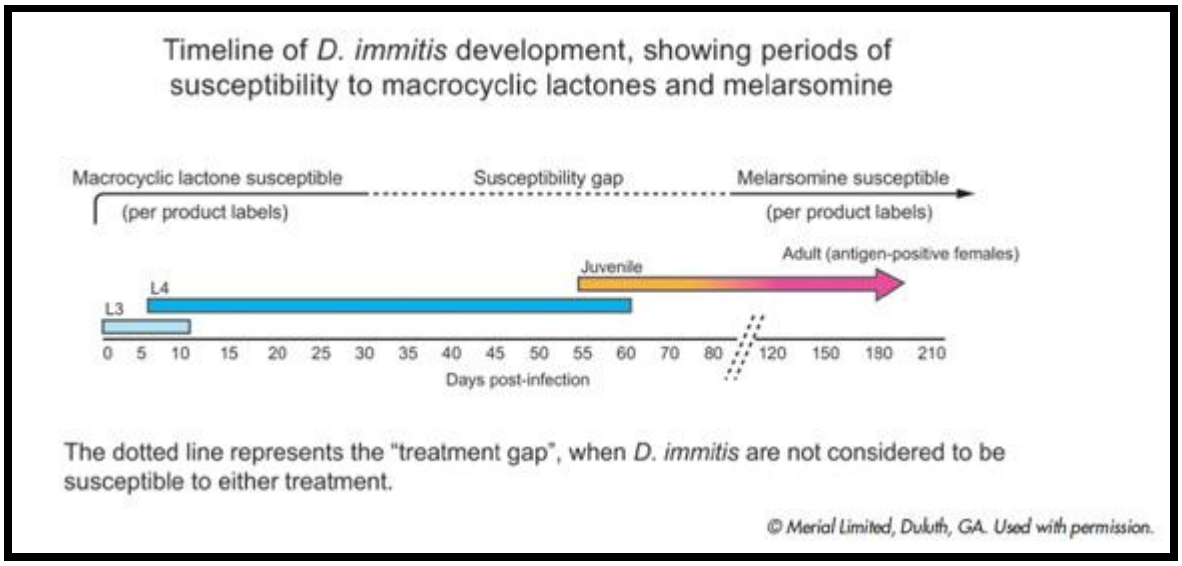


Figure 6. Susceptible and non-susceptible ages of heartworm to macrocyclic lactones and melarsomine based upon the age of the worm in days.[6]

Current knowledge about MLs as heartworm preventives were summarized by McCall (24). Many preventives are currently available against heartworm lifestages (not against adults), and ivermectin has the highest safety-net (efficacy duration and rate) but a miserable adulticidal activity (approximately only 20%), reducing the risk for the dog.

MLs, preventives are highly effective against the majority of *D. immitis* isolates. However, a lack of effectiveness of macrolide heartworm chemoprophylactic agents was described recently and appears to have a genetic basis (25). A successful prophylaxis requires drugs to be administered at the appropriate dose and time for the duration of the period of exposure to infection. For an appropriate treatment it is recommended to test dogs annually. The growing lack of efficacy of MLs' is mainly due to failure in prophylactic administration. To reduce the infection rate, it is important to maintain annual heartworm tests and to reduce the time dogs spend outdoors (26).

Early 2011, Bourguinat et al. (27) described for the first time a resistant case of heartworm to MLs. A naturally infested dog remained positive in microfilarial tests despite repeated treatments with high doses of MLs. The authors highlighted the

[6] http://www.heartwormsociety.org/veterinary-resources/canine-guidelines.html

presence of two single nucleotide polymorphisms in the P-glycoprotein gene which correlate with a reduction of the sensitivity to MLs.

To overcome resistance to the main chemical classes utilized against the heartworm, a new preventive chemical class, with a new mode of action, would be most helpful. In the discovery of a new class of anthelmintics, the amino-acetonitrile derivative (AAD) is very important (28). AADs are active against various species of gastro-intestinal nematodes, have a low toxicity to mammals, and have favourable pharmacokinetic properties. AAD anthelmintic action requires a unique subtype of nicotinic acetylcholine receptor (nAChR) which is specific to nematodes. In *C. elegans* this receptor is *acr-23*, and loss of function mutations in ACR-23 cause AAD resistance (28). In *Haemonchus contortus*, AAD sensitivity is mediated by the receptors MPTL-1 and DES-2. All these proteins (ACR-23, MPTL-1, and DES-2) belong to the DEG-3 group that is a nematode specific subfamily of nAChR. AADs are allosteric activators of these receptor channels (29).

Dirofilaria genomics raises hopes for a better understanding of the parasite, the disease, its prevention and treatment (30). Sequencing of 4005 expressed sequence tags (ESTs) of *D. immitis* adults identified about 1800 genes (31). 70% of these identified *D. immitis* genes showed significant similarity to a *B. malayi* gene. However, there was no DEG-3 type nAChR gene among the *D. immitis* ESTs, while *B. malayi* possesses two such genes (32; 33). ESTs sequencing only identifies expressed genes. So ultimately, the question whether *D. immitis* possess DEG-3 type nAChR genes can only be resolved by whole genome sequencing.

Old and new sequencing strategies

DNA sequencing

DNA sequencing refers to various methods to determine the nucleotide sequence of a DNA molecule. DNA sequencing is useful in different ways such as diagnosis, biotechnology, forensic biology, systematics, evolution and identification of drug and vaccine targets (35). The increase in sequencing speed with novel methods has been accelerating biological research and discoveries.

At the end of the 1970s the first biological sequencing methods were developed (36-40). In 1980, Sanger F., Gilbert W. and Berg P. were awarded with the Nobel Prize in Chemistry for their brilliant inventions in the DNA-sequencing area (41). Dideoxy sequencing in particular allowed the determination of the complete DNA sequence of thousands of genes and over 250 genomes (42).

Premises - Sanger, Maxam and Gilbert

In 1977, Maxam and Gilbert (39) developed another sequencing method based on chemical modification of DNA followed by subsequent cleavage at specific bases. In the Maxam-Gilbert technology, DNA fragments to be sequenced need to be purified and are radioactively labeled at one end. Chemical treatment generates breaks at a small proportion of one or two of the four nucleotide bases in each of the four reactions, G, A+G, C, and C+T. This ends in a series of labeled fragment of different length that are arranged side by side in an electrophoresis gel and they are separated by size. The sequence is reconstructed after exposing the gel to X-ray film (39).

The Sanger sequencing is based on the incorporation of dideoxyribonucleotides (Figure 7) (34; 37). Dideoxyribonucleotides do not have a 3’ hydroxyl being chemically synthesized DNA template strand, DNA polymerase, short primers (to initiate the synthesis of DNA by DNA polymerase), one of the four tri-phosphate dideoxyribonucleotides, ddATP, ddGTP, ddCTP or ddTTP. If a dideoxyribonucleotide is incorporated in the growing DNA chain, due to the lack of 3’ hydroxyl group the addition of the following nucleotide is blocked and the DNA chain is terminated. This template produces a mix of DNA of different lengths, complementary to the DNA

sequenced and terminated at each of the different dideoxynucleotides included (A, T, C or G) (37; 38).

The products of the four reactions are separated by electrophoresis on a polyacrylamide gel in parallel columns. The synthesized fragments are detected by an incorporated, fluorescent or radioactive marker (in the primer or in one of the tri-phosphate deoxyribonucleotides). Reading the bands in ascending order, beginning at the bottom of the gel, the DNA sequence is reconstructed (34; 37; 38).

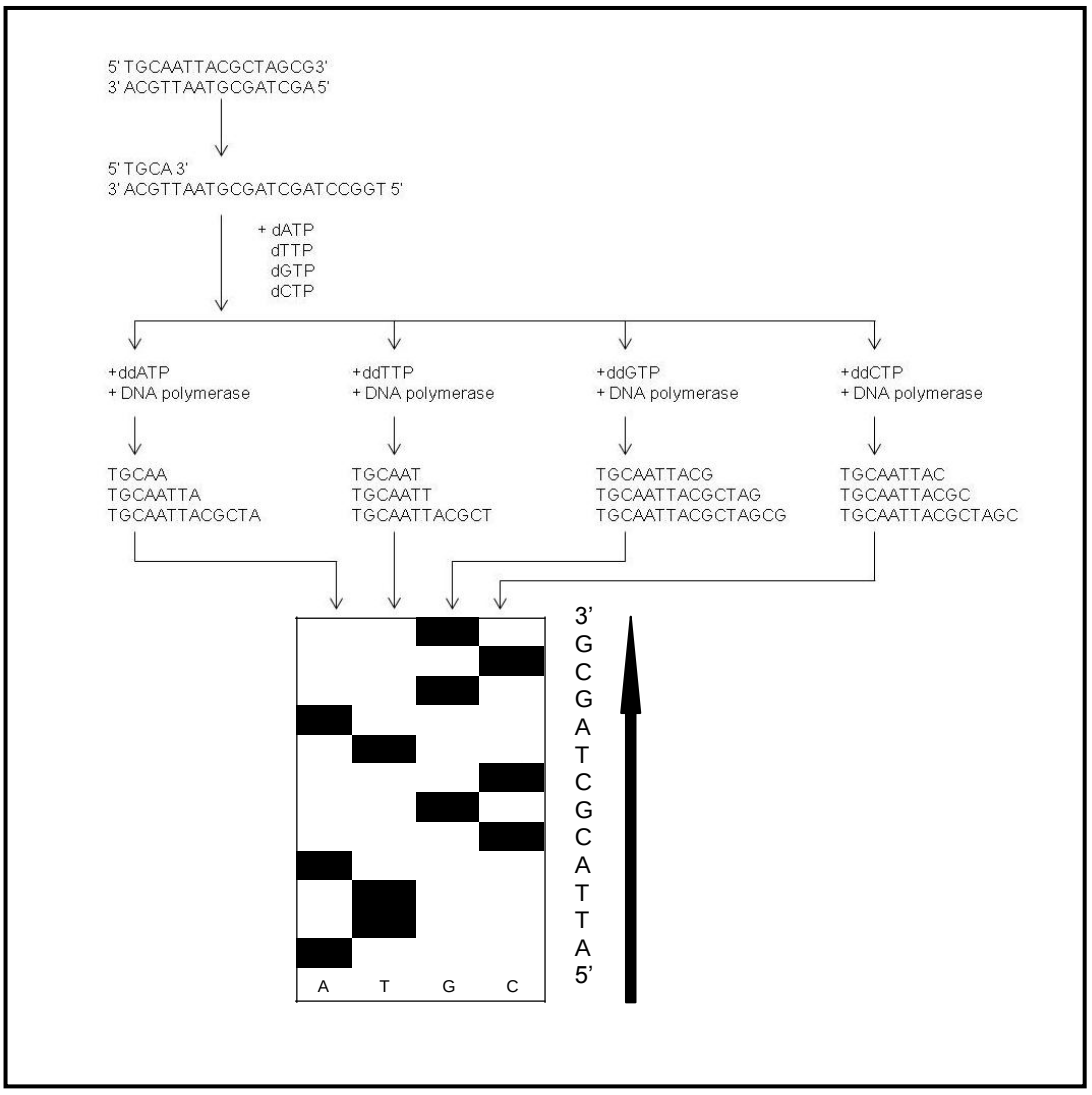


Figure 7. Illustration of the Sanger dideoxy method for DNA sequencing. Addition of primers, DNA polymerase and labeled dideoxyribonucleotides, lacking the 3' hydroxyl group, to a mix of DNA simple strand. Without the 3' hydroxyl group, reactions are stopped producing a group of DNA strands of various sizes. These products are separated by size on electrophoresis gels. The labeled nucleotides allow the reading of the sequences over the four columns of the gel. See text for explanation.[7]

[7] Figure drawn by C. Godel (2012).

Dye-labeling terminator sequencing

In the dye-labeling terminator sequencing method, each of the four tri-phosphate dideoxyribonucleotides carries a different fluorescent label. This allows sequencing in a single reaction rather than four (44-46). This method relies on the use of a modified DNA polymerase that accepts the fluorescently labeled dideoxynucleotide as a substrate (34). The automation of the dye-terminator sequencing technology by Smith et al. (1986) (44) is now, along with high-throughput DNA sequence analyzers and next generation sequencing, used for most of the sequencing projects. The obtainable sequence length is about 1000 nucleotides (47).

Automated sequencing and high-throughput sequencing development

In 1986, Applied Biosystems and the laboratory of Leroy Hood announced a revolution in the sequencing world by launching the automated ABI 370A DNA sequencer (44). This technology allowed sequencing of expressed sequenced tags (ESTs), first successfully done by Craig Venter, as an approach to discover new genes. Today the EST database contains over 43 million ESTs from more than 1300 different organisms (47).

Craig Venter also introduced shotgun sequencing of whole genomes. This uses restriction enzymes or mechanical forces to shear large DNA fragments into smaller ones, which are cloned into a vector and amplified in *E. coli*. Then they are individually sequenced and assembled into contiguous sequence (contigs). Gaps in the sequence are filled with the PCR (48).

The assembly is challenging and may be complicated by sequence repeats (49). However, the first human genome was sequenced with a shotgun approach (50).

In the last decades the sequencing speed increased and the costs dropped drastically with the invention of new machines and technologies (Figure 8) (51; 62).

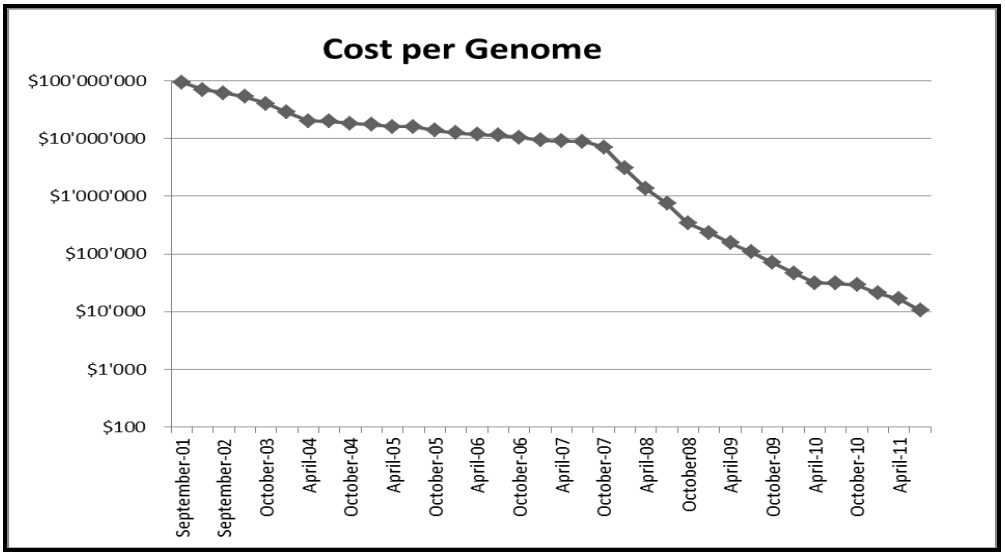


Figure 8. Decreasing costs to sequence a human genome over 10 years.[8]

Next generation sequencing technologies comprise 454 sequencing, SOLID, Illumina sequencing, Ion Torrent, and PacBio. In the following, I shall concentrate on 454 and Illumina, which are presently the most widely used.

New technologies allow to sequence up to one billion bases in a single day at low cost. They lay out millions of DNA fragments on a single chip and sequence all the fragments in parallel. However, these new sequencing technologies produce read lengths as short as 50-400 nucleotides, increasing the challenge for genome assembly (52).

454-Pyrosequencing

In 2005, 454 Life Sciences released the first genome sequenced by pyrosequencing, the genome of *Mycoplasma genitalium* (53). The company 454, founded by J. Rothberg (54), was purchased by Roche Diagnostics in 2007. 454 sequencing is based on real-time pyrophosphate sequencing (55). DNA is fragmented, ligated to specific adapters, which serve as binding sites for primers in a polymerase chain reaction (PCR) in water droplets in an oil solution, known as emulsion PCR. Each droplet contains a single DNA fragment attached to a single primer-coated bead. After amplification, the template DNA covered beads are loaded into picolitre-volume wells etched into the surface of a fiber optic slide (54; 56), one bead per well.

[8] Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program Available at: www.genome.gov/sequencingcosts

Reaction mix and sequencing enzyme are added. Luciferase is used to generate light from ATP, which in turn is built from the pyrophosphate fread upon nucleotide insertion (53). The accuracy of the 454 sequencing technology coupled with the genome sequencer FLX reaches 99.99%. This technology does not only have a high accuracy, it also works fast; producing more than 100 million bases per 8 hour run (56). The longest reads are over 1kb in length (57), which will help mapping in repetitive regions. The main drawback of this technology is the inability to resolve mononucleotide stretches longer than about 8. Pyrosequencing has been involved in numerous de novo genome projects as bacterial, fungal and viral genomes, sequencing of small RNA populations or sequencing of genome from ancient organisms, mammoth (58) or Neanderthal (59).

Illumina sequencing

Like the 454 sequencing technology the Illumina genome analyzer is based on the concept of “sequencing by synthesis” to produce reads of 36-150 bp from tens of millions of surface amplified DNA fragments simultaneously. Using blocked nucleotides, it overcomes the mononucleotide stretch problem of 454 and yields over a million of high-quality short reads per run totalling several gigabases of aligned sequences (62). Illumina sequencing also known as Solexa sequencing (63) used for whole genome sequencing, genome resequencing, transcriptomics and small RNA identification and quantification (64). The fragments and all four nucleotides are added simultaneously to the flow cell channels, along with DNA polymerase, and grown in clusters to provide a stronger fluorescent signal (48-49). The sequencing reaction takes place in flow cell surfaces which contain oligonucleotides. Flow cell surfaces are divided into eight sealed glass lanes where bridge amplification takes place on the inside surface. The addition of DNA polymerase to the surface produces multiple “polonies” (DNA copies). One cluster contains about a million copies of the original fragment (48). For additional details see Figure 9.

The sequencing workflow of the Illumina genome analyzer is initiated by the fragment preparation: random fragmentation of the template DNA and ligation of adapters to both ends of the fragment. Then, the DNA single stranded fragments bind to the surface of flow cell channels, which are lined with a complementary sequence of the adapter. To initiate the bridge amplification, unlabeled nucleotides and DNA polymerase are added, resulting in double stranded fragments. A PCR type reaction

is performed and several million of DNA clusters are generated in each channel of the flow cell (DNA polonies). These are sequenced over multiple cycles adding one sort of labeled dNTP per cycle. Laser excitation detects the image in which of the polonies the nucleotides were incorporated. Before the initiation of the next chemistry cycle, the blocked 3' terminus and the fluorophore from each incorporated bases are removed. Additional chemistry cycles are performed to obtain the complete sequencing (48; 60).

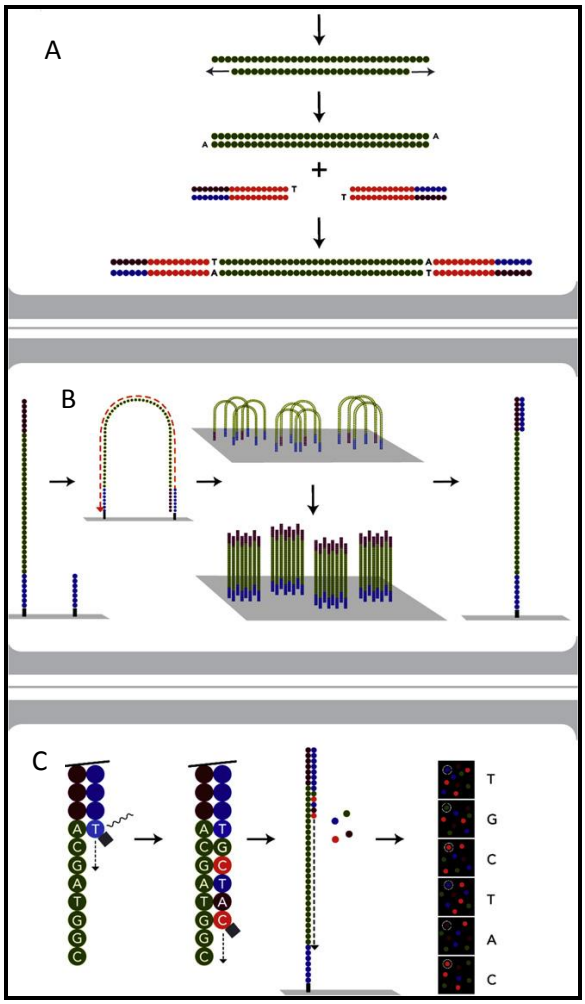


Figure 9. Illumina Hiseq. Sequencing workflow of the Genome Analyzer. A. Generation of sequencing libraries by random fragmentation of DNA strand and ligation of adapters at both ends of the fragment. B. Addition of fragments to the flow cell coated with complementary oligonucleotides. It is followed by the formation of bridges by hybridization and the amplification of the 3' to 5' in order to form clusters. C. A denaturation step occurs followed by the addition of primers, polymerase and labeled nucleotides. The surface is imaged and the process can be repeated.[9]

[9] Tucker T. et al. (2009) Massively parallel sequencing: the next big thing in genetic medicine *The American Journal of Human Genetics* 85, 142-147).

The Illumina technique is limited by the short read length between 36 to 150 bases. However, it is faster and cheaper than other methods. The mate-pair method was developed to obtain longer intervals between reads. As an example to demonstrate the feasibility of the whole genome shotgun sequencing strategy, the genome of the giant panda was sequenced using Illumina Genome Analyzer sequencing technology (65).
Paired-end libraries are short reads and mate-pair reads are between 2 to 5 kb apart and thus very helpful for scaffolding (68).

Third generation of high-throughput sequencing

The “one-base-at-a-time” sequencing, also named third generation sequencing (68), refers to the sequencing from a single DNA molecule and therefore does not need any amplification of the DNA template before sequencing (48).

In 2003, Helicos Biosciences and Braslavsky et al. (70) described and licensed a Single Molecule Real-Time (SMRT) technology sequencing a single DNA molecule at a time (69). Commercially available in 2007, this method relies on the “true single molecule sequencing” (60). The fluorescently labeled nucleotides are detected followed by chemical cleavage of fluorophores enabling the next fluorescent labeled nucleotide addition and DNA elongation (71). Up to 28 Gb over 8 days can be sequenced with this method in a single run with short reads length of 55 bp approximately. This technology brings more accurate homopolymer and direct RNA sequencing (72). The current accuracy is greater than 99%. Its ability to explore very long reads and its fast sequencing speed will dramatically decrease the sequencing costs (68).

Ion semiconductor sequencing, known as Ion Torrent sequencing, came to market in 2010 (75). The Ion Torrent technology measures the proton which is released upon incorporation of a nucleotide into the growing strand of DNA (76-77). The main advantages are that it does not require modified nucleosides and fluorescence images (which use up enormous amounts of disc spaces).

Summary and further analysis

Table 1 summarizes the different high-throughput sequencing methodologies.

Table 1. Summary of different methods of high-throughput sequencing and their specific characteristics.

Sequencing methods	ABI 3730xl Genome analyzer	Roche 454 FLX	Illumina genome analyzer and HiSeq	ABI SOLiD (Life technologies)	Helicos Biosciences	Ion Torrent	Pacific Biosciences
Platform	ABI 3730xl Genome analyzer	GS FLX Titanium, GS Junior	HiSeq 2000, Genome Analyzer IIX, Genome Analyzer IIE, iScanIQ	ABI SOLiD, SOLiD4	HeliScope	Ion Torrent ion semiconductor sequencer	PACBIO RS
Sample requirements		1µg for shotgun library, 5µg for paired-end	<1g for single of paired-end libraries	<2µg for shotgun library, 5-10µg for paired-end	<2µg, single end only		10-20 ng for whole genome amplification, 500ng for 1kb or smaller fragments
Chemistry	Automated Sanger sequencing	Pyrosequencing on solid support	Sequencing-by-synthesis with reversible terminators	Sequencing by ligation	Reversible dye terminators	Ion semiconductor sequencing	Single molecule detection by light pulses emitted as a byproduct of nucleotide incorporation
Method of amplification of the sample	In vivo amplification via cloning	Bead-based / Emulsion PCR	Bridge PCR	Bead-based / Emulsion PCR	Single molecule sequencing	Emulsion PCR	Single molecule sequencing
Template preparation		Clonal-ePCR on bead surface	Clonal bridge enzymatic amplification on glass surface	Clonal-ePCR on bead surface	Single molecule detection		Single molecule detection
Detection method		Light emitting from secondary reactions initiated by release of pyrophospate	Fluorescent emission from incorporated dye-labeled nucleotides	Fluorescent emission from ligated dye-labeled oligonucleotides	Real-time detection of fluorescent dye in polymerase active site during incorporation	Real-time detection of released hydrogen ions by ion sensor	Real-time detection
Read lengths	700-1000 bp	500-1000 bp	36-150 bp	35-75 bp	55 bp	200bp	200bp -10kb
Accuracy		99.99%	>98.5%	99.99%	>99%	99.6%	99.999%
Total throughput, bases per run		0.40-0.60 Gb	3-6 Gb	10-20 Gb (180Gb per run)	28 Gb	100 Mb	3Gb
Run time	2Mb / day	8 hours	4-9* days	7-14* days	8 days	1.5 hours	30min to 1 day
Pros		Long reads helps mapping in repetitive regions; quick runs	Most broadly used	Two-base encoding provides inherent correction error. High accuracy due to two-base encoding technology	No amplification required, lower cost, more accurate homopolymer and direct RNA sequencing	Rapid sequencing speed, low costs, small bench machine	Longer reads length, rapid sequencing speed, easy use, protocol flexibility, low costs
Cons	Less sequences generated, considerably more expensive	High costs (reagent) and high error rate in homopolymer repeats	Low multiplexing capability of samples	Long run time	Asynchronous sequencing, higher raw error rate increasing the cost per base	Difficult to enumerate long repeats, short read length and low throughput	Quite new technology so it is hard to define cons.

Use		Whole genome, transcriptome, metagenome and deep sequencing	Whole genome or targeted resequencing, transcriptome, gene regulation, de novo and metagenomics analysis	Whole genome resequencing (fragment library or mate-pair library), comparative resequencing, enrichment technology, comparative transcriptomic sequencing	Whole genome resequencing, transcriptome resequencing and analysis, small RNA sequencing, Quantitative polyadenylation site mapping/Digital Gene Expression,...	Small scale applications (microbial genome /transcriptome sequencing, targeted or amplicon sequencing, quality testing	Whole genome amplification, target resequencing, pathogen application, transcript analysis, can resolve SNPs and large scale structural rearrangements
References	78, 79, 81	49, 56, 60, 62, 80, 81, 82	49, 60, 62, 80, 81, 84	49, 60, 61, 62, 80, 81, 83	61, 70, 72, 85, 86, 88	87	73, 74, 89

Assembly methods

The first step after sequencing is usually a check of the reads, discarding short or ambiguous reads. Then the reads that pass the quality check need to be assembled into contigs, and these into scaffolds. In theory, read assembly is perfectly solvable by alignment algorithms. In praxis, however this fails because of the enormous numbers of reads. Today, the most efficient assemblers rely on graphs where the reads (= nodes) are linked by common words (k-mers = edges). The most of these assemblers consider the reads as network called a graph. The mainly used short read assemblers are Velvet (de Bruijn graph methods), SOAPdenovo, Forge or ABySS. The field is developing very fast. The quality of the assembly is measured based on different criteria, such as the contig size distribution (N50, longest contig/scaffold), the number of bases in contigs, length of the longest contig, etc.

The algorithms for assembly are based on Eulerian paths, Hamiltonian paths, or both (de Bruijn path). Leonhard Euler is a Swiss mathematician who founded graph theory by solving the problem of the seven bridges of Könisberg in 1736. An Eulerian path visits every edge (link) in a graph exactly once. In practice, it is much more complicated to construct a Hamiltonian path, which visits every vertex (node) exactly once.

Velvet uses the De Bruijn graph methods as do most of the next-generation sequencing assemblers. A de Bruijn graph reduces the computational effort by breaking reads into k-mers; the parameter k denotes the word length (90; 91). The de Bruijn graph captures perfect overlaps of length k-1 between these k-mers and not between the actual reads (92). The best k-mer size for a particular assembly is

determined empirically. Repeats in the genome are collapsed (92). However, some of the collapsed repeats could be biologically meaningful.

ABySS (Assembly By Short Sequences) uses a distributed representation of the de Brujin graph. The technology is based on a two stage algorithm, generation of k-mers from the sequence reads and building of contigs, and extension of the contigs based on sequence pairs (93).

Example of de novo genome sequencing

More and more sequenced genomes become available due to the drastic drop of the cost and time of sequencing. Table 2 lists species with fully sequenced genome. Most of the bigger genomes have been still sequenced by classical Sanger sequencing but this is changing.

Table 2. Comparison between various animal species of the genome size, number of annotated ligand-gated ion channels and the sequencing method for the genome studies.

Species	Phylum	NGS	Coverage	Genome (Mb)	LGIC	Genome status	References
Dirofilaria immitis	Nematoda Clade III	HiSeq2000 (Illumina)	~170-fold	~90	24	Draft	This work
Brugia malayi	Nematoda Clade III	Whole Genome Shotgun	~9-fold	~95	24	Complete	(108) (113)
Onchocerca volvulus	Nematoda Clade III	Genome Analyzer II (Illumina) //Roche 454	-	25.99	?	In progress	(98) (99) (101)
Caenorhabditis elegans	Nematoda Clade V	Fingerprint – clones/Solexa	~20-fold	100	84	Complete	(95)(96)
Ascaris suum	Nematoda Clade III	HiSeq2000 (Illumina)	~80-fold	273	5	Draft	(97)
Meloidogyne hapla	Nematoda Clade IVb	ABI 3730 and MegaBase Sequencers	10.4	54	32		(94)
Haemonchus contortus	Nematode Clade V	HiSeq2000 (Illumina)	-	60	41	In progress	(98)
Schistosoma mansoni	Platyhelminth Trematoda	Genome Analyzer II (Illumina)	~6-fold	270	13	Draft	(98)
Echinococcus multilocularis	Platyhelminth Cestoda	Genome Analyzer II (Illumina)	-	150	13	In progress	(98)
Danio rerio	Vertebrata		~21-fold	1500	56	Draft	(98) (107)
Xenopus tropicalis	Vertebrata	High shotgun seq	~7.6-fold	1500 ?	38	Draft	(106)
Homo sapiens sapiens	Vertebrata	Whole Genome Shotgun /Sanger /454	~7.4-fold	2865	36	Draft	(110) (112)
Canis familiaris	Vertebrata	High shotgun seq	~7.5-fold	2041	36	Draft	(100)
Felis catus	Vertebrata	High shotgun seq	~1.9-fold	~3000	21	Draft	(109)

Drosophila melanogaster	Diptera	Whole Genome Shotgun	~23-fold	180	21	Draft	(103)
Aedes aegypti	Diptera	Whole Genome Shotgun	~12-fold	1376	16	Draft	(105) (111)
Anopheles gambiae	Diptera	Whole Genome Shotgun	~10.2-fold	260	20	Draft	(104)
Culex quinquefasciatus	Diptera	Whole Genome Shotgun	-	540	19	In progress	(102)

Anthelmintics and targets

A major driving force for the sequencing of pathogen genomes has been the identification of novel drug targets. Until now, however, empirical screening for new drugs has been more successful than rational, post-genomic approaches.

Most current anthelmintics target the ion channels. Avermectins hyperactivate glutamate chloride gated channels, Cyclodeopsipeptides activate the SLO-1 K^+ channel, Imidazothiazoles and AADs hyperactivate the nAChR and Praziquantel inhibits calcium channels (Figure 10). The next paragraph will focus on the ligand gated-ion channels and in particular on nAChR.

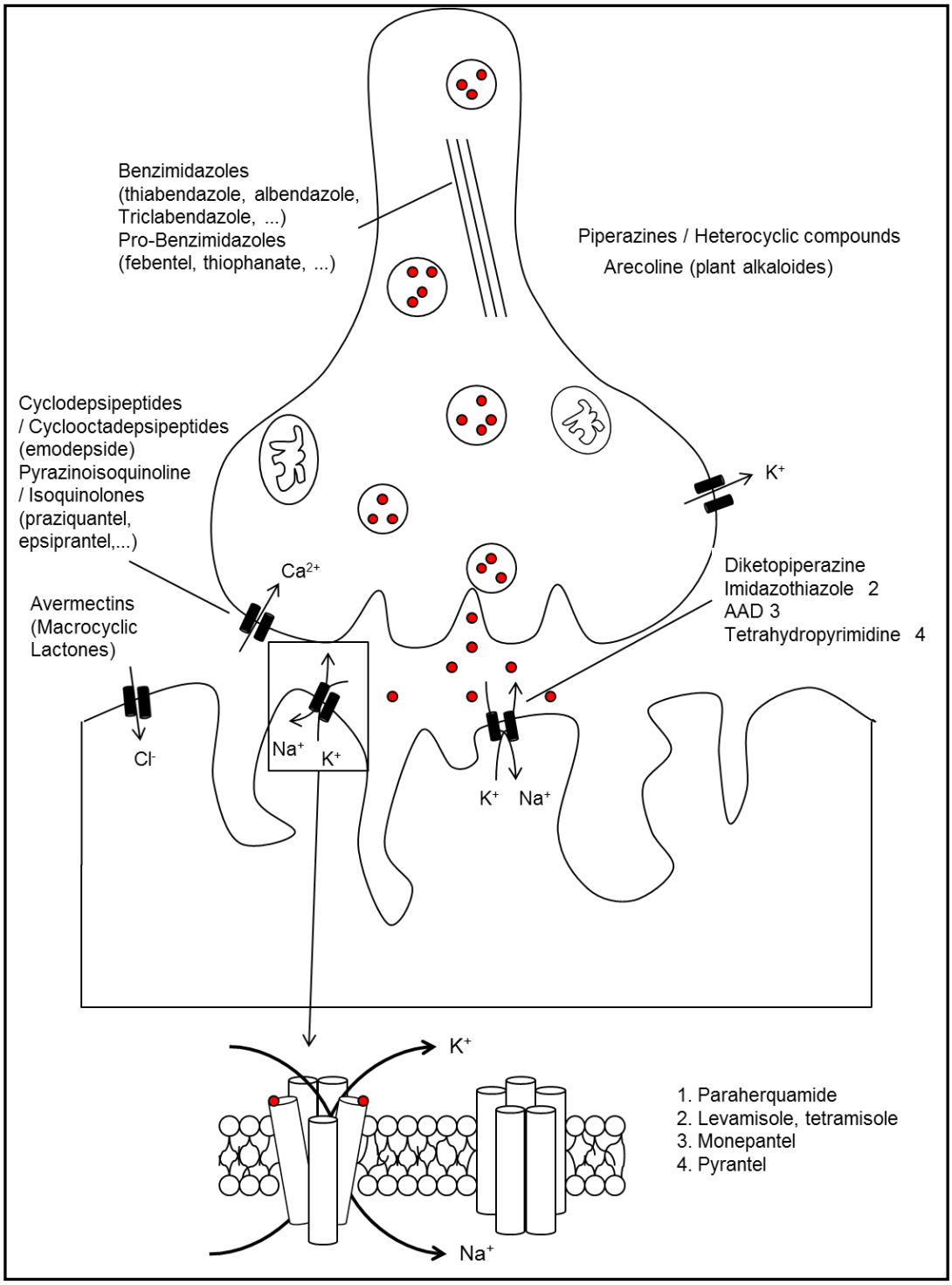


Figure 10. Illustration representing the mode of action and localization of numerous different anthelmintics at the neuromuscular junction. The different class of drug represented have specific mode of action, targeting for example tubulins for the benzimidazoles, calcium channels for the cyclodepsipeptides and chloride gated channels for avermectins. Related to each class, some drug names are given as example.[10]

[10] C. Godel (2012) , unpublished.

Nicotinic acetylcholine receptors and cys-loop ligand-gated ion channels

Ligand-gated ion channels (LGIC) are widely spread across the Kingdom animals, bacteria, unicellular eukaryotes and plants. The importance and pharmacological potential of LGIC in animals is demonstrated by the fact that many natural toxin act on them. The superfamily of cys-loop ligand-gated ion channels represents the most important targets of the current anthelmintics in parasitic nematodes (114-115).

Nicotinic acetylcholine receptors belong to the class of the LGIC as well as serotonin 5-HT, $GABA_A$ and $GABA_C$, and glycine receptors (117-118).

Located in postsynaptic membranes of central nervous system synapses and at the neuromuscular endplates, the nAChR are cholinergic receptors forming ion channels activated by acetylcholine but also nicotine. These glycoproteins are cation channels selective for Na^+, K^+ and less frequently to Ca^{2+}. The channels structure is pentameric with 2 α, one β, γ, and δ subunit surrounding a central pore. The subunits are named in relation to their increasing molecular weights. The channels open by extra-cellular binding of a chemical messenger (agonist). The nicotinic acetylcholine receptors are the link between nerve cell and muscle fiber. When a receptor is stimulated by a chemical messenger, it causes depolarization muscular contractions by opening the channel. This muscular stimulation is caused by the flow of cations through the receptor after the binding of neurotransmitters. The signal activated by the neurotransmitter is received from the pre-synaptic neuron (116). The pentameric glycoprotein presents a cys-loop, a cysteine doublet in the N-teminal part that participates in the acetylcholine (ACh) binding site. The receptor is divided into nine α neuronal subunits encoding the nAChR additionally to the three β subunits (119-120). Next to the cys-loop the nAChR possess four specific, conserved transmembrane domains (TMI-TMIV). The receptor is structurally divided in two main parts; the ligand-binding domain with a long extracellular N-terminal chain presenting the glycolysation sites, and the large transmembrane channel domain, about 100-200 amino acids, phosphorylation sites between TMIII and TMIV and a short C-terminal domain about 4-28 amino acids (117-118). The nicotinic acetylcholine receptors are illustrated in Figures 11 and 12.

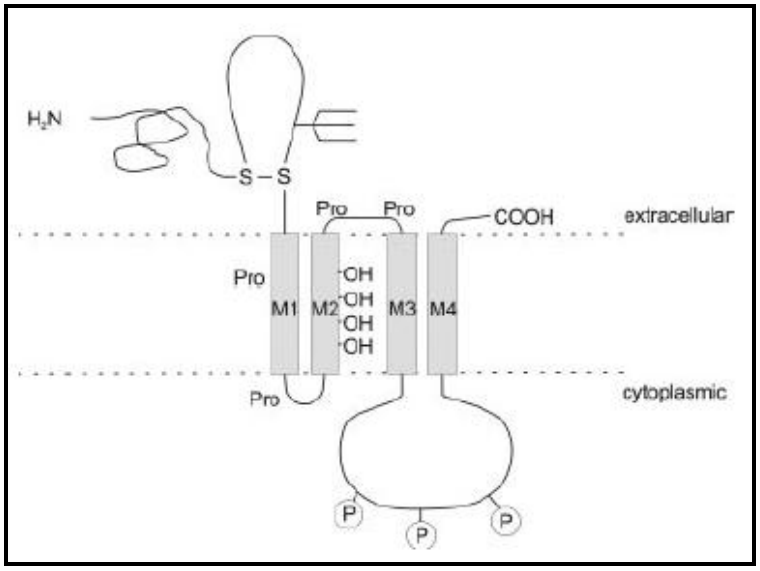


Figure 11. Schematic representation of nicotinic acetylcholine receptors protein with the four transmembrane domains (M1-M4), the extracellular chain with the N-terminal and glycolysation sites and the intracellular domain with the cys-loop. [11]

The free-living nematode *C. elegans* is the model worm for parasitic species and has the most LGIC genes of any known organism. With the growing numbers of sequenced genomes available, phylogenomic studies are possible of LGIC families in diverse species. As an example, in 2007 Williamson et al. presented the LGICs of *Brugia malayi* and *Trichinella spiralis* (120). For both parasites showed fewer LGIC than *C. elegans*. Rufener et al. (2010), illustrated the phylogeny of LGIC genes of nematodes, vertebrates, insects and platyhelminths, based on the lingand binding domains (29). This paper correlated the inventory of LGIC genes with monepantel susceptibility. All the monepantel sensitive species possessed an ACR-23 or MPTL-1 ortholog, the presumed target of AADs (121). The LGIC inventory of the heartworm, however, is a black box since no single LGIC has been cloned from *D. immitis*. The closely related filaroidea *B. malayi* does not possess any ortholog to *acr-23*. However, *deg-3* and *des-2* genes are present (108; 121); they may play a role in the mode of action of the AAD.

[11] Schematic representation of the nAChRs. (78) (Hucho F. and Weise C. (2001) Ligand-gated ion channels *Angew. Chem. Int. Ed.* 40:3100-3116)

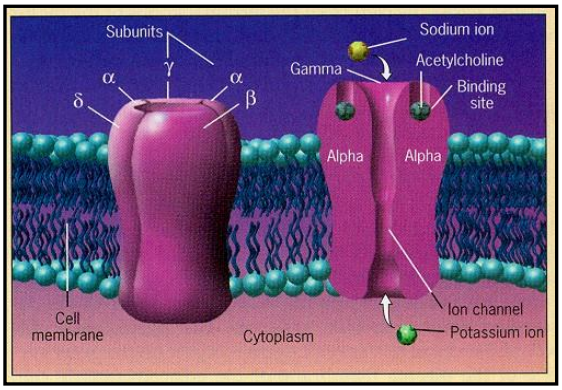


Figure 12. Pentameric nicotinic acetylcholine receptor. The neurotransmitter binds to the receptor and allows the cation flow through the channel causing muscular contractions.[12]

[12] http://chemweb.calpoly.edu/cbailey/377/PapersSp2000/Kevin/nicotine.html

Aims and objectives

The main goal of this PhD project was to identify and characterize potential drug targets in the animal parasitic nematode, *Dirofilaria immitis.* Of particular interest were ligand-gated ion channels of the DEG-3 subfamily as potential targets of monepantel.

The specific objectives were chronologically (1) to standardize and implement in vitro drug sensitivity tests with *D. immitis*, (2) to sequence the genome of *D. immitis* and predict its proteome in particular and (3) to identify target LGIC, characterize these in silico based on phylogenetic comparisons, (4) potential drug targets to correlate the drug susceptibility with the presence or absence of the predicted target(s).

Chapter 1. Understanding the magic bullet: molecular opportunities for antiparasitic drug selectivity

Publication 1

Manuscript under construction

Review Article

Understanding the magic bullet: molecular opportunities for antiparasitic drug selectivity

Christelle Godel[1-3] and Pascal Mäser[1,2*]

[1]Swiss Tropical and Public Health Institute, 4002 Basel, Switzerland

[2]University of Basel, 4002 Basel, Switzerland

[3]Novartis Animal Health, 1566 St. Aubin, Switzerland

[*]Corresponding author

In this publication, my contribution focused on the writing of the introduction and literature researches as well as participation to the writing of the main parts.

Understanding the magic bullet: molecular mechanisms of antiparasitic drug selectivity

Pascal Mäser[1,2*], Marcel Kaiser[1,2], Christelle Godel[1-3]

[1]Swiss Tropical and Public Health Institute, 4002 Basel, Switzerland

[2]University of Basel, 4002 Basel, Switzerland

[3]Novartis Animal Health, 1566 St. Aubin, Switzerland

[*]Corresponding author Swiss Tropical and Public Health Institute
Socinstrasse 57
4002 Basel
Switzerland
Phone +41 61 284 8338
Fax +41 61 284 8101
E-mail pascal.maeser@unibas.ch

Parasite chemotherapy has a simple rationale: to kill the pathogen but not its host by applying compounds of selective toxicity. The possibility of selective activity was first inferred by Paul Ehrlich who, graduating in histology in 1887, had observed that certain organic dyes exhibit greater affinity for parasites than for host cells. Ehrlich concluded that selective antiparasitic chemotherapy is possible if such dyes are cytotoxic, and only four years later he put the idea into practice by successful treatment of two malaria patients with methylene blue (Guttmann et al., 1891). Ehrlich coined the term *chemotherapia specifica*, also known as *magic bullet*, for a chemical compound like methylene blue that exhbits discriminating activity against the pathogen and thus opens a therapeutic window for treatment of an infected host. The degree of selectivity of a drug in vivo is expressed as the therapeutic index, the LD_{50} of a drug divided by its ED_{50} (where LD_{50} is the lethal dose that kills 50% of test animals and ED_{50} is the effective dose which cures 50% of the test animals). The larger the therapeutic index, the greater the margin of safety. To a first estimate the selectivity of a drug may also be assessed in vitro by comparing its 50% inhibitory concentrations (IC_{50}) towards parasites and mammalian cells (Kaminsky et al., 1996). Table 1 shows the in vitro selectivities of selected parasiticides as determined in our laboratory. Some of the drugs have selectivity indices of several hundreds or thousands. What is the molecular basis of such striking parasite-specific activity? Here we try to dissect the molecular mechanisms of antiparasitic drug selectivity. Concentrating on the nature of the difference between the drug target in the parasite and its closest orthologue in the mammalian host, we distinguish six different scenarios (A to F, summarized in Table 2) which create opportunities for antiparasitic drug selectivity, all of which are documented by examples of antiparasitic drugs.

A) There is no orthologue of the antiparasitic drug target in the host proteome. This scenario corresponds to the classical definition of an antibiotic drug target, i.e. an enzyme that is essential to the pathogen but absent in the host. The most prominent example comes from prokaryotes: transpeptidase, the target of penicillin. Transpeptidase builds the stabilizing cross-links in the murein cell wall of bacteria. Mammals lacking murein as well as transpeptidase, they are not susceptible to

penicillins. Absence of a human orthologue is a criterion that fits antibacterial drug targets in particular, given the large evolutionary distance between host and pathogen. However, drug targets of bacterial origin also occur in eukaryotic parasites: in endosymbionts or remnants thereof.

Wolbachia are obligate intracellular bacteria which colonize arthropods and filarial nematodes (Saridaki et al., 2010). Living in tight association with their host cell, *Wolbachia* display reduced genomes and may represent a transitory stage between parasite and symbiont. While in arthropods, *Wolbachia* frequently exert harmful effects including male-killing and cytoplasmic incompatibility (Saridaki et al., 2010), their relationship with nematodes is mutualistic. Wolbachia possibly provide parasitic nematodes with purines and heme, and they are required for the maturation of oocytes. Doxycycline treatment of dogs infected with the heartworm *Dirofilaria immitis* reduced the numbers of circulating microfilariae, caused degeneration of oocytes in adult female, and necroses in adult male worms (Bazzocchi et al., 2008). However, the doxycycline treatment did not completely clear the *Wolbachia* from *D. immitis* (Bazzocchi et al., 2008; Rossi et al., 2010). The *Wolbachia* genomes of *Brugia malayi* (Foster et al., 2005) and *D. immitis* (Godel et al., submitted) encode for about 800 predicted proteins. Many of these are likely to be essential – for the simple reason that if not, they would have been lost in the course of evolution. Potential antifilarial drug targets encoded by endosymbiont *Wolbachia* comprise the cell division protein FtsZ (Li et al., 2011), phosphoglycerate mutase (Foster et al., 2009), pyruvate phosphate dikinase (Raverdy et al., 2008) and, somewhat surprisingly for an intracellular bacterium, also the gene products of the *mur* operon for murein biosynthesis (Godel et al., submitted).

Drug targets of bacterial origin are also found in the phylum Apicomplexa, which contains the causative agents of malaria, toxoplasmosis, babesiosis, east cost fever and coccidioses. All these parasites possess an apicoplast, the remnant of a secondary endosymbiont as indicated by the four surrounding membranes and bipartite targeting signals of imported proteins (Foth et al., 2003). Only the innermost compartment contains DNA, which is of prokaryotic provenance. The metabolic

pathways in the apicoplast of *Plasmodium falciparum* were reconstructed in silico based on the prediction of import signals (Foth et al., 2003; Ralph et al., 2004). Promising antimalarial targets in the apicoplast are enzymes of the deoxyxylulose-phosphate (i.e. non-mevalonate) pathway for isoprenoid synthesis such as DOX-P reductoisomerase, the target of fosmidomycin (Jomaa et al., 1999) (Table 2). Enzymes of the type II pathway for fatty acid synthesis have been proposed as drug targets too; however, initial reports of triclosan efficacy in a malaria mouse model (Surolia et al., 2001) were not reproducible later on (Baschong et al., 2011).

There is a number of further target proteins, besides those that are relics from prokaryotes, which are absent from mammals but essential to parasites. The presence of the folate synthetic enzymes dihydropteroate synthase (EC 2.5.1.15) and dihydrofolate synthase (EC 6.3.2.12) in *Plasmodium* spp. and other apicomplexans renders them susceptible to sulfonamides. Trypanosomatids possess a variety of biochemical peculiarities (Opperdoes, 1985) which lend themselves to chemotherapeutic intervention: trypanothione and ergosterol are shown in Figure 1. While initial leads against *T. cruzi* were azoles targeting Δ24 sterol methyl transferase (E.C. 2.1.1.43) (Urbina et al., 1996), presently the most promising candidates for new drugs against *T. cruzi* are triazole inhibitors of cytochrome P-450-dependent C14 sterol demethylase (CYP51) (Urbina, 2009). Ergosterol in their membranes is also thought to render *Leishmania* susceptible to amphotericin B, enabling the docking of amphotericin B and formation of pores (Ramos et al., 1996). Chitin synthesis is another pathway of pharmacological importance. Chitin is made by a plethora of pathogens and pests including fungi, nematodes, and arthropods (Figure 1) but not by mammals. Inhibitors of chitin synthase (EC 2.4.1.16) therefore have a great potential as broad-spectrum parasiticides. Lufenuron is used in the veterinary sector against fleas and heartworm, nikkomycin Z is in clinical development for human mycoses (Chaudhary et al., 2012).

B) There is an orthologous target in the host, but it is less sensitive. Tubulins fulfill essential functions in cell division and vesicular trafficking in all eukaryotic cells, and belong to the most

conserved proteins of eukaryotes. The β-tubulins 2B from *Homo sapiens* (NP_821080) and tbb-4 from *Caenorhabditis elegans* (NP_509585), for instance, are over 93% identical. And yet, benzimidazole drugs selectively bind the *C. elegans* protein, blocking the polymerization of microtubules in nematodes but not in mammals. Benzimidazoles are widely used against phytopathogenic fungi, gastrointestinal nematodes of lifestock, soil-transmitted helminths of humans, cestodes, and *Giardia intestinalis*. Drug resistance studies have demonstrated that a single point mutation in a nematode β-tubulin, phenylalanine-200 to tyrosine (Kwa et al., 1994), suffices to convey benzimidazole resistance (Kwa et al., 1995). The same mutation was also reported from benzimidazole resistant field isolates of phytopathogenic fungi (Ma et al., 2003; Banno et al., 2008; Suga et al. 2011). Intriguingly, mammalian β-tubulins naturally carry a tyrosine at position 200, suggesting that a single hydroxy group determines over drug susceptibility. Tyrosine-200 was proposed to be involved in a hydrogen bond preventing benzimidazoles from binding to β-tubulin (Robinson et al., 2004). Additional point mutations described from benzimidazole resistant nematodes and fungi line out the presumed benzimidazole-binding pocket of β-tubulin (Robinson et al., 2004). In particular, mutation of glutamate-198 (which is conserved in nematodes and fungi) to alanine was found in high-level resistant isolates (Ma et al., 2003; Ghisi et al., 2007; Rufener et al., 2009).

C) There is an orthologous target in the host proteome, but it is not essential. This is a rather theoretical possibility which is poorly supported by actual drugs. Purine interconversion enzymes might be candidate targets. Since all obligate endoparasitic protozoa studied so far lack the enzymes for purine de novo synthesis; presumably, the corresponding genes were lost in adaptation to a parasitic lifestyle. The mammalian host being able to make purines de novo, the salvage is unlikely to be essential. However, due to the high degree of redundancy in purine interconversion pathways, parasite are unlike to be essential either. Even in *Toxoplasma gondii*, whose purine metabolic network is very simple, knock-out parasites of single purine salvage genes were still viable (Chaudhary et al., 2004). One example for an enzyme that is a drug target because it is essential to

the pathogen but not to the host is reverse transcriptase. This enzyme is an attractive antiretroviral drug target that is also encoded in the human genome, in retrotransposons (de Parseval et al., 2005). 'Human' reverse transcriptases are unlikely to be essential, even though the enzyme has been implicated in repair of chromosomal breaks (Teng et al., 1996).

D) There is an orthologous, sensitive target in the host, but it is less accessible. Drug transport phenomena are increasingly being recognized as determinants of drug susceptibility (Dobson et al., 2009; Lanthaler et al., 2011). The presence or absence of a particular transporter may decide on whether a drug reaches its target tissue, cell, or subcellular compartment. The importance of drug transport was illustrated in a tragical way by the pyrimidine analogue fialuridine, which had been in clinical development as an antiviral agent. Fialuridine had a inconspicuous safety profile in the tested animal models but killed five hepatitis B patients in a clinical trial (McKenzie et al., 1995). The reason for the selective toxicity turned out to be the presence of the nucleoside transporter ENT1 in the human mitochondrial inner membrane, importing fialuridine into mitochondria (Lee et al., 2006). The toxicity of fialuridine could not be anticipated in mice since the mouse ENT1 orthologue does not carry a mitochondial targeting signal.

Another example of how the absence of a transporter may be protective is provided by the melamine-based arsenical melarsoprol, still the only approved therapy for the treatment of late-stage *T. b. rhodesiense* sleeping sickness. Melarsoprol and its metabolite melarsen oxide are mainly imported into trypanosomes via an aminopurine permease termed P2 (Carter et al., 1993), encoded by the gene *TbAT1* (Mäser et al., 1999; Geiser et al., 2005). The absence of an aminopurine-selective transporter from mammalian cells may contribute to the therapeutic window of melarsoprol. Homozygous deletion of *TbAT1* reduced the sensitivity of *T. b. brucei* to melarsoprol (Matovu et al., 2003), but not to the membrane-permeable phenylarsine (Lüscher et al., 2006b). Expression of *TbAT1* in *Saccharomyces cerevisiae* rendered the transformants susceptible to melarsen (Mäser et al., 1999). The amidine-based substrate recognition motif of TbAT1 was mapped (de Koning et al.,

1999; Lüscher et al., 2006a), and covalent linkage of a P2-recognition haptophore (*sic* Ehrlich) such as melamine was proposed to selectively deliver toxophores to trypanosomes (Tye et al., 1998; Barrett et al., 2006).

Transport barriers may also decide over drug susceptibility. The mammalian brain is protected from hydrophobic xenobiotics by P-glycoprotein, a detoxification efflux pump expressed by the endothelial cells of the blood-brain barrier. Among the many substrates of P-glycoprotein is also ivermectin. Ivermectin is a semisynthetic avermectin that activates glutamate-gated chloride channels of the postsynaptic membrane in helminths and arthropods, thereby hyperpolarizing the target cell (Dent et al., 2000; Yates et al., 2004). Mammals do not possess such channels. However, ivermectin also acitvates the GABA-gated chloride channels of the mammalian brain (Adelsberger et al., 2000; Wolstenholme). Thus the principal reason for the selectivity of ivermectin appears to be that ivermectin is a P-glycoprotein substrate and cannot reach its targets in the mammalian brain. The protective role of P-glycoprotein is highlighted by the fact that *MDR1* deficiency in dogs is linked to ivermectin toxicity (Wolstenholme, 2010). Homozygous deletion of *MDR1* caused ivermectin hypersusceptibility also in mice (Schinkel et al., 1994). In parasitic nematodes, ivermectin resistance was accompanied by loss of heterogeneity at a nematode P-glycoprotein locus (Blackhall et al., 1998; Xu et al., 1998).

E) There is an orthologous, sensitive, and accessible target in the host, but it is replenished. Dihydrofolate reductase, required to make thymidine from uridine, is an essential enzyme in all cells. *Plasmodium falciparum* DHFR is the target of the antimalarial pyrimethamine (combined with sulfadoxine to Fansidar). The in vitro selectivity index of pyrimethamine (Table 1) is much larger than would be anticipated from the different affinities of recombinant DHFR enzymes from *H. sapiens* and *P. falciparum* (Zhang et al., 2002). The intricate regulation of DHFR was shown to contribute to the hypersensitivity of *P. falciparum* to inhibitors. Mammalian as well as *Plasmodium* DHFR levels are under the control of negative feed-back: the protein binds its own mRNA, preventing it from being

translated (Zhang et al., 2002). This block is releived upon inhibition of human DHFR by folate analogues such as pyrimethamine, because the active site of the enzyme participates in binding of the mRNA. This is not the case for *P. falciparum* DHFR, which continues to bind its mRNA even when inhibited by pyrimethamine. *P. falciparum* expressing human DHFR were as tolerant to DHFR inhibitors as human cells (Zhang et al., 2002).

F) A prodrug is only activated in the parasite. Nucleoside and nucleobase analogues are used as antiviral and anticancer drugs; many are also active against protozoan parasites. Typically nucleoside or nucleobase analogues need to be converted to the corresponding nucleotides. Once in the nucleotide pool, the analogues may inhibit nucleic acid polymerases or be mis-incorporated into nucleic acids themselves. Thus nucleoside kinases and nucleobase phosphoribosyltransferases are activating enzymes and potential determinants of drug susceptibility. This was demonstrated for *Toxoplasma gondii*, where inosine analogues were identified that are subversive substrates of *T. gondii* adenosine kinase but not of human adenosine kinase (Yadav et al., 2004; Al Safarjalani et al., 2010). The identified compounds were selectively toxic to the parasites (el Kouni et al., 1999). Adenosine kinase is of pharmacological importance also in trypanosomes since RNAi-mediated silencing of adenosine kinase in *T. brucei* reduced the sensitivity to cordycepin (3'-deoxyadenosine) and adenine arabinoside (Lüscher et al., 2007; Vodnala et al., 2008).

The nitroimidazoles are a widely used class of prodrugs. Fexinidazole and nifurtimox are used against *Trypanosoma* spp., metronidazole against *Giardia*, *Trichomonas*, and *Entamoeba*. Nitroimidazoles need to undergo single electron reduction to become active, converting to $–NO_2$ group to the $–NO_2^{-}\cdot$ radical (Figure 2). This requires a strong reducing agent to produce the nitroradical anion and an anaerobic environment to prevent its re-oxidation in a futile cycling process (Tocher, 1997). The nitroradical anion then reacts with thiol groups, covalently attaching the drug to target proteins. In *Entamoeba histolytica* and *Trichomonas vaginalis*, thioredoxin reductase was shown to be activator and target of metronidazole at the same time (Leitsch et al., 2007; Leitsch et

al., 2009). *T. vaginalis* mutants selected for metronidazole resistance had lost thioredoxin enzymatic activity, probably as a consequence of lack of the cofactor flacin adenine dinucleotide (FAD) as the susceptibility was restored upon exogenous addition of FAD (Leitsch et al., 2009). Trypanosomatids possess NADH-dependent, bacterial-like type I nitroreductase in their mitochondrium, which are thought to be crucial for the activation of 'nitro-drugs' such as benzimidazole, nifurtimox, and fexinidazole. Selection for nifurtimox resistance in *T. cruzi* (Wilkinson et al., 2008) and *T. brucei* (Baker et al., 2011) resulted in loss of nitroreductase, and heterozygous deletion of the *NTR* gene in *T. brucei* conferred nifurtimox resistance (Wilkinson et al., 2008). Mammalian cells, like most eukaryotes, do not possess type I nitroreductases (Tweats et al., 2012).

Another prominent class of parasiticides that are activated by reduction are the artemisinins. Semi-synthetic artemisinins are used as antimalarials and they are also active against schistosomes (Xiao et al., 1989; Utzinger et al., 2007). The characteristic feature of artemisinins is their endoperoxide bridge. Opened by chemical reduction, it creates radicals that alkylate heme and target proteins. The reducing agent is thought to be Fe^{2+} heme resulting from hemoglobin digestion in the parasite's food vacuole (Fügi et al., 2010). This would explain the selective activity of artemisinins against hemoglobin-feeders such as the *Plasmodium* spp. and *Schistosoma* spp. However, this model was challenged by the finding that the in vitro efficacy of artemisinin against *P. falciparum* increased under a 2% carbon monoxide atmosphere (Parapini et al., 2004).

Conclusion. In summary, the question of how a drug can selectively kill the endoparasite but not the host has at least six mechanistically different answers. Importantly, not all of them are approachable in silico. Comparative genomics between host and parasite is best suited to detect cases of scenarios A and F (Table 2). Most straightforward are targets originating from a bacterial endosymbiont, such as those in the apicoplast or in *Wolbachia*; these are readily recognized based on phylogeny or signal sequences, and likely to be essential since they have been preserved during evolution. Other proven antiparasitic targets, in particular those of scenarios D and E, can hardly be predicted by

bioinformatics and are usually found through cell-based screening. We conclude that it is theoretically impossible to replace phenotypic screening with comparative and structural genomics, and advocate carrying out the different approaches in parallel in order to maximize the likelyhood of discovering novel kinds of parasiticides.

ACKNOWLEDGEMENT

CG is the recipient of the Novartis PhD Student Fellowship program. Research in our laboratory is supported by the Swiss National Science Foundation.

TABLES

Drug	Target parasite	IC_{50} against target parasite [µg/ml]	IC_{50} against L6 fibroblasts [µg/ml]	In vitro selectivity
Chloroquine	*P. falciparum*	0.0030	34	11,000
Mefloquine	*P. falciparum*	0.0034	3.1	910
Pyrimethamine	*P. falciparum*	0.0047	3.0	640
Artemisinin	*P. falciparum*	0.0020	99	49,000
Melarsoprol	*T. b. rhodesiense*	0.0025	7.3	2,900
Pentamidine	*T. b. rhodesiense*	0.00082	2.3	2,800
Nifurtimox	*T. b. rhodesiense*	0.31	37	120
Fexinidazole	*T. b. rhodesiense*	0.61	88	150
Metronidazole	*G. duodenalis*	0.49	79	160
Albendazole	*G. duodenalis*	0.25	0.11	0.44

Table 1. Selected parasiticides and their in vitro activities against the blood-stages of *P. falciparum* (strain NF53) and *T. b. rhodesiense* (STIB901), *G. duodenalis* (G1) trophozoites, and *R. norvegicus* fibroblasts (L6). IC_{50} values were determined after 72 h of incubation in serial drug dilutions using the redox dye Alamar blue as an indicator of cell number and viability (Räz et al., 1997).

The drug target in the host is ...

A)	**absent**	Sulfadoxin	=>	Dihydropteroate synthase
		Fosmidomycin	=>	DOXP reductoisomerase
		Azoles	=>	Ergosterol synthesis
		Nikkomycin Z	=>	Chitin synthase
B)	**not essential**	Didanosine	=>	Reverse transcriptase
C)	**less sensitive**	Benzimidazoles	=>	ß-Tubulin
D)	**less accessible**	Ivermectin	=>	Ligand-gated ion channels
		Melarsoprol	=>	Diverse intracellular targets
E)	**replenished**	Eflornithine	=>	Ornithine decarboxylase
		Pyrimethamine	=>	Dihydrofolate reductase
F)	**conserved, but the drug is not activated**	Nifurtimox	<=	Nitroreductase type I
		Artemisinin	<=	Fe (II) heme
		Metronidazole	<=	Thioredoxin

Table 2. Molecular mechanisms of antiparasitic drug selectivity.

FIGURES

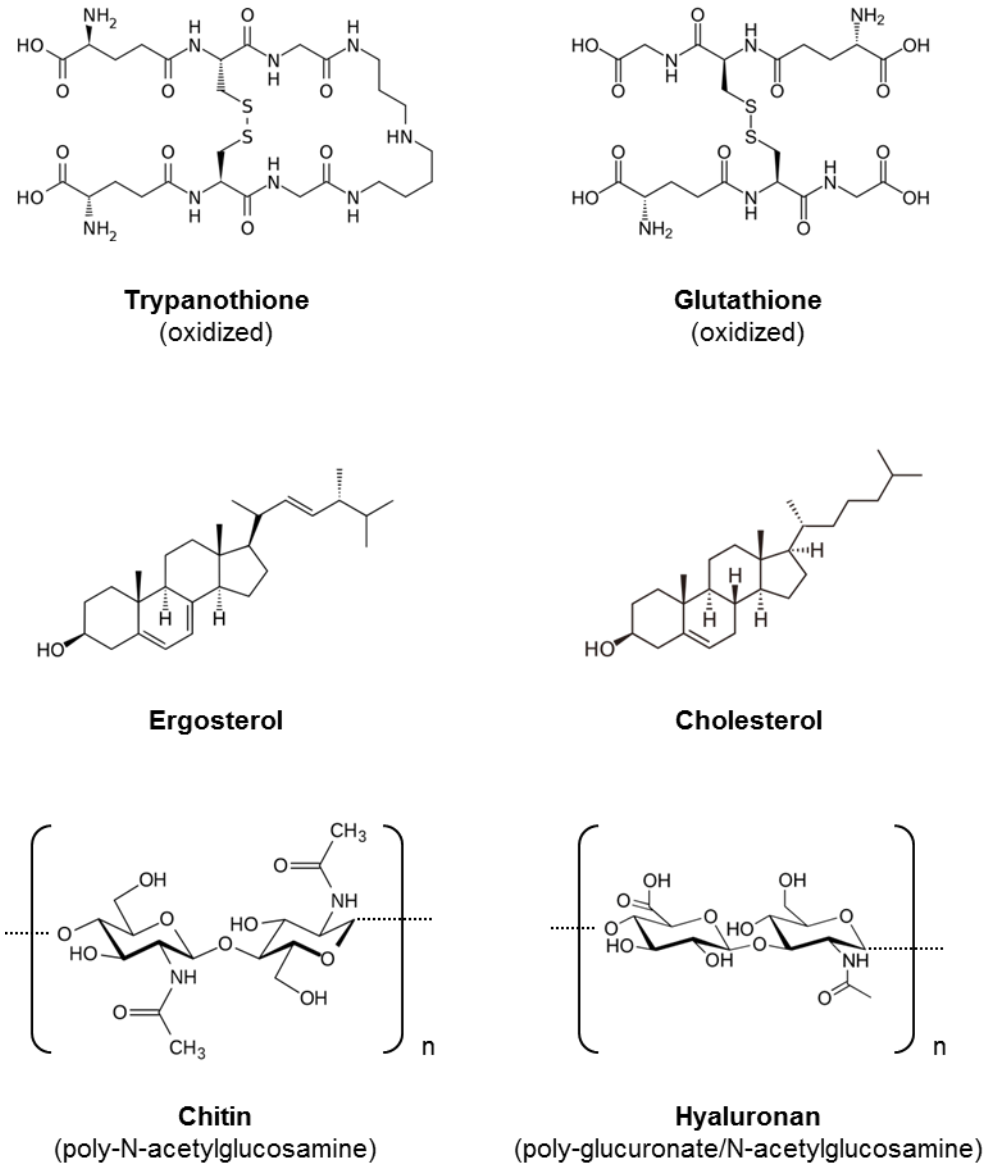


Figure 1. Essential and parasite-specific metabolites and their closest structures in mammalian cells.

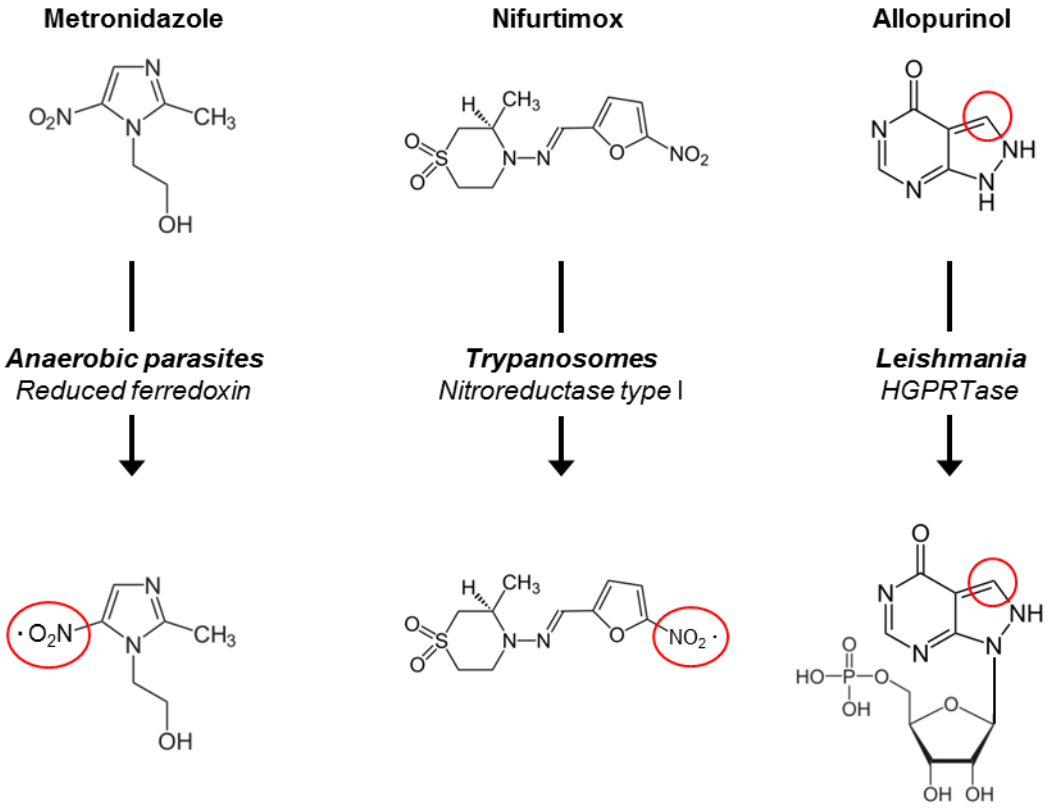


Figure 2. Parasite-specific activation of prodrugs. Several of the currently used parasiticides are prodrugs that are only activated inside the parasite.

Adelsberger H., Lepier A., Dudel J. (2000) Activation of rat recombinant alpha(1)beta(2)gamma(2S) GABA(A) receptor by the insecticide ivermectin. Eur J Pharmacol 394: 163-70

Al Safarjalani O.N., Rais R.H., Kim Y.A., Chu C.K., Naguib F.N., El Kouni M.H. (2010) Carbocyclic 6-benzylthioinosine analogues as subversive substrates of Toxoplasma gondii adenosine kinase: biological activities and selective toxicities. Biochem Pharmacol 80: 955-63

Baker N., Alsford S., Horn D. (2011) Genome-wide RNAi screens in African trypanosomes identify the nifurtimox activator NTR and the eflornithine transporter AAT6. Mol Biochem Parasitol 176: 55-7

Banno S., Fukumori F., Ichiishi A., Okada K., Uekusa H., Kimura M., Fujimura M. (2008) Genotyping of benzimidazole-resistant and dicarboximide-resistant mutations in Botrytis cinerea using real-time polymerase chain reaction assays. Phytopathology 98: 397-404

Barrett M.P., Gilbert I.H. (2006) Targeting of toxic compounds to the trypanosome's interior. Adv Parasitol 63: 125-83

Baschong W., Wittlin S., Inglis K.A., Fairlamb A.H., Croft S.L., Kumar T.R., Fidock D.A., Brun R. (2011) Triclosan is minimally effective in rodent malaria models. Nat Med 17: 33-4; author reply 34-5

Bazzocchi C., Mortarino M., Grandi G., Kramer L.H., Genchi C., Bandi C., Genchi M., Sacchi L., McCall J.W. (2008) Combined ivermectin and doxycycline treatment has microfilaricidal and adulticidal activity against Dirofilaria immitis in experimentally infected dogs. Int J Parasitol 38: 1401-10

Blackhall W.J., Liu H.Y., Xu M., Prichard R.K., Beech R.N. (1998) Selection at a P-glycoprotein gene in ivermectin- and moxidectin-selected strains of Haemonchus contortus. Mol Biochem Parasitol 95: 193-201

Carter N.S., Fairlamb A.H. (1993) Arsenical-resistant trypanosomes lack an unusual adenosine transporter. Nature 361: 173-75

Chaudhary K., Darling J.A., Fohl L.M., Sullivan W.J., Jr., Donald R.G., Pfefferkorn E.R., Ullman B., Roos D.S. (2004) Purine salvage pathways in the apicomplexan parasite Toxoplasma gondii. J Biol Chem 279: 31221-7

Chaudhary P.M., Tupe S.G., Deshpande M.V. (2012) Chitin Synthase Inhibitors as Antifungal Agents. Mini Rev Med Chem

de Koning H.P., Jarvis S.M. (1999) Adenosine Transporters in Bloodstream Forms of Trypanosoma brucei brucei: Substrate Recognition Motifs and Affinity for Trypanocidal Drugs. Mol Pharmacol 56: 1162-70

de Parseval N., Heidmann T. (2005) Human endogenous retroviruses: from infectious elements to human genes. Cytogenet Genome Res 110: 318-32

Dent J.A., Smith M.M., Vassilatis D.K., Avery L. (2000) The genetics of ivermectin resistance in Caenorhabditis elegans. Proc Natl Acad Sci U S A 97: 2674-9

Dobson P.D., Lanthaler K., Oliver S.G., Kell D.B. (2009) Implications of the dominant role of transporters in drug uptake by cells. Curr Top Med Chem 9: 163-81

el Kouni M.H., Guarcello V., Al Safarjalani O.N., Naguib F.N. (1999) Metabolism and selective toxicity of 6-nitrobenzylthioinosine in Toxoplasma gondii. Antimicrob Agents Chemother 43: 2437-43

Foster J., et al. (2005) The Wolbachia genome of Brugia malayi: endosymbiont evolution within a human pathogenic nematode. PLoS biology 3: e121

Foster J.M., Raverdy S., Ganatra M.B., Colussi P.A., Taron C.H., Carlow C.K. (2009) The Wolbachia endosymbiont of Brugia malayi has an active phosphoglycerate mutase: a candidate target for anti-filarial therapies. Parasitology research 104: 1047-52

Foth B.J., Ralph S.A., Tonkin C.J., Struck N.S., Fraunholz M., Roos D.S., Cowman A.F., McFadden G.I. (2003) Dissecting apicoplast targeting in the malaria parasite Plasmodium falciparum. Science 299: 705-8

Fügi M.A., Wittlin S., Dong Y., Vennerstrom J.L. (2010) Probing the antimalarial mechanism of artemisinin and OZ277 (arterolane) with nonperoxidic isosteres and nitroxyl radicals. Antimicrob Agents Chemother 54: 1042-6

Geiser F., Luscher A., de Koning H.P., Seebeck T., Mäser P. (2005) Molecular pharmacology of adenosine transport in Trypanosoma brucei: P1/P2 revisited. Mol Pharmacol 68: 589-95

Ghisi M., Kaminsky R., Mäser P. (2007) Phenotyping and genotyping of Haemonchus contortus isolates reveals a new putative candidate mutation for benzimidazole resistance in nematodes. Vet Parasitol 144: 313-20

Guttmann P., Ehrlich P. (1891) Über die Wirkung des Methylenblau bei Malaria. Berliner Klinische Wochenschrift 39: 953-56

Jomaa H., et al. (1999) Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. Science 285: 1573-6

Kaminsky R., Schmid C., Brun R. (1996) An 'in vitro selectivity index' for evaluation of cytotoxicity of antitrypanosomal compounds. Tropical Medicine & International Health 1: A36-A36

Kwa M.S., Veenstra J.G., Roos M.H. (1994) Benzimidazole resistance in Haemonchus contortus is correlated with a conserved mutation at amino acid 200 in beta-tubulin isotype 1. Mol Biochem Parasitol 63: 299-303

Kwa M.S., Veenstra J.G., Van Dijk M., Roos M.H. (1995) Beta-tubulin genes from the parasitic nematode Haemonchus contortus modulate drug resistance in Caenorhabditis elegans. J Mol Biol 246: 500-10

Lanthaler K., Bilsland E., Dobson P.D., Moss H.J., Pir P., Kell D.B., Oliver S.G. (2011) Genome-wide assessment of the carriers involved in the cellular uptake of drugs: a model system in yeast. BMC Biol 9: 70

Lee E.W., Lai Y., Zhang H., Unadkat J.D. (2006) Identification of the mitochondrial targeting signal of the human equilibrative nucleoside transporter 1 (hENT1): implications for interspecies differences in mitochondrial toxicity of fialuridine. J Biol Chem 281: 16700-6

Leitsch D., Kolarich D., Binder M., Stadlmann J., Altmann F., Duchene M. (2009) Trichomonas vaginalis: metronidazole and other nitroimidazole drugs are reduced by the flavin enzyme thioredoxin reductase and disrupt the cellular redox system. Implications for nitroimidazole toxicity and resistance. Mol Microbiol 72: 518-36

Leitsch D., Kolarich D., Wilson I.B., Altmann F., Duchene M. (2007) Nitroimidazole action in Entamoeba histolytica: a central role for thioredoxin reductase. PLoS Biol 5: e211

Li Z., Garner A.L., Gloeckner C., Janda K.D., Carlow C.K. (2011) Targeting the Wolbachia cell division protein FtsZ as a new approach for antifilarial therapy. PLoS Negl Trop Dis 5: e1411

Lüscher A., de Koning H.P., Mäser P. (2006a) Chemotherapeutic strategies against Trypanosoma brucei: Drug targets vs. drug targeting. Current Drug Targets 13: 555-67

Lüscher A., Nerima B., Mäser P. (2006b) Combined contribution of TbAT1 and TbMRPA to drug resistance in Trypanosoma brucei. Mol Biochem Parasitol 150: 364-6

Lüscher A., Onal P., Schweingruber A.M., Mäser P. (2007) Adenosine kinase of Trypanosoma brucei and its role in susceptibility to adenosine antimetabolites. Antimicrob Agents Chemother 51: 3895-901

Ma Z., Yoshimura M.A., Michailides T.J. (2003) Identification and characterization of benzimidazole resistance in Monilinia fructicola from stone fruit orchards in California. Appl Environ Microbiol 69: 7145-52

Mäser P., Sütterlin C., Kralli A., Kaminsky R. (1999) A nucleoside transporter from Trypanosoma brucei involved in drug resistance. Science 285: 242-4

Matovu E., et al. (2003) Mechanisms of arsenical and diamidine uptake and resistance in Trypanosoma brucei. Euk Cell 2: 1003-8

McKenzie R., et al. (1995) Hepatic failure and lactic acidosis due to fialuridine (FIAU), an investigational nucleoside analogue for chronic hepatitis B. N Engl J Med 333: 1099-105

Opperdoes F.R. (1985) Biochemical peculiarities of trypanosomes, African and South American. Br Med Bull 41: 130-6

Parapini S., Basilico N., Mondani M., Olliaro P., Taramelli D., Monti D. (2004) Evidence that haem iron in the malaria parasite is not needed for the antimalarial effects of artemisinin. FEBS Lett 575: 91-4

Ralph S.A., van Dooren G.G., Waller R.F., Crawford M.J., Fraunholz M.J., Foth B.J., Tonkin C.J., Roos D.S., McFadden G.I. (2004) Tropical infectious diseases: metabolic maps and functions of the Plasmodium falciparum apicoplast. Nat Rev Microbiol 2: 203-16

Ramos H., Valdivieso E., Gamargo M., Dagger F., Cohen B.E. (1996) Amphotericin B kills unicellular leishmanias by forming aqueous pores permeable to small cations and anions. J Membr Biol 152: 65-75

Raverdy S., Foster J.M., Roopenian E., Carlow C.K. (2008) The Wolbachia endosymbiont of Brugia malayi has an active pyruvate phosphate dikinase. Molecular and biochemical parasitology 160: 163-6

Räz B., Iten M., Grether-Bühler Y., Kaminsky R., Brun R. (1997) The Alamar Blue assay to determine drug sensitivity of African trypanosomes in vitro. Acta Trop 68: 139-47

Robinson M.W., McFerran N., Trudgett A., Hoey L., Fairweather I. (2004) A possible model of benzimidazole binding to beta-tubulin disclosed by invoking an inter-domain movement. J Mol Graph Model 23: 275-84

Rossi M.I., Paiva J., Bendas A., Mendes-de-Almeida F., Knackfuss F., Miranda M., Guerrero J., Fernandes O., Labarthe N. (2010) Effects of doxycycline on the endosymbiont Wolbachia in Dirofilaria immitis (Leidy, 1856)--naturally infected dogs. Vet Parasitol 174: 119-23

Rufener L., Kaminsky R., Maser P. (2009) In vitro selection of Haemonchus contortus for benzimidazole resistance reveals a mutation at amino acid 198 of beta-tubulin. Mol Biochem Parasitol 168: 120-2

Saridaki A., Bourtzis K. (2010) Wolbachia: more than just a bug in insects genitals. Curr Opin Microbiol 13: 67-72

Schinkel A.H., et al. (1994) Disruption of the mouse mdr1a P-glycoprotein gene leads to a deficiency in the blood-brain barrier and to increased sensitivity to drugs. Cell 77: 491-502

Suga H., Nakajima T., Kageyama K., Hyakumachi M. (2011) The genetic profile and molecular diagnosis of thiophanate-methyl resistant strains of Fusarium asiaticum in Japan. Fungal Biol 115: 1244-50

Surolia N., Surolia A. (2001) Triclosan offers protection against blood stages of malaria by inhibiting enoyl-ACP reductase of Plasmodium falciparum. Nat Med 7: 167-73

Teng S.C., Kim B., Gabriel A. (1996) Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. Nature 383: 641-4

Tocher J.H. (1997) Reductive activation of nitroheterocyclic compounds. Gen Pharmacol 28: 485-7

Tweats D., Bourdin Trunz B., Torreele E. (2012) Genotoxicity profile of fexinidazole--a drug candidate in clinical development for human African trypanomiasis (sleeping sickness). Mutagenesis

Tye C., Kasinathan G., Barrett M.P., Brun R., Doyle V.E., Fairlamb A.H., Weaver R., Gilbert I.H. (1998) An approach to use an unusual adenosine transporter to selectively deliver polyamine analogues to trypanosomes. Bioorg Med Chem Lett 8: 811-16

Urbina J.A. (2009) Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches. Acta Trop 115: 55-68

Urbina J.A., Vivas J., Lazardi K., Molina J., Payares G., Piras M.M., Piras R. (1996) Antiproliferative effects of delta 24(25) sterol methyl transferase inhibitors on Trypanosoma (Schizotrypanum) cruzi: in vitro and in vivo studies. Chemotherapy 42: 294-307

Utzinger J., Xiao S.H., Tanner M., Keiser J. (2007) Artemisinins for schistosomiasis and beyond. Curr Opin Investig Drugs 8: 105-16

Vodnala M., Fijolek A., Rofougaran R., Mosimann M., Mäser P., Hofer A. (2008) Adenosine kinase mediates high affinity adenosine salvage in Trypanosoma brucei. J Biol Chem 283: 5380-8

Wilkinson S.R., Taylor M.C., Horn D., Kelly J.M., Cheeseman I. (2008) A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. Proc Natl Acad Sci U S A 105: 5022-7

Wolstenholme A.J. (2010) Recent progress in understanding the interaction between avermectins and ligand-gated ion channels: putting the pests to sleep. Invert Neurosci 10: 5-10

Xiao S.H., Catto B.A. (1989) In vitro and in vivo studies of the effect of artemether on Schistosoma mansoni. Antimicrob Agents Chemother 33: 1557-62

Xu M., Molento M., Blackhall W., Ribeiro P., Beech R., Prichard R. (1998) Ivermectin resistance in nematodes may be caused by alteration of P-glycoprotein homolog. Mol Biochem Parasitol 91: 327-35

Yadav V., Chu C.K., Rais R.H., Al Safarjalani O.N., Guarcello V., Naguib F.N., el Kouni M.H. (2004) Synthesis, biological activity and molecular modeling of 6-benzylthioinosine analogues as subversive substrates of Toxoplasma gondii adenosine kinase. J Med Chem 47: 1987-96

Yates D.M., Wolstenholme A.J. (2004) An ivermectin-sensitive glutamate-gated chloride channel subunit from Dirofilaria immitis. Int J Parasitol 34: 1075-81

Zhang K., Rathod P.K. (2002) Divergent regulation of dihydrofolate reductase between malaria parasite and human host. Science 296: 545-47

Chapter 2. Efficacy testing of AADs against *D. immitis* (unpublished)

An in vitro drug sensitivity assay for *D. immitis* was developed (122) and further refined here. In brief, microfilariae are isolated from infested dogs and incubated in a medium developed from Hank's Balanced Salt Solution (HBSS) for up to 72 hours. The efficacy, absence of motility, was measured by the evaluation of AADs against microfilariae of *D. immitis* and *Acanthocheilonema viteae* after 72 hours of incubation.

Preliminary in vitro tests were performed with the AAD, Monepantel. It showed activity against the microfilariae of the filarial nematodes *D. immitis* and of the model parasite, *A. viteae*. L_3 of *A.viteae* were also sensitive in vitro to monepantel.

There was activity at 1 ppm each for *A. viteae* microfilariae and L_3 and *D. immitis* microfilariae (Figure 13).

There was also a high activity of monepantel in the LDA (larval development assay) (123), against *Trichostrongylus colubriformis*, at 0.01 ppm. This is a standard test for the current selection of compounds described by Gill at al (123). It assesses drug efficacy during the development from egg to L_3.

The in vitro tests showed good efficacy against filarial nematodes, *A. viteae* and *D. immitis*. However, in vivo in the gerbil *Meriones unguiculatus*, monepantel exhibited no antifilarial activity (Figure 13).

Milbemycin oxime, a ML, was used as a positive control. Milbemycin oxime is used as a preventive against heartworm disease; medication as Sentinel® combined with praziquantel and lufenuron, and Interceptor®, in combination with Praziquantel.

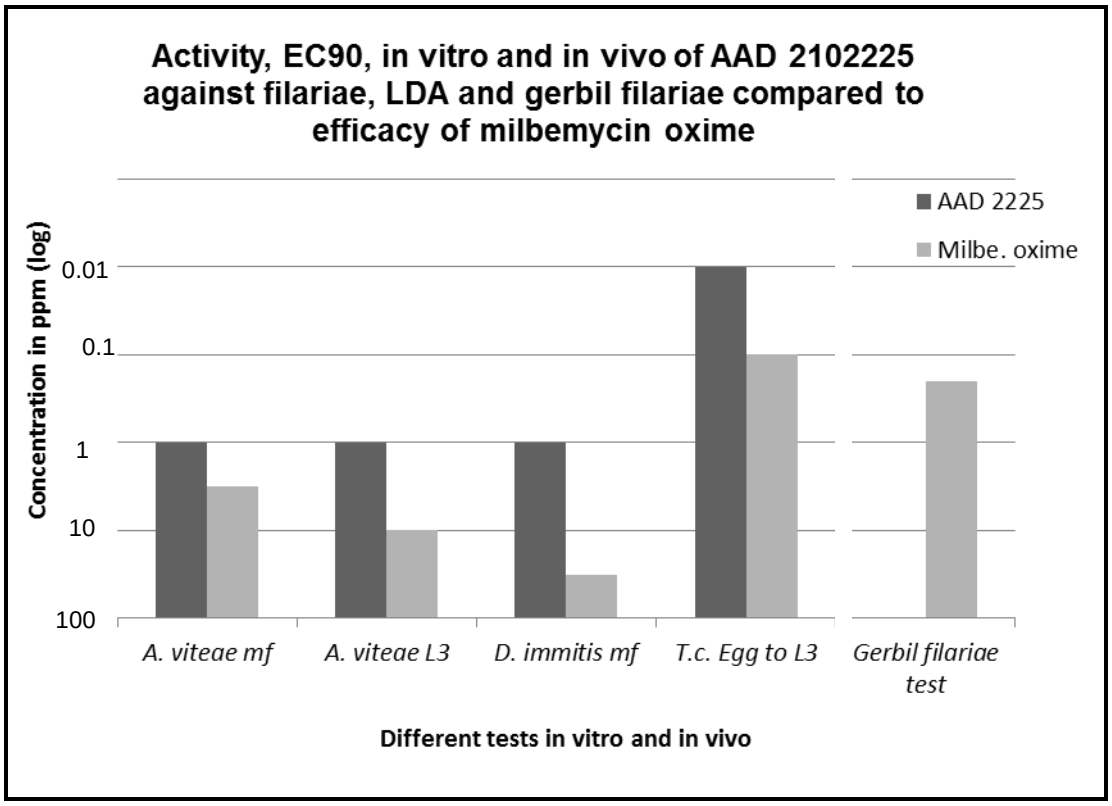


Figure 13. Monepantel (AAD 2225) activity (EC 90) in four in vitro tests (absence of motility) and one in vivo test. Concerning the in vitro tests, the efficacy against microfilariae or L_3 is evaluated after 48 hours of incubation with the compounds. In the LDA, egg to L_3, the incubation time is longer with duration of seven days. The results are represented in ppm (particle per million). In vivo, compounds are applied per os at 40 mg/kg. The efficacy is evaluated at necropsy, by count of the adult parasites. Efficacy is defined by comparison to positive control, milbemycin oxime, and a low concentration in ppm.

In further studies, 29 AADs were evaluated against microfilariae in vitro. Ivermectin and milbemycin oxime were used as positive controls; albendazole, as negative control and a benzithiazole compound was included as well. The aim was to compare the in vitro efficacy of these 29 AADs with existing results against similar or different parasites.

The choice of the compounds was based on chemical structure, availability and efficacy against gastro-intestinal nematodes. The test was validated by the good efficacy of the positive controls, ivermectin and milbemycin oxime.

The AADs exhibited global efficacy against microfilariae of *D. immitis* and *A. viteae* in vitro (Figure 14). Microfilariae of *D. immitis* and *A. viteae* have more or less similar sensitivity. There was a mean efficacy of 82.75% of all AADs.

However, only a small number of compounds showed activity in vivo, even if they had shown high activity in vitro.

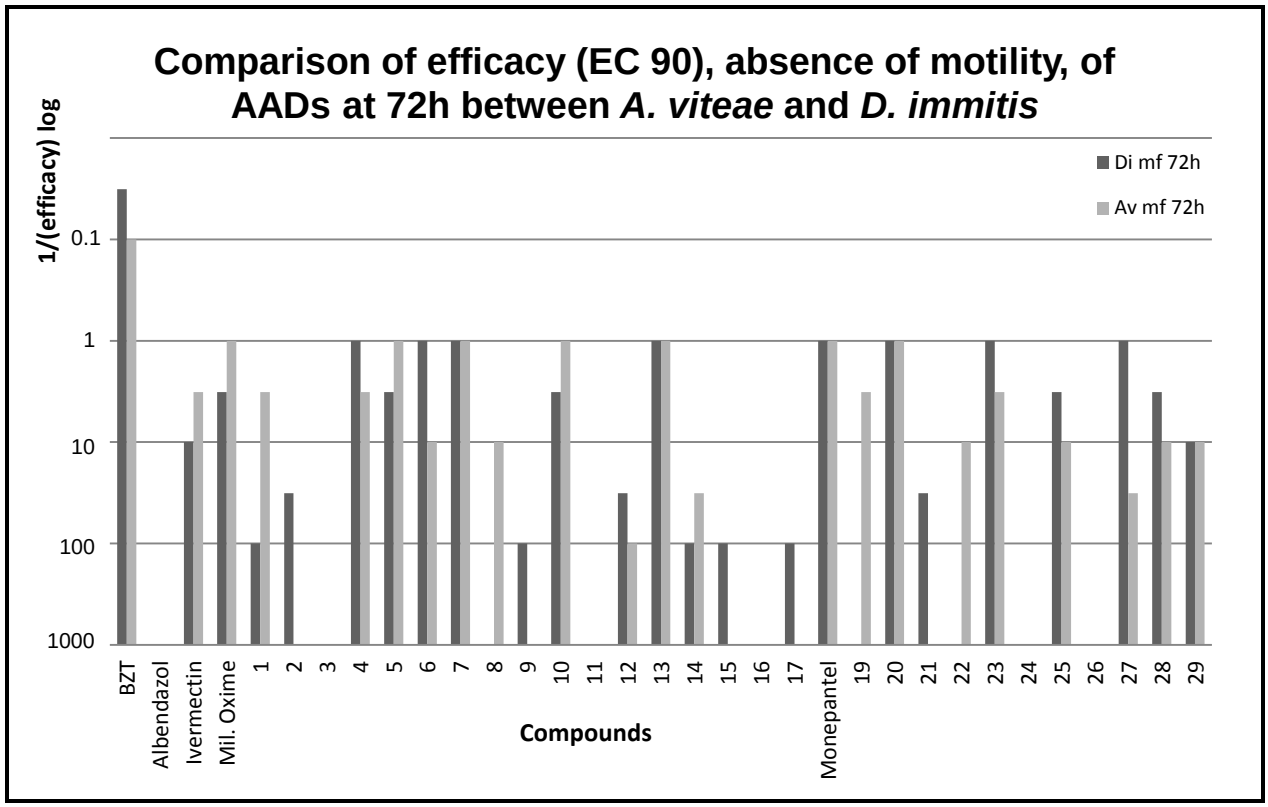


Figure 14. In vitro efficacies, EC 90, of AADs against *D. immitis* and *A. viteae* microfilariae; the tests were performed in triplicates. Results are represented in ppm (particle per million). Efficacy (absence of motility) is defined by comparison to positive control, milbemycin oxime, and a low concentration in ppm.

Figure 15 comparises the efficacies of in vitro microfilarial tests and LDA.

Overall microfilariae of *D. immitis* and *A. viteae* were less sensitive to AAD than the larvae of GI nematodes. However, the detection of compounds' efficacy is more specific if microfilarial test is used to select compounds for the filarial in vivo test.

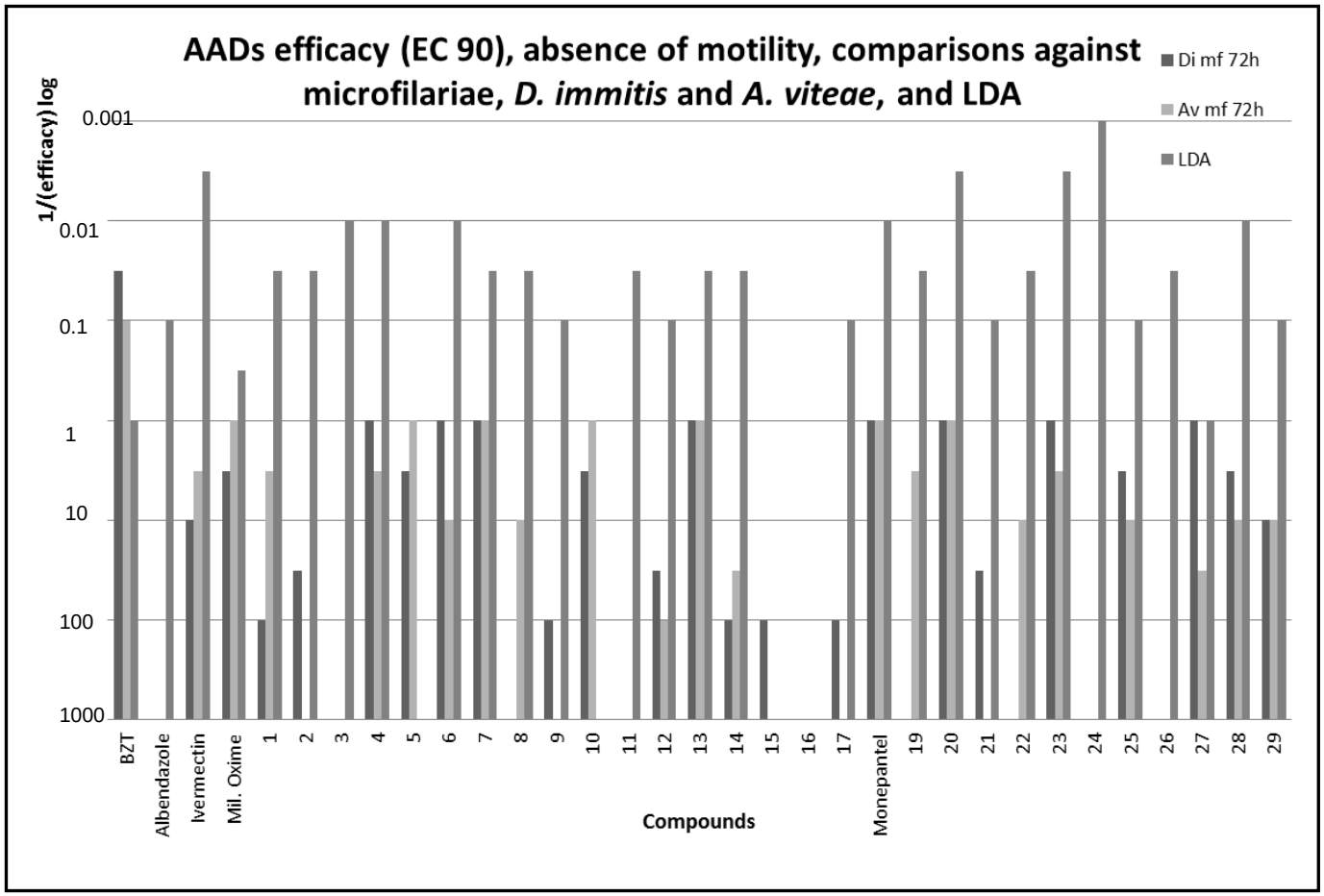


Figure 15. In vitro efficacies, EC 90, of AADs against gastro-intestinal nematodes in the LDA, from egg to L_3 were evaluated after seven days of incubation. In vitro efficacies of AADs were evaluated against *D. immitis* and *A. viteae* microfilariae after 72 hours of incubation all tests were performed in triplicates. Results are represented in ppm (particle per million). Efficacy (absence of motility) is defined by comparison to positive control, milbemycin oxime, and a low concentration in ppm.

In summary, AADs are active in vitro against microfilariae of *D. immitis* and of *A. viteae*. However, AADs so far did not show efficacy in vivo against *D. immitis* in dogs and *A. viteae* in gerbils (unpublished results).

These results raise further questions. Why are AAD active against microfilariae in vitro? Presumably because they express MPTL-1-like targets. Why are they not active in vivo? Maybe because such targets are only expressed in the microfilarial stage, or because of pharmacokinetic problems.

The appropriate method to get closer to the receptor is to learn more information about the genome of the heartworm. To follow this way, refer to the third chapter “The genome of the heartworm, *Dirofilaria immitis*”.

Chapter 3. The genome of the heartworm, *Dirofilaria immitis*

Publication 2

In review at FASEB Journal.

The genome of the heartworm, *Dirofilaria immitis*

Christelle Godel[1-3*], Sujai Kumar[4*], Georgios Koutsovoulos[4‡], Philipp Ludin[1,2‡], Daniel Nilsson[5], Nicola Wrobel[6], Marian Thompson[6], Christoph Schmid[1,2], Frédéric Bringaud[7], Adrian Wolstenholme[8], Claudio Bandi[9], Christian Epe[3], Ronald Kaminsky[3], Mark Blaxter[4], Pascal Mäser[1,2†]

[1]Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland

[2]University of Basel, Petersplatz 1, 4000 Basel, Switzerland

[3]Novartis Animal Health, Centre de Recherche Santé Animale, 1566 St. Aubin, Switzerland

[4]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

[5]Department of Molecular Medicine and Surgery, Science for Life Laboratory, Karolinska Institutet, Tomtebodavägen 23A, 17165 Solna, Sweden

[6]The GenePool Genomics Facility, School of Biological Sciences, University of Edinburgh, EH9 3JT, UK

[7]Centre de Résonance Magnétique des Systèmes Biologiques (RMSB), UMR 5536, University Bordeaux Segalen, CNRS, 146 rue Léo Saignat, 33076 Bordeaux, France

[8]Department of Infectious Diseases & Center for Tropical and Emerging Global Disease, University of Georgia, Athens, GA 30602, USA

[9]Dipartimento di Scienze Veterinarie, Università degli studi di Milano, Italy

[*]Contributed equally

[‡]Contributed equally

[†]Corresponding author

In this publication, I coordinated the genome consortium, leading telephone conferences, scientists from diverse institutions and the advancement of the whole project. Additionally, I performed the DNA extraction and the reciprocal blast searches bringing information about the existing drug targets of *D. immitis* as well as I generated an interesting table about new drug target candidates.

The genome of the heartworm, *Dirofilaria immitis*

Christelle Godel[1-3*], Sujai Kumar[4*], Georgios Koutsovoulos[4‡], Philipp Ludin[1,2‡], Daniel Nilsson[5], Nicola Wrobel[6], Marian Thompson[6], Christoph Schmid[1,2], Frédéric Bringaud[7], Adrian Wolstenholme[8], Claudio Bandi[9], Christian Epe[3], Ronald Kaminsky[3], Mark Blaxter[4], Pascal Mäser[1,2†]

[*]Contributed equally

[‡]Contributed equally

[1]Swiss Tropical and Public Health Institute, Socinstrasse 57, 4002 Basel, Switzerland

[2]University of Basel, Petersplatz 1, 4000 Basel, Switzerland

[3]Novartis Animal Health, Centre de Recherche Santé Animale, 1566 St. Aubin, Switzerland

[4]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

[5]Department of Molecular Medicine and Surgery, Science for Life Laboratory, Karolinska Institutet, Tomtebodavägen 23A, 17165 Solna, Sweden

[6]The GenePool Genomics Facility, School of Biological Sciences, University of Edinburgh, EH9 3JT, UK

[7]Centre de Résonance Magnétique des Systèmes Biologiques (RMSB), UMR 5536, University Bordeaux Segalen, CNRS, 146 rue Léo Saignat, 33076 Bordeaux, France

[8]Department of Infectious Diseases & Center for Tropical and Emerging Global Disease, University of Georgia, Athens, GA 30602, USA

[9]Dipartimento di Scienze Veterinarie, Università degli studi di Milano, Italy

[†]Corresponding author

Address	Swiss Tropical and Public Health Institute
	Socinstrasse 57, 4002 Basel, Switzerland
Tel.	+41 61 284 8338
Fax	+41 61 284 8101
E-mail	pascal.maeser@unibas.ch

Abstract

The heartworm *Dirofilaria immitis* is an important nematode parasite of dogs and other mammals, including humans. Transmitted by mosquitoes in warmer climatic zones, it is spreading across Southern Europe and the Americas at an alarming pace. There is no vaccine and chemotherapy is prone to complications. To learn more about this elusive parasite, we have sequenced the genomes of *D. immitis* and its endosymbiont *Wolbachia* using Illumina technology. In the 84.2 Mb of nuclear genomic assembly, we predict 10,179 protein coding genes, and 823 genes in the 0.9 Mb *Wolbachia* genome. The *D. immitis* genome harbors neither DNA transposons nor active retrotransposons, and there is extraordinarily little genetic variation between two sequenced isolates from Europe and the USA. Analysis of 990 orthologous genes resolves filarial nematode phylogeny, placing *D. immitis* next to *Onchocerca* species. Differential presence of anabolic pathways such as heme biosynthesis or purine *de novo* synthesis hints at the intricate relationship between the heartworm and *Wolbachia*. Comparing the *D. immitis* predicted proteome to other nematodes, and to mammalian hosts, identifies families of potential drug targets, immune modulators and vaccine candidates. The genome sequence will both support the development of new tools to combat dirofilariasis and aid efforts to control the related human filariases, which belong to the most neglected of all parasitoses.

Introduction

The heartworm *Dirofilaria immitis* (Leidy, 1856) is an onchocercid parasitic nematode of mammals that is transmitted by various mosquitoes including *Aedes*, *Anopheles* and *Culex* spp. The definitive host is the dog, yet it also infects cats, foxes, coyotes, and even humans (1). Dirofilariasis of dogs is a severe and potentially fatal disease. Adult nematodes of 20 to 30 cm reside in the pulmonary arteries, and the initial damage is to the lung. The spectrum of subsequent pathologies related to chronic heartworm infection is broad, the most serious manifestation being heart failure. Recent rapid spread of *D. immitis* through the US and southern Europe (2,3) is being favored by multiple factors. Global warming is expanding the activity season of vector mosquitoes, increasing their abundance and the likelihood of transmission of the parasite, and there are growing numbers of pets, reservoir animals, and 'traveling' dogs (2,3).

D. immitis' lifecycle is typical for Onchocercidae: microfilariae, shed into the bloodstream by adult females, are ingested by a mosquito where they develop into L_3 larvae and migrate to the labium. The prepatent period in the newly bitten dog is six to nine months, during which the injected larvae undergo two further molts and migrate via muscle fibers to the pulmonary vasculature, where the adult nematodes develop. Currently, diagnosis is based on the detection of circulating microfilariae or antigens from mature females, which is problematic in cats and humans. Treatment of dirofilariasis is problematic too, since the only adulticide approved by the FDA, the arsenical melarsomine dihydrochloride, can cause adverse neurological reactions. Treatment carries a significant risk of lethality, due to blockage of the pulmonary artery by dead nematodes. No vaccine is available. These issues, together with the alarming increasing spread of *D. immitis*, prompted the American Heartworm Society to recommend year-round chemoprophylactic treatment of all dogs (4), to kill the larval stages before they develop into adults. This requires monthly administration of anthelmintics, predominantly macrocyclic lactones such as ivermectin, milbemycin, or moxidectin.

D. immitis is closely related to human-pathogenic, onchocercid nematodes such as *Brugia malayi* and *Wuchereria bancrofti*, the etiological agents of lymphatic filariasis, and *Loa loa* and *Onchocerca volvulus*, which cause subcutaneous filariasis and river blindness. These disabling parasites are endemic in tropical and sub-tropical regions around the globe, with an estimated 380 million people affected (5). Improved diagnostics, new drugs, and, ultimately, effective vaccines are sorely needed. The sequencing of the *B. malayi* genome provided insights into the molecular mechanisms of immune evasion and a rational basis for drug design (6). Onchocercid nematodes are commonly infected with *Wolbachia*, an alphaproteobacterial endosymbiont that is essential for the fecundity, development, and survival of the nematodes, and this is an additional source of potential drug targets (7-10). However, the human-pathogenic Onchocercidae does not represent an attractive

market for the pharmacological industry, as projected income would not cover costs for drug development. The heartworm may hold a possible solution to this dilemma: the market for novel canine anthelmintics is big, and the drug targets are likely to be conserved in human-pathogenic species, given their high phylogenetic relatedness. Thus the *D. immitis* genome is of interest for veterinary as well as human medicine. Here we present the draft genome sequences of *D. immitis* and its *Wolbachia* endosymbiont (*w*Di), and use these data to clarify its phylogenetic position among the Onchocercidae, investigate the relationships between nematode and endosymbiont, and identify new drug and vaccine targets.

Results

Sequencing and genome assembly. Genomic DNA was isolated from single individuals of two *D. immitis* lines originally isolated from naturally infected dogs, one from Georgia (USA) and the other from Pavia (Italy). Several libraries of different insert sizes were constructed and sequenced on Illumina GA and HiSeq instruments. We also sequenced adult male and adult female transcriptomes from the Georgia nematodes using Illumina RNA-Seq. A total of 16 gigabase (Gb) of genomic sequence passed rigorous quality checks. As the *D. immitis* genome is likely to be the same size as related Onchocercidae (~95 mega bases [Mb]), this corresponds to ~170-fold coverage. We tested different assembly tools and parameters and chose ABySS (11) with a k-mer size of 35, as this performed best based on statistical and biological measures (see Table S2 for details). We filtered contigs mapping to the mitochondrial genome (12) and those with significant sequence similarity to *Wolbachia* genomes from *B. malayi* and arthropods. The final nuclear assembly, which was obtained after redundancy reduction with CD-Hit (13), contained 84.2 Mb of sequence in 31,291 contigs with an N50 of 10,584 bases (Table 1 and Table S1). We assembled a draft genome of the *Wolbachia* of *D. immitis* (*w*Di) by collecting all reads and their pairs mapping to *Wolbachia*-like contigs and reassembling them using tuned parameters in Velvet (14). The *w*Di genome assembled in 62 contigs spanning 902,182 bases with a contig N50 size of 23,251 bases.

Lack of genetic diversity between the sequenced *D. immitis* isolates. Even though the two *D. immitis* samples were isolated 8,000 km apart, on two different continents, the raw data assembled well together. We tested genetic diversity between the strains by mapping the reads from each strain back to the draft assembly, and found only 32,729 high quality variations, a very low per-nucleotide diversity rate of 0.04%. This hypodiversity is consistent with the hypothesis that the heartworm arrived in the New World with European immigrants (it was first described in the Americas 164 years ago (15)). Assuming one generation per year, the Italian and American isolates in our study would be a minimum of 328 generations apart, and the rate of genomic change $\approx 1.2 \times 10^{-9}$ per site per generation, comparable to the single-nucleotide mutation estimation rate of 3.5×10^{-9} per site per generation found in *Drosophila* (16).

A metazoan genome without active transposable elements. The *D. immitis* genome was surveyed for the three main classes of transposable elements (DNA transposons, LTR retrotransposons, and non-LTR retrotransposons). No traces of DNA transposons or non-LTR retrotransposons were found. However, a total of 376 fragments of BEL family PAO LTR retrotransposons (17) were identified. None

of these fragments were functional, as all contained frame shifts and stop codons in the coding sequence (Figure S1). These *D. immitis* PAO pseudogenes are most similar to PAO retrotransposons from *B. malayi* (6). Several of the *B. malayi* PAO retrotransposons are likely to be active, but *B. malayi* has a lower density of PAO elements and fragments overall (3.4 PAO per Mb, 8.3% of which are predicted to be active) compared to *D. immitis* (4.6 per Mb, none of which active). As far as we are aware, *D. immitis* is the first metazoan lacking active transposable elements.

Prediction of genes and proteins. Three parallel strategies were followed to predict and validate protein-coding genes from the assembled *D. immitis* contigs: (i) prediction with the *ab initio* gene finders SNAP (18) and Augustus (19) using a *B. malayi* training set; (ii) direct alignment to *B. malayi* proteins; and (iii) alignment to an assembly of the RNA-Seq data. The three lines of evidence were integrated by the Maker annotation pipeline (20), resulting in 11,375 gene models, 897 of which possessed alternate transcripts. The total number of predicted proteins of length 100 amino acids or more was 10,179, which is similar to the 9,807 predicted for *B. malayi* (Table 1). Based on matches to *D. immitis* expressed sequence tags and core eukaryotic gene completeness (21), the *D. immitis* proteome is near-complete. Gene density is similar in *D. immitis* and *B. malayi*: in both about 18% of the genome is protein-coding (Table 1). This contrasts with *C. elegans*, which has about twice as many genes across 30% of its genome. We used this proteome prediction for a series of further analyses.

Phylogeny of the Onchocercidae. The human-parasitic Onchocercidae have been grouped into two clades: *Onchocerca* species and the group *Brugia/Wuchereria/Loa*. To make best use of model species it is necessary to understand their relationships to the human targets. In addition to *D. immitis*, *Litomosoides sigmodontis*, a parasite of rodents, is increasingly used as a model for human disease (22). The relationships of these species to human-infective onchocercids has been poorly resolved (23,24). We assembled a dataset of 990 putatively orthologous genes using inParanoid (25) and QuickParanoid (http://pl.postech.ac.kr/QuickParanoid/) from the proteomes of *D. immitis*, *B. malayi*, *Onchocerca ochengi* (a cattle parasite closely related to human-infective *O. volvulus*), *Litomosoides sigmodontis*, *L. loa*, *W. bancrofti*, and, as an outgroup, the non-onchocercid spirurian *Ascaris suum* (26)(Figure S2). A Muscle-aligned (27) supermatrix was analyzed using RAxML (28), and a robustly supported tree derived (Figure 1). *D. immitis* is sister to *O. ochengi* , and the murine model *L. sigmodontis* is sister to the *Brugia* clade. *D. immitis* thus stands as a specific model for human and other onchocerciases.

The *Wolbachia* genome and its contribution to metabolism. The relationship between filarial nematode *Wolbachia* endosymbionts and their hosts has been the subject of intense scrutiny, as treatment of infected mammal hosts with tetracycline and other antibiotics results in clearance of the nematodes, suggesting a mutualistic symbiosis (23). The metabolic basis of this symbiosis remains unclear. It has been proposed that the *Wolbachia* endosymbiont of *B. malayi* (*w*Bm) provides with additional sources of critical metabolites such as heme and riboflavin (29). We annotated the *w*Di genome using the RAST server (30), and interrogated these annotations to examine the symbiont's biochemical capabilities. *C. elegans* and other nematodes (including *B. malayi*, and, on the basis of the genome sequence presented here, *D. immitis*) are deficient in heme synthesis but *w*Bm has an intact heme pathway (Figure S3) and a *ccmB* heme exporter suggesting it may support its host by providing heme. *w*Bm has a complete pathway from succinyl-CoA to heme (one apparently missing component, *hemG* may be substituted by a functional *hemY*). *w*Di lacks both *hemY* and *hemG* (and the recently described *hemJ* that can perform the same transformation). This step in the heme pathway is apparently absent in other bacteria, and so this may not indicate a non-functional heme synthesis pathway. Further anabolic pathways absent in *D. immitis* but present in *w*Di are purine and pyrimidine *de novo* synthesis. While *D. immitis* has but two enzymes, PRPP synthetase and adenylosuccinate lyase, *w*Di has the full complement of 10 enzymes required to synthesize inosine monophosphate (Figure S3).

*w*Bm is deficient in folate synthesis, as it lacks dihydrofolate reductase and dihydroneopterin aldolase. *w*Di has both these genes, suggesting it can metabolize 7,8-dihydrofolate to folate and 5,6,7,8-tetrahydrofolate, and utilize dihydroneopterin as an input to folate metabolism. *Wolbachia* *w*Mel from *Drosophila melanogaster* has both these enzymes, and they are variably present in other alphaproteobacteria. Whether this pathway contributes to the nematode symbiosis is unclear, but it does highlight another component of *Wolbachia* metabolism that may be accessible to drug development. Further *w*Di gene products that might be exploited as drug targets include nucleic acid synthesis and cell division proteins such as FtsZ and DnaB, the fatty acid synthesis enzymes *FabZ* and *AcpS*, components of the *Sec* protein secretion system, and, possibly, the peptidoglycan synthesis enzymes of the *Mur* operon. All these are unique proteins in *w*Di, do not have counterparts in mammals, and are being developed as antibiotic drug targets for bacterial infections such as tuberculosis.

The *w*Bm genome is rich in repeats, many associated with WO bacteriophage. We identified only one region of the *w*Di genome that has similarity to WO, spanning a widely conserved hypothetical protein of unknown function. *w*Di appears to be devoid of bacteriophage. As phage genes comprise 14% of the *w*Bm genome (and 134 ORFs) this may explain the smaller genome size of

*w*Di. Genes with ankyrin repeats have been associated with parasitic phenotypes in *Wolbachia* of insects (31), and ankyrin repeat genes are found in phage segments. *w*Di also has 3 fewer ankyrin repeat containing proteins than *w*Bm.

Horizontal gene transfer from *Wolbachia* to the nuclear genome is common in hosts harboring this endosymbiont (29), and it has been proposed that these transfers may confer new functionality to the nuclear genome, though this is unlikely (32). We identified 775 elements (spanning 170,324 bp) of over 100 bases with 80% identity or greater to *w*Di, but only 5 of these elements matched more than 90% of the length of a *w*Di ORF and were not interrupted by frame-shifting indels or stop codons. Only one of these putative lateral gene transfers had a match to a known *Wolbachia* protein (tRNA pseudouridine synthase) but we found no evidence of transcription of this gene in our RNA-Seq data set and thus conclude that although elements from *w*Di have been inserted into the nuclear genome, as observed in other intracellular symbioses, none of them are likely to produce functional proteins. Frequency patterns of short palindromes can serve as a specific finger-print for DNA (33). We compared the presumed nuclear insertions, *D. immitis* genomic sequence, and *w*Di genomic sequence based on the actual occurrence of palindromes of length 4 in relation to their expected occurrence given the GC-content (Table 1) of the target sequences. The suspected DNA fragments clearly clustered with the Wolbachia genome (Figure S4), providing further evidence that they had been acquired by the nucleus through horizontal transfer.

Existing drug targets of *D. immitis*. Many of the molecular targets of currently used anthelmintics were identified by forward genetics in the model nematode *C. elegans* (Table 2) (34): the macrocyclic lactones (ivermectin), benzimidazoles (albendazole), imidazothiazoles (levamisole), cyclodepsipeptides (emodepside), and the aminoacetonitrile derivatives (monepantel). Most prominent are neural membrane proteins, highlighting the importance of the neuromuscular junction as a hotspot of anthelmintic drug action. Two superfamilies of membrane proteins are among the prime targets of known pharmacophores and natural toxins: the G protein-coupled receptors (GPCR) and the ligand-gated ion channels (LGIC). LGIC are ionotropic receptors of acetylcholine, GABA, glutamate, or other neurotransmitters. GPCR are coupled to heterotrimeric G proteins and respond to a large variety of odors, pheromones, hormones, or neurotransmitters. The *C. elegans* genome encodes more members of these two families than any other animal. This is illustrated by the frequency distribution of predicted transmembrane domains (TM) per protein (Figure 2), where peaks of four and seven TM indicate LGICs and GPCRs, respectively. Such peaks are absent in the TM frequency distribution of *D. immitis* proteins (Figure 2). Another pharmacologically interesting family of transmembrane proteins apparently underrepresented in *D. immitis* compared

to *C. elegans* is the twelve TM domain ABC transporter family. ABC transporters carry out a multitude of active transport processes including drug efflux.

However, some proven drug targets appear to be missing from *D. immitis*. There are no members of the DEG-3 subfamily of acetylcholine receptors, which contains the presumed targets of monepantel (35). This contrasts with *B. malayi*, which has DEG-3 and DES-2 orthologues. The target space of levamisole and ivermectin in *C. elegans* comprises a large number of different LGIC, some of which do not have an orthologue in *D. immitis* (Table 2), indicating that those present are sufficient to confer drug susceptibility. Of the drugs listed in Table 2, flubendazole (36), mebendazole (37), levamisole (38), ivermectin, milbemycin, moxidectin, and selamectin (39,40) have been demonstrated to be active against *D. immitis*.

New drug target candidates in *D. immitis*. New potential drug targets were identified *in silico* through an exclusion pipeline following previous schemes for anthelmintic drug discovery (41,42). Starting from the complete set of predicted *D. immitis* proteins, we applied the following inclusion criteria for candidate drug targets: (i) absence of an orthologue in the dog or human proteome, (ii) presence of a *C. elegans* orthologue that is essential for survival or development (based on RNAi phenotypes), (iii) absence of additional *D. immitis* paralogues and (iv) predicted function as an enzyme or receptor. Among the 20 identified candidates (Table 3) are several proven drug targets, such as RNA-dependent RNA polymerase (antiviral), apurinic/apyrimidinic endonuclease and hedgehog proteins (anticancer), UDP-galactopyranose mutase (antimycobacterial), sterol-C24-methyltransferase (antifungal), and the insecticide targets chitin synthase. The identified *D. immitis* orthologues of these enzymes might serve as the starting points for the development of new anthelmintics.

Immune modulators and vaccine candidates. Filarial nematodes modulate the immune systems of their hosts in complex ways that result in an apparently intact immune system that ignores a large parasite residing, sometimes for decades, in tissues or the bloodstream. They also require intact immune systems to develop properly (43). Often, immune responses result in pathology for the host in addition to parasite clearance, and *Wolbachia* may exacerbate these responses (44). Strategies for development of a vaccine against filariases thus depend on delivering the correct antigens to the right arm of the immune system, avoiding induction of dangerous responses, and deflecting or stopping immune suppression by the parasite. We surveyed the *D. immitis* genome for molecules currently proposed as vaccine candidates in other onchocercids (45,46), and identified homologues for all fourteen classes of molecules (Table 4).

The mechanisms of immune modulation by onchocercid nematodes remain enigmatic, but involvement of proteases and protease inhibitors (such as leucyl aminopeptidase (LAP), serpins and cyctatins), homologues of ancient immune system signalling molecules (such as transforming growth factor beta [TGFb] and macrophage migration inhibition factor [MIF]), and possible mimics of immune molecules have all been proposed. In *D. immitis*, we identified LAP, three cystatins, many serpins, two MIF genes, orthologous to the MIF-1 and MIF-2 genes of *B. malayi* and *O. volvulus*, and four TGFb homologues (Table 4). We also identified a homologue of suppressor of cytokine signalling 5 (SOCS5). SOCS proteins consist of a central SH2 domain, a C-terminal SOCS box, and a variable N-terminal domain. SOCS5 negatively regulates the JAK/STAT pathway and inhibits the IL-4 pathway in T helper cells, promoting TH1 differentiation (47). The SH2 domain recognizes target proteins and the SOCS box recruits an ubiquitin complex that mediates proteosomal degradation of the target. SOCS5 homologues can be identified in animal-parasitic nematodes, including *B. malayi*, *D. immitis*, *L. loa*, *A. suum* and *Trichinella spiralis*, but SOCS5 is absent from the free-living *C. elegans*, the necromenic *Pristionchus pacificus*, and the plant parasitic *Meloidogyne* spp. Several viruses induce host SOCS protein expression for immune evasion and survival (48). *D. immitis* and other filarial nematodes (49) may use SOCS5 homologues to mimic host SOCS5. We also identified a homologue of IL16, a PDZ-domain containing, pleiotropic cytokine (50). In mammals IL16 acts via the CD4 receptor to modulate the activity of a wide range of immune effector cells, including T-cells and dendritic cells (51). Again, this molecule is only present in parasitic nematodes (including *A. suum* (26)) and absent from free-living and plant-parasitic species' genomes. We suggest that these molecules, and perhaps other mimics of cytokines and modulators, are the effector toolkit used by filarial nematodes to build an immunologically compromised niche.

In parasitic nematodes, secreted proteins have been found to be more rapidly evolving than other classes (52), and are thus likely to harbor novelties associated with host manipulation and invasion. The *D. immitis* proteome was divided into transmembrane (n = 1416; presence of TM domains (53)), secreted (638; N-terminal export signals (53)), GPI anchored (151; C-terminal GPI anchor attachment signals (54)) and cytosolic (8133; all remaining), sets. Each protein was aligned to its best match from *B. malayi* and mean alignment scores compared between sets. Secreted and GPI anchored proteins had significantly lower alignment scores ($p<0.01$; 1-way Anova followed by Kruskal-Wallis' post test) than the cytosolic or transmembrane proteins (Figure S2), indicating that the divergence between the two parasites is more pronounced in proteins potentially exposed to the host.

Discussion

The *D. immitis* genome sequence described here is only the second to be determined for an onchocercid nematode, despite the social and economic importance of these parasites. We used solely short-read Illumina technology, and the high throughput of this platform allowed us to use stringent quality-filtering and assembly metrics to derive a good draft genome. Three genomes were co-sequenced: the mitochondrial (at about 4000-fold read coverage; this was determined previously (12)), the *Wolbachia* symbiont *w*Di (about 1000-fold coverage of the 0.9 Mb genome) and the nuclear genome (about 150-fold coverage). After redundancy reduction the total size of the nuclear assembly was 84.2 Mb, slightly smaller than the 93.6 Mb published for *B. malayi* (6). Two peculiarities of the assembled *D. immitis* genome are striking: the apparent genetic homogeneity and the lack of active transposable elements. The lack of heterogeneity was convenient in that it allowed us to pool data obtained from two different *D. immitis* isolates, one from Pavia, Italy, and the other from Georgia, USA. Polymorphisms called from the independent sequencing of the two isolates yielded a per-nucleotide diversity of 0.04%. Startling as this hypovariability may be, it is in agreement with microsatellite population genetics (55) and with the hypothesis that the heartworm arrived in the New World only recently with European immigrants (15). The first report of dirofilariasis in the US dates from 1847, as opposed to a 1626 observation from Italy. Both our nematodes fall within the single USA population defined by microsatellite analyses (55). The lack of genetic diversity in the nuclear genome will make identification of mutations conferring drug resistance much easier. The lack of DNA transposons and active retrotransposons in *D. immitis* is a strong negative result since potentially active elements are relatively easy to identify (they are present in multiple, highly similar copies). To our knowledge, this is the first metazoan genome devoid of active transposable elements.

The *Wolbachia* *w*Di genome, with 823 predicted proteins, complements the *D. immitis* nuclear genome in that it encodes enzymes for anabolic pathways which are missing in the latter, e.g. heme biosynthesis and purine de novo synthesis (Figure S3). In contrast to *wBm*, wDi also carries the genes for folate synthesis, suggesting that folate too might be supplied by the endosymbiont. However, folate, purines, or heme could also be taken up from the mammalian or insect host, and so it remains to be shown whether such metabolites are actually delivered from *w*Di to *D. immitis*. In any case the endosymbiont, being essential for proliferation of *D. immitis*, represents a target for control of the heartworm. Screening the predicted *w*Di proteome returned expected antibiotic drug targets such as Fts and Sec proteins, but also the products of the Mur operon required for peptidoglycan synthesis. The role of peptidoglycan in the obligate intracellular *Wolbachia* remains to be elucidated.

All the anthelmintics used in human medicine were originally developed for the veterinary

sector. We have pursued two approaches to identify potential drug targets in *D. immitis*: top-down, starting with the known anthelmintic targets of *C. elegans* (Table 2), and bottom-up, narrowing down the predicted *D. immitis* proteome to a list of essential, unique, and druggable targets (Table 3). While the majority of the current anthelmintics activate their target (thereby interfering with synaptic signal transduction), the second approach aimed to identify inhibitable targets. The applied criteria – presence of an essential orthologue in *C. elegans*, absence of any significantly similar protein in human or dog, and absence of paralogues in *D. immitis* – are obviously also excluding suitable targets such as novel and filaria-specific proteins (see Figure 1). Our goal was to end up with a manageable, rather than complete, list of unique *D. immitis* proteins that are likely to be essential and druggable. Some of the candidates identified are worth further investigation based on their presumed role in signal transduction, e.g. the nematode-specific GPCRs or hedgehog proteins (Table 3). Others have already been validated as drug targets in other systems: sterol-C24-methyltransferase (EC 2.1.1.41) is a target of sinefungin, chitin synthase (2.4.1.16) is the target of the insecticide lufenuron, and the mannosyltransferase bre-3 is required for interaction of *Bacillus thuringiensis* toxin with intestinal cells (56). The discovery of new *D. immitis* drug targets would be timely since resistance to macrocyclic lactones has recently been reported from the Southern USA (57,58).

Two common properties of filarid parasites are (i) a fascinating biology involving immune evasion, arthropod vectors and *Wolbachia* endosymbionts, and (ii) a pressing need for new drugs, improved diagnostics and – ideally – vaccines. We hope that the genome sequence of *D. immitis* presented here will contribute to an increased understanding and to new leads for control and eradication.

Materials and Methods

***Dirofilaria immitis* isolates and maintenance.** The *D. immitis* Pavia isolate is an inbred strain originally isolated in a veterinary practice in northern Italy. Adult nematodes were recovered after necropsy of infected dogs (permit number FR401e/08 from the Veterinary Office Canton de Fribourg) that were the control group of on-going investigations and not sacrificed for the sake of the present study. The nematodes were extracted from the dogs' right pulmonary artery and from the heart. Adult heartworms were immediately sexed, counted, measured, and washed in Hank's Balanced Salt Solution (Gibco) supplemented with streptomycin/penicillin (Invitrogen) and fungizone/amphotericin (Invitrogen). Individual nematodes were snap-frozen in liquid nitrogen and cryopreserved at -80°C.

Isolation of nucleic acids and sample preparation. To extract genomic DNA from individual heartworms, segments of approximately 2 cm were ground under liquid nitrogen with a micropestle, incubated at 56°C for three hours in lysis buffer supplemented with 100 mg/ml RNase A (QIAamp DNA extraction kit, Qiagen), and further processed following the supplier's protocol. The DNA was washed three times before elution. DNA quality and quantity were assessed by electrophoresis in 1.5% agarose gels (agarose LE analytic grade, Promega) supplemented with Sybr safe DNA stain (Invitrogen), and by measuring absorbance (Nanodrop 2000c, Thermo Scientific) or fluorescence (Qubit, Invitrogen). RNA was extracted from the nematodes using the protocol described in the Qiagen miRNeasy Mini Handbook October 2007. Poly-A containing mRNA molecules were purified using poly-T oligo-attached magnetic beads. Following purification, the mRNA was fragmented into small pieces using divalent cations under elevated temperature. Cleaved RNA fragments were then copied into first strand cDNA using reverse transcriptase and a high concentration of random hexamer primers. This was followed by second strand cDNA synthesis using DNA Polymerase I and RNaseH.

Sequence generation and assembly. Whole genome shotgun sequences were generated at the GenePool (University of Edinburgh) and at Fasteris SA (Geneva, Switzerland) using Illumina GA-IIx and HiSeq 2000 instruments. In addition to standard short-insert (100-400 bp) paired end libraries, long-insert (3-4 kb) libraries were also attempted and used as described below. Sequences were basecalled using the Illumina CASAVA pipeline. Four libraries (Table S1) were used in the final assembly totalling 28 Gb in 295 M reads, and have been deposited in the EBI Short Read Archive with accession ID ERA032353. We trimmed low quality reads aggressively as we estimated we had more than 200 fold coverage. Low quality bases were trimmed from the 3' end of all reads if they were

below a Phred quality equivalent of 20 using FASTQ Quality Trimmer. Reads containing uncalled bases ('N's) and reads shorter than 35 bases after trimming were discarded along with their pairs. The long-insert mate-pair reads were used as single-ended sequences in the release candidate assembly because the proportion of paired-end contamination was deemed excessive. In addition, the reads in this library were truncated at 50 bases to minimise the likelihood of chimeric joins as a result of reads spanning the junction point of the circularised long-insert fragments, leaving a total of 16 Gb in 271 M reads across all 4 libraries. The reads were assembled into contiguous sequences and scaffolded using ABySS after evaluating a range of values for the k-mer size (k) and the minimum number of pairs connecting 2 contigs (n). We also tested Velvet and CLC-bio with a range of parameters but chose ABySS with k=35 and n=4 based on primary observations such as N50 value, maximum contig length, total number of bases in contigs, total number of Ns in contigs (all evaluated for a minimum contig length of 200 bp and 1000 bp), as well as a suite of biological completeness tests described below (Table S2).

***D. immitis* mitochondrion.** All contigs in the best ABySS assembly were screened for mitochondrial contigs using the published *D. immitis* mitochondrial genome.

***Wolbachia* genome assembly.** We screened our best ABySS assembly for known *Wolbachia* DNA using BLAST against complete *Wolbachia* genomes from EMBL/GenBank/DDBJ. To obtain a more accurate assembly of the *Wolbachia* of *D. immitis* (wDi) genome, we extracted all the read pairs where at least one read mapped to any likely *Wolbachia* contig, and reassembled these 6,912,659 read pairs using more accurate k-mer and coverage parameters with Velvet version 1.1.03 (k-mer size 45, expected k-mer coverage 400, k-mer coverage cutoff 150). The assembly was checked for numerical and biological optimality as outlined above. Although not a finished genome, we believe this draft is essentially complete because a MEGABLAST search of 107 existing genome survey sequences (GSS) from wDi on GenBank (Accession numbers ET041559 - ET041665) found 106 of them in our wDi assembly at an e-value of 1e-50.

Assembly evaluation and final assembly freeze. For each assembly generated using different programs and varying parameters, we tallied a series of contig size metrics as well as biological metrics such as alignments to known EST sequences, genes predicted, and CEGMA completeness. 4005 *D. immitis* EST sequences were obtained from NCBI dbEST and assembled into 1810 contigs

using PHRAP. BLASTn was used to determine how many of these longer transcripts were found in the candidate assembly. A transcript was deemed reasonably complete if 80% of its length was found in BLASTn matches and the number of fragmented (i.e. on different contigs) and unfragmented hits were counted. Augustus was used to predict the number of genes and total exon length for each assembly using the HMM for *Brugia malayi*. Table S2 shows the results of these metrics, and the ABySS assembly with k=35, n=4 was chosen for further processing. To obtain the *D. immitis* nuclear genome, we screened the contig set for *D. immitis* mitochondrial sequence (GenBank ID AJ537512) and for *Wolbachia* sequences (GenBank IDs AM999887, CP001391, AE017321, and AE017196, and our re-assembled *Wolbachia* genome). We also reduced contig redundancy by using the program CD-HIT-EST to cluster those contigs that were at least 97% identical. This reduced contig set was frozen for all further analyses in this paper.

Comparing *Wolbachia* of *D. immitis* and *B. malayi*. Comparing the wDi genome to that of wBm using PROmer revealed significant breakage of synteny between the two genomes. Repeats in the *Wolbachia* of *D. immitis* (wDi) and *Wolbachia* of *B. malayi* (wBm) genomes were determined using self against self NUCMER (with –maxmatch parameter). The wDi draft genome was compared to the *Wolbachia* of *B. malayi* (wBm), the most closely related sequenced *Wolbachia* genome sequence available (GenBank ID: NC_006833). The wDi genome was as much as 16% smaller than the 1,080,084 bp wBm genome. One reason for the smaller genome size is that wDi had been assembled using Velvet, a de-Bruijn graph assembler that collapsed repeated regions into single contigs. As a result, only one 89 bp repeat was found in the wDi assembly compared to 49 repeats in wBm ranging in size from 226 to 3082 bp.

RNAseq assembly. We created a transcriptional profile of *D. immitis* by generating RNAseq libraries from an adult male and an adult female, and sequenced a total of 11,019,886 (male) and 21,643,293 (female) read pairs of length 54 bases each on the Illumina GAIIx platform (ArrayExpress accession: E-MTAB-714; ENA study accession: ERP000758). The ends of these reads were trimmed if they were low quality (below an equivalent Phred score of 20) and the resulting 31,396,183 read pairs were assembled using Trans-ABySS with k-mer values from 23 to 47 in steps of 4. We also tried using Oases to assemble the transcriptome, and although the resulting transcripts were slightly longer, they had lower CEGMA completeness values and aligned less well to the set of 1810 *D. immitis* EST contigs and *B. malayi* peptides that we used in our genome assembly evaluation (Table S3).

Gene prediction and annotation. The MAKER annotation pipeline was used for a first pass automatic annotation of the assembly freeze. The pipeline first masked repeats using RepeatMasker 3.2.9, searching for all "*Nematoda*" repeats, with RepBase libraries for RepeatMasker version 20110419. MAKER predicted genes using evidence from several sources: (a) EST alignments from *D. immitis* (RNAseq assembly); (b) Protein alignments from *Brugia malayi* (WormBase release WS220); (c) Gene predictions from the ab-initio gene finders SNAP and Augustus, using the shipped HMM profiles for *B. malayi*. MAKER predicted 11,895 gene models that resulted in a total of 12,872 transcripts and peptides. Selected pathways were annotated with HMMer. For this purpose, all entries for a given E.C. number were downloaded from the manually curated section (SwissProt) of UniProt, redundancy reduced over a threshold of 90% identity, and aligned with ClustalW. Profiles were generated and searched against the *D. immitis* predicted proteome with HMMer 3.0.

Prediction of TM domains and membrane proteins. Transmembrane domains were predicted with Phobius. Ligand-gated ion channels were predicted with PFAM profiles PF02931 (neurotransmitter-gated ion-channel ligand binding domain) and PF02932 (neurotransmitter-gated ion-channel transmembrane region), considering only proteins that scored for both profiles. Similarly, proteins were annotated as ABC transporters if they scored for both PF00664 (ABC transporter transmembrane region) and PF00005 (ABC transporter, ATP-binding cassette). Putative G protein-coupled receptors were identified with a HMM library consisting of PF00001, PF00002, PF00003, PF02949, PF08395, PF10292, PF02117, PF10316, PF10320, PF10322, PF10323, PF10324, PF10325, PF10326, and PF02116.

Phylogenetics. Proteins for *Ascaris suum* were downloaded from EMBL-EBI (http://www.ebi.ac.uk/), transcripts for *Loa loa* and *Wuchereria bancrofti* from the Broad Institute (http://www.broadinstitute.org/), and the transcriptome of *Brugia malayi* (version WS220) from Wormbase (ftp://ftp.wormbase.org/pub/wormbase/). Protein predictions for *Litomosoides sigmodontis* and *Onchocerca ochengi* were obtained from genome sequence projects at the Blaxter Lab. Finally, the protein predictions for *Dirofilaria immitis* were added to create a dataset of 7 species and 103,849 proteins. Orthology groups were constructed with inParanoid and QuickParanoid. 990 clusters that shared one gene per species were identified, and the amino acid sequences for each cluster were aligned using Muscle. The nucleotide sequences were aligned based on the corresponding protein alignments with the program *tranalign* of the EMBOSS package. Maximum likelihood analysis was performed with RAxML 7.2.8, with a GTR+ (generalised time-reversible)

model, where base frequencies and gamma shape distribution were estimated from the data. The support values generated from RAxML using identical settings for 100 bootstrap replicates were 100 for all branches.

Summary of software used. The following programs were used (or at least tested) for genome assembly, analysis, and annotation.

Program / Resource	Reference
ABySS 1.2.0	Robertson et al. (2010) Nat Methods 7, 909-912
Augustus 2.5.5	Stanke & Morgenstern (2005) Nuc Acids Res 33, W465-467
BLAST	Altschul et al. (1990) J Mol Biol 215, 403-410
CASAVA 1.5	www.illumina.com/software/genome_analyzer_software.ilmn#casava
CD-HIT	Huang et al. (2010) Bioinformatics 26, 680-682
CEGMA	Parra et al. (2007) Bioinformatics 23, 1061-1067
ClustalW	Thompson et al. (1994) Nuc Acids Res 22, 4673-4680
EMBOSS 6.3.1	Rice et al. (2000) Trends Genet 16: 276-277
FASTQ Quality Trimmer	hannonlab.cshl.edu/fastx_toolkit/
HMMer 3.0	Eddy SR (1995) Proc Int Conf Intell Syst Mol Biol 3, 114-120
MAKER 2.08	Cantarel et al. (2008) Genome Res 18, 188-196
Muscle 3.8.31	Edgar (2004) Nuc Acids Res 32: 1792-1797
NUCMER	Kurtz et al. (2004) Genome Biol 5: R12
Phobius	Käll et al. (2004) J Mol Biol 338: 1027-36
QuickParanoid	Ostlund et al. (2010) Nuc Acids Res 38, D196-203
RAxML 7.2.8	Stamatakis et al. (2005) Bioinformatics 21, 456-463
RepeatMasker 3.2.9	Chen N (2004) Curr Prot Bioinf, Chapter 4, Unit 4 10
RepBase	Kapitonov & Jurka (2008) Nature Rev Gen 9, 411-412; author reply 414
SNAP	Korf I (2004) BMC Bioinf 5, 59
Velvet 1.1.0.3	Zerbino & Birney (2008) Genome Res 18: 821-829

Acknowledgments

We thank Claudio and Marco Genchi for providing *D. immitis* from Pavia and Andrew Moorhead for the *D. immitis* from Athens, GA. We are very grateful for prepublication access to filarial nematode genomic data. Unpublished *Loa loa* and *Wuchereria bancrofti* data were sourced from the Filarial Worms Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/). Unpublished *Litomosoides sigmodontis* and *Onchocerca ochengi* preliminary genomic data were generated as part of the European collaboration Enhanced Protective Immunity Against Filariasis (EPIAF), a Focused research project (SICA) of the EU (award 242131), and we thank collaborators J. Allen, S. Babayan, D. Taylor and B. Makepeace for access to these. *D. immitis* genomic sequencing was performed by Fasteris SA (Geneva) and The GenePool Genomics Facility (Edinburgh).

References

1. Lee AC, Montgomery SP, Theis JH, Blagburn BL, Eberhard ML (2010) Public health issues concerning the widespread distribution of canine heartworm disease. Trends Parasitol 26: 168-173.

2. Genchi C, Rinaldi L, Mortarino M, Genchi M, Cringoli G (2009) Climate and Dirofilaria infection in Europe. Vet Parasitol 163: 286-292.

3. Traversa D, Di Cesare A, Conboy G (2010) Canine and feline cardiopulmonary parasitic nematodes in Europe: emerging and underestimated. Parasit Vectors 3: 62.

4. Anonymous (2010) Diagnosis, prevention, and management of heartworm (Dirofilaria immitis) infection in dogs. American Heartworm Society.

5. Anonymous (2008) Progress report on Global programme to eliminate lymphatic filariasis and Conclusions of the meeting of the Technical Advisory Group on the Global Elimination of Lymphatic Filariasis. Weekly Epidemiological Record 83: 333-348.

6. Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, et al. (2007) Draft genome of the filarial nematode parasite Brugia malayi. Science 317: 1756-1760.

7. Slatko BE, Taylor MJ, Foster JM (2010) The Wolbachia endosymbiont as an anti-filarial nematode target. Symbiosis 51: 55-65.

8. Hoerauf A, Nissen-Pahle K, Schmetz C, Henkle-Duhrsen K, Blaxter ML, et al. (1999) Tetracycline therapy targets intracellular bacteria in the filarial nematode *Litomosoides sigmodontis* and results in filarial infertility. J Clin Invest 103: 11-18.

9. Bandi C, McCall JW, Genchi C, Corona S, Venco L, et al. (1999) Effects of tetracycline on the filarial worms Brugia pahangi and Dirofilaria immitis and their bacterial endosymbionts Wolbachia. Int J Parasitol 29: 357-364.

10. Bazzocchi C, Mortarino M, Grandi G, Kramer LH, Genchi C, et al. (2008) Combined ivermectin and doxycycline treatment has microfilaricidal and adulticidal activity against Dirofilaria immitis in experimentally infected dogs. Int J Parasitol 38: 1401-1410.

11. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, et al. (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19: 1117-1123.

12. Hu M, Gasser RB, Abs El-Osta YG, Chilton NB (2003) Structure and organization of the mitochondrial genome of the canine heartworm, Dirofilaria immitis. Parasitology 127: 37-51.

13. Huang Y, Niu B, Gao Y, Fu L, Li W (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics 26: 680-682.

14. Zerbino DR (2010) Using the Velvet de novo assembler for short-read sequencing technologies. Current protocols in bioinformatics Chapter 11: Unit 11.15.

15. Bowman DD, Atkins CE (2009) Heartworm biology, treatment, and control. Vet Clin North Am Small Anim Pract 39: 1127-1158, vii.

16. Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, et al. (2009) Analysis of the genome sequences of three Drosophila melanogaster spontaneous mutation accumulation lines. Genome Res 19: 1195-1201.

17. Eickbush TH, Malik HS (2002) Origins and evolution of retrotransposons. In: Craig AG, Craigie R, Gellert M, Lambowitz AM, editors. Mobile DNA II. Washington, D.C.: ASM Press.

18. Korf I (2004) Gene finding in novel genomes. BMC bioinformatics 5: 59.

19. Stanke M, Morgenstern B (2005) AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res 33: W465-467.

20. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, et al. (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18: 188-196.

21. Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23: 1061-1067.

22. Allen JE, Daub J, Guiliano D, McDonnell A, Lizotte-Waniewski M, et al. (2000) Analysis of genes expressed at the infective larval stage validates utility of Litomosoides sigmodontis as a murine model for filarial vaccine development. Infection and immunity 68: 5454-5458.

23. Fenn K, Blaxter M (2004) Are filarial nematode Wolbachia obligate mutualist symbionts? Trends Ecol Evol 19: 163-166.

24. Casiraghi M, Bain O, Guerrero R, Martin C, Pocacqua V, et al. (2004) Mapping the presence of Wolbachia pipientis on the phylogeny of filarial nematodes: evidence for symbiont loss during evolution. Int J Parasitol 34: 191-203.

25. Ostlund G, Schmitt T, Forslund K, Kostler T, Messina DN, et al. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. Nucleic Acids Res 38: D196-203.

26. Wang J, Czech B, Crunk A, Wallace A, Mitreva M, et al. (2011) Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. Genome Res 21: 1462-1477.

27. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics 5: 113.

28. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688-2690.

29. Fenn K, Blaxter M (2006) Wolbachia genomes: revealing the biology of parasitism and mutualism. Trends in parasitology 22: 60-65.

30. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, et al. (2008) The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9: 75.

31. Sinkins SP, Walker T, Lynd AR, Steven AR, Makepeace BL, et al. (2005) Wolbachia variability and host effects on crossing type in Culex mosquitoes. Nature 436: 257-260.

32. Blaxter M (2007) Symbiont genes in host genomes: fragments with a future? Cell Host Microbe 2: 211-213.

33. Lamprea-Burgunder E, Ludin P, Maser P (2010) Species-specific typing of DNA based on palindrome frequency patterns. DNA Res 18: 117-124.

34. Sangster NC, Song J, Demeler J (2005) Resistance as a tool for discovering and understanding targets in parasite neuromusculature. Parasitology 131 Suppl: S179-190.

35. Kaminsky R, Ducray P, Jung M, Clover R, Rufener L, et al. (2008) A new class of anthelmintics effective against drug-resistant nematodes. Nature 452: 176-180.

36. Guerrero J, Campbell Seibert BP, Newcomb KM, Michael BF, McCall JW (1983) Activity of flubendazole against developing stages of Dirofilaria immitis in dogs. Am J Vet Res 44: 2405-2406.

37. McCall JW, Crouthamel HH (1976) Prophylactic activity of mebendazole against Dirofilaria immitis in dogs. J Parasitol 62: 844-845.

38. Carlisle CH, Atwell RB, Robinson S (1984) The effectiveness of levamisole hydrochloride against the microfilaria of Dirofilaria immitis. Aust Vet J 61: 282-284.

39. Campbell WC (1982) Efficacy of the avermectins against filarial parasites: a short review. Vet Res Commun 5: 251-262.

40. McCall JW (2005) The safety-net story about macrocyclic lactone heartworm preventives: a review, an update, and recommendations. Vet Parasitol 133: 197-206.

41. Doyle MA, Gasser RB, Woodcroft BJ, Hall RS, Ralph SA (2010) Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. BMC Genomics 11: 222.

42. Holman AG, Davis PJ, Foster JM, Carlow CK, Kumar S (2009) Computational prediction of essential genes in an unculturable endosymbiotic bacterium, Wolbachia of Brugia malayi. BMC microbiology 9: 243.

43. Babayan SA, Read AF, Lawrence RA, Bain O, Allen JE (2010) Filarial parasites develop faster and reproduce earlier in response to host immune effectors that determine filarial life expectancy. PLoS Biol 8: e1000525.

44. Saint Andre A, Blackwell NM, Hall LR, Hoerauf A, Brattig NW, et al. (2002) The role of endosymbiotic Wolbachia bacteria in the pathogenesis of river blindness. Science 295: 1892-1895.

45. Lustigman S, James ER, Tawe W, Abraham D (2002) Towards a recombinant antigen vaccine against Onchocerca volvulus. Trends Parasitol 18: 135-141.

46. Makepeace BL, Jensen SA, Laney SJ, Nfon CK, Njongmeta LM, et al. (2009) Immunisation with a multivalent, subunit vaccine reduces patent infection in a natural bovine model of onchocerciasis during intense field exposure. PLoS Negl Trop Dis 3: e544.

47. Yoshimura A, Naka T, Kubo M (2007) SOCS proteins, cytokine signalling and immune regulation. Nat Rev Immunol 7: 454-465.

48. Akhtar LN, Benveniste EN (2011) Viral exploitation of host SOCS protein functions. J Virol 85: 1912-1921.

49. Ludin P, Nilsson D, Mäser P (2011) Genome-wide identification of molecular mimicry candidates in parasites. PLoS One 6: e17546.

50. Baier M, Bannert N, Werner A, Lang K, Kurth R (1997) Molecular cloning, sequence, expression, and processing of the interleukin 16 precursor. Proc Natl Acad Sci U S A 94: 5273-5277.

51. Cruikshank WW, Kornfeld H, Center DM (2000) Interleukin-16. J Leukoc Biol 67: 757-766.

52. Harcus YM, Parkinson J, Fernandez C, Daub J, Selkirk ME, et al. (2004) Signal sequence analysis of expressed sequence tags from the nematode Nippostrongylus brasiliensis and the evolution of secreted proteins in parasites. Genome Biol 5: R39.

53. Käll L, Krogh A, Sonnhammer EL (2004) A combined transmembrane topology and signal peptide prediction method. J Mol Biol 338: 1027-1036.

54. Fankhauser N, Mäser P (2005) Identification of GPI anchor attachment signals by a Kohonen self-organizing map. Bioinformatics 21: 1846-1852.

55. Belanger DH, Perkins SL, Rockwell RF (2010) Inference of Population Structure and Patterns of Gene Flow in Canine Heartworm (Dirofilaria immitis). J Parasitol 97: 602-609.

56. Griffitts JS, Huffman DL, Whitacre JL, Barrows BD, Marroquin LD, et al. (2003) Resistance to a bacterial toxin is mediated by removal of a conserved glycosylation pathway required for toxin-host interactions. J Biol Chem 278: 45594-45602.

57. Bourguinat C, Keller K, Bhan A, Peregrine A, Geary T, et al. (2011) Macrocyclic lactone resistance in Dirofilaria immitis. Vet Parasitol 181: 388-392.

58. Bourguinat C, Keller K, Blagburn B, Schenker R, Geary TG, et al. (2011) Correlation between loss of efficacy of macrocyclic lactone heartworm anthelmintics and P-glycoprotein genotype. Vet Parasitol 176: 374-381.

59. Robertson G, Schein J, Chiu R, Corbett R, Field M, et al. (2010) De novo assembly and analysis of RNA-seq data. Nat Methods 7: 909-912.

60. Chen N (2004) Using RepeatMasker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics Chapter 4: Unit 4.10.

61. Kapitonov VV, Jurka J (2008) A universal classification of eukaryotic transposable elements implemented in Repbase. Nature reviews Genetics 9: 411-412; author reply 414.

62. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

63. Stamatakis A, Ludwig T, Meier H (2005) RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21: 456-463.

64. Eddy SR (1995) Multiple alignment using hidden Markov models. Proc Int Conf Intell Syst Mol Biol 3: 114-120.

65. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, et al. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 27: 29-34.

66. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22: 4673-4680.

Figure legends

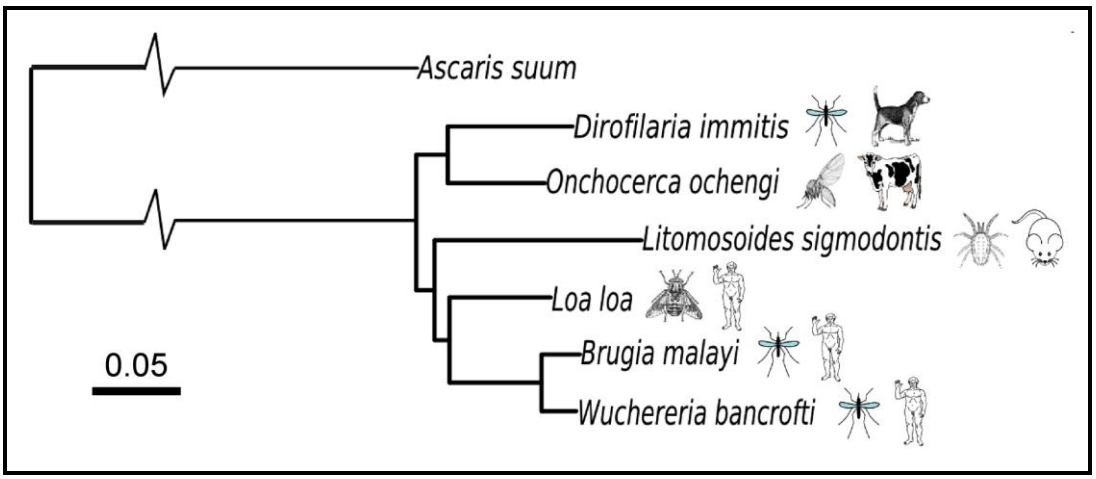


Figure 1. Phylogeny of the Onchocercidae. Phylogenetic consensus tree constructed from a set of 990 genes that are present as single-copy in all available genome drafts from Onchocercidae and, as an outgroup, *Ascaris suum*. Icons represent the vectors and definite hosts. The bootstrap support was 100% for each node.

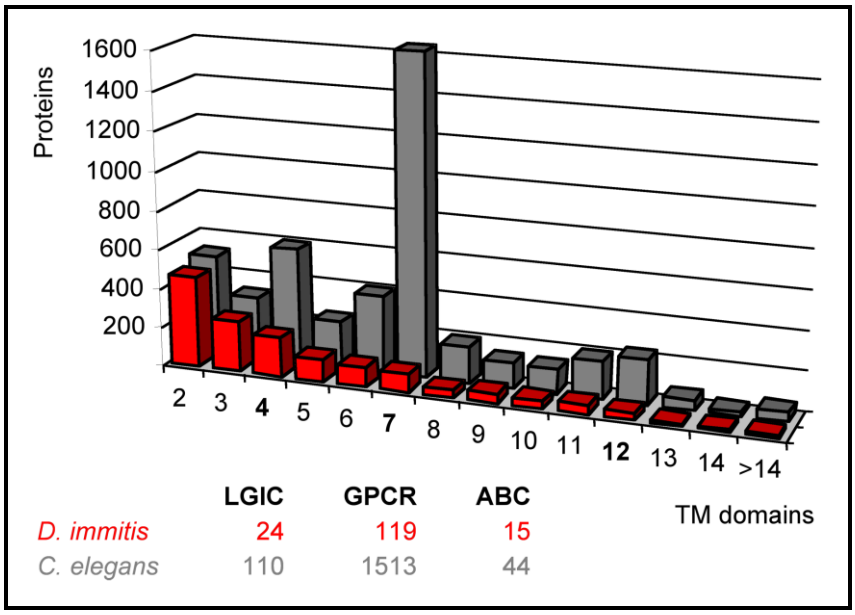


Figure 2. Frequency distribution of predicted transmembrane domains. Transmembrane (TM) domains were predicted with Phobius for all proteins of *C. elegans* (grey) and *D. immitis* (red). In parallel, ligand-gated ion channels (LGIC, 4 TM domains per monomer), G protein coupled receptors (GPCR, 7 TM domains) and ATP-binding cassette transporters (ABC, 12 TM domains per transporter) were identified with hidden Markov model-based profiles from Pfam (see supplementary Methods); the number of obtained hits is plotted below the graph.

Tables

	C. elegans	*B. malayi*	*D. immitis*
Estimated genome size	100 Mb	95 Mb	95 Mb
Assembly size	100 Mb	70,8 Mb	84,2 Mb
Gene models	21,193	11,515	11,375
Genes per Mb	235	162	135
Number of predicted proteins	19'762	11'508	12,344
Percent-protein-coding sequence	30.2	17.8	18.0
Median number of exons per gene	6	5	5
Median exon size	147 b	140 b	142 b
Median intron size	68 b	219 b	226 b
Overall GC content	35.4%	30.5%	28.3%
Exons GC content	42.9%	39.6%	37.4%
Introns GC content	29.1%	27.6%	26.6%

Table 1. Genome metrics. Comparison of the genome assemblies of *C. elegans*, *B. malayi* and *D. immitis*.

Chemical class	Drugs	Target	*C. elegans*	*D. immitis*
Benzimidazole	Albendazole Flubendazole Mebendazole	Beta-tubulin	BEN-1	DimmContig2035_DIMM36740
Imidazothiazole	Levamisole	nACh receptor	LEV-1	-
			LEV-8	-
			UNC-29	DimmContig1332_DIMM30000
			UNC-38	DimmContig3506_DIMM45965
			UNC-63	DimmContig160_DIMM08405
Macrocyclic lactone	Ivermectin Milbemycin Moxidectin Selamectin	Glutamate receptor	AVR-14	DimmContig464_DIMM16610
			AVR-15	-
			GLC-1	-
			GLC-2	DimmContig953_DIMM25280, DimmContig692_DIMM21120
			GLC-3	-
			GLC-4	DimmContig747_DIMM22030
		GABA receptor	EXP-1	DimmContig8852_DIMM57890-
			GAB-1	DimmContig1631_DIMM33210
			UNC-49	
Cyclodepsipeptide	Emodepside	K^+ channel	SLO-1	DimmContig1682_DIMM33710
		Latrophilin GPCR	LAT-1	DimmContig2098_DIMM37270, DimmContig2098_DIMM37275
			LAT-2	DimmContig514_DIMM17690
Aminoacetonitrile derivative	Monepantel	nACh receptor	ACR-23	-
			DES-2	-

Table 2. Candidate drug targets, top-down search. Current anthelmintics, their known targets in *C. elegans*, and orthologues in *D. immitis* as identified with reciprocal BLAST searches using InParanoid.

D. immitis protein	Predicted function	*H. sapiens* $\log_{10}$ (E)	*C. lupus* $\log_{10}$ (E)	C. elegans RNAi
Nucleic acid synthesis and repair				
DimmContig1353_DIMM30225	RNA-dependent RNA polymerase	>1	0.89	lethal
DimmContig188_DIMM09370	RNA-dependent RNA polymerase	0.28	0.11	lethal
DimmContig829_DIMM23395	Apurinic/apyrimidinic endonuclease	>1	-0.12	lethal
Glycosylation and sugar metabolism				
DimmContig417_DIMM15580	dTDP-4-dehydrorhamnose 3,5-epimerase	0.23	0.04	lethal
DimmContig51_DIMM03355	Beta-1,4-mannosyltransferase	0.95	0.94	lethal
DimmContig3217_DIMM44525	UDP-galactopyranose mutase	0.36	0.08	molt defective
DimmContig2055_DIMM36945	Chitin synthase	-3.52	-4.00	lethal
Lipid metabolism				
DimmContig5517_DIMM52545	Lipase	0.08	-1.60	lethal
DimmContig338_DIMM13730	Sterol-C24-methyltransferase (Erg11)	-3.10	-4.00	lethal
DimmContig1280_DIMM28375	Methyltransferase	-0.03	-0.15	lethal
Transport				
DimmContig689_DIMM21065	Aquaporin	-2.05	-0.52	lethal
Signal transduction				
DimmContig333_DIMM13570	Nuclear hormone receptor	-2.40	-4.52	lethal
DimmContig245_DIMM11130	G-protein coupled receptor	-1.06	-0.59	lethal
DimmContig1550_DIMM32415	G-protein coupled receptor	-1.26	-1.57	lethal
DimmContig2384_DIMM39455	G-protein coupled receptor	-2.70	-1.96	lethal
DimmContig335_DIMM13630	Groundhog protein	>1	0.04	lethal
DimmContig3787_DIMM47150	Warthog protein	>1	0.78	lethal
DimmContig48_DIMM03220	Warthog protein	>1	0.32	lethal
DimmContig253_DIMM11410	Haloacid dehalogenase-like hydrolase	-3.22	-4.00	lethal
DimmContig326_DIMM13420	Apoptosis regulator ced-9	0.77	-1.02	lethal

Table 3. Candidate drug targets, bottom-up search. Potential drug targets were filtered from the predicted *D. immitis* proteome using the following criteria: (i) presence of an orthologue in *C. elegans* which has as an RNAi phenotype *lethal*, *L_3_arrest*, or *molt_defective*; (ii) absence of a significant BLAST match (E-value below 10^{-5}) in the predicted proteomes of *H. sapiens* and *C. lupus familiaris*, and (iii) predicted function as an enzyme or receptor.

D. immitis protein(s)	*B. malayi* orthologue(s)	Description	VC	IM
DIMM39040; DIMM39045	BM18548	Pi-class glutathione S-transferase (GSTP)	X	
DIMM29150	BM02625	Tropomyosin (TMY)	X	
DIMM29270	BM00759; BM19824	Fatty acid and retinoic acid binding protein (FAR)	X	
DIMM47055	BM03010	Fructose bisphosphate aldolase (FBA)	X	
DIMM59360		Astacin metalloprotease MP1	X	
DIMM37935; DIMM46475	BM01859; BM09541; BM14520	Chitinases (CHI)	X	
DIMM48695	BM21967; BM08119	Abundant larval transcript 1 (ALT-1; also known as SLAP)	X	
DIMM48700	BM20051	"RAL-2"; unknown function; DUF148 superfamily (also known as SXP-1)	X	
DIMM62215; DIMM45570; DIMM58880	BM03177; BM05783; BM16294	Activation associated proteins (ASP, also known as venom allergen homologues, VAH)	X	
DIMM58690	BM02480	"OV103" *Onchocerca* vaccine candidate of unknown function	X	
DIMM12355	BM07484; BM22082	"B8" *Onchocerca* vaccine candidate of unknown function	X	
DIMM55190; DIMM50565; DIMM48395	BM00175; BM14240; BM04930; BM07956	"B20" *Onchocerca* vaccine candidate of unknown function	X	
DIMM56580	BM05118	Cysteine proteinase inhibitor 2 (CPI-2)	X	X
DIMM18905	BM04900	Cysteine proteinase inhibitor 3 (CPI-3)	X	X
DIMM11425	BM21284	Interleukin-16 (IL16)		X
DIMM57180	BM00325	Leucyl aminopeptidase (LAP)		X
DIMM28945	BM06847	Suppressor of cytokine signaling 5 (SOCS5)		X
DIMM42430	BM07480	Macrophage migration inhibitory factor (MIF-1)		X
DIMM40455	BM16561	Macrophage migration inhibitory factor 2 (MIF-2)		X
DIMM23225	BM17713	Transforming growth factor beta (TGF) homologue of *C. elegans* TIG-2		X
DIMM37585	BM20852	TGF homologue of *C. elegans* DAF-7		X
DIMM29335	BM21753	TGF homologue of *C. elegans* DBL-1/CET-1		X
DIMM61250	BM18112	TGF homologue of *C. elegans* UNC-129		X

Table 4. Potential immune modulators and vaccine candidates. *D. immitis* proteins that are potential immune modulators (IM) or that have been proposed as vaccine candidates (VC) in other Onchocercidae.

Supplementary tables

Library name	fas1	fas2	blx	mp
Type	76 bp paired-end	76 bp paired-end	101 bp paired-end	100 bp mate-paired
Insert length	343 (58)	342 (57)	108 (16)	-
#Reads (raw)	23,527,084	35,637,504	47,085,106	188,325,900
#Bases (raw)	1,788,058,384	2,708,450,304	4,755,595,706	18,832,590,000
#Reads (trimmed)	15,725,202	28,369,128	45,328,600	181,344,222
#Bases (trimmed)	1,083,296,217	2,006,850,353	4,446,959,717	9,049,499,979

Table S1. Sequencing libraries. Counts of raw and quality-trimmed genome sequencing reads. Insert length is given as the median with standard deviation in parentheses.

	ABySS	Velvet
Parameters	Kmer 35 using mp library as se, n=4	Kmer 35 se, -exp_cov auto -cov_cutoff auto
Max contig size	167,771	117,878
Number of contigs >=200	31,979	31,982
Number of bases for contigs >=200	87,011,407	76,294,902
GC for contigs >=200	28.3	28.8
N50 for contigs >=200	10,474	7,125
D.immitis ESTs recovered	1,533	1,442
CEGMA completeness	Complete 95% Partial 96%	Complete 89% Partial 93%

Table S2. Genome assembly comparisons. Based on the above data, the ABySS assembly was selected for all further analyses.

Assembler	TransABySS	Velvet-Oases
Parameters	Kmer range 23-47	Kmer 35, insert length 120
Max. contig size	14,946	10,465
Number of contigs >=300 bp	31,643	27,935
Total bases in contigs >=300 bp	34,603,394	34,479,062
N50 of contigs >=300 bp	1,477	1,702
Mean length of contigs >=300 bp	1,094	1,234
Number of contigs >=1000 bp	12,391	12,844
Total bases in contigs >=1000 bp	23,720,523	25,873,800
Alignments to sanger ESTs	1,030	991
(>80% of EST covered)		
CEGMA completeness	Complete 94%	Complete 88%
	Partial 96%	Partial 93%
B. malayi peptide blastx	13,487	12,719
(>80% of ref. sequence covered)		

Table S3. Transcriptome assembly comparisons. Based on the above data, the TransABySS assembly was selected for all further analyses.

Supplementary figures

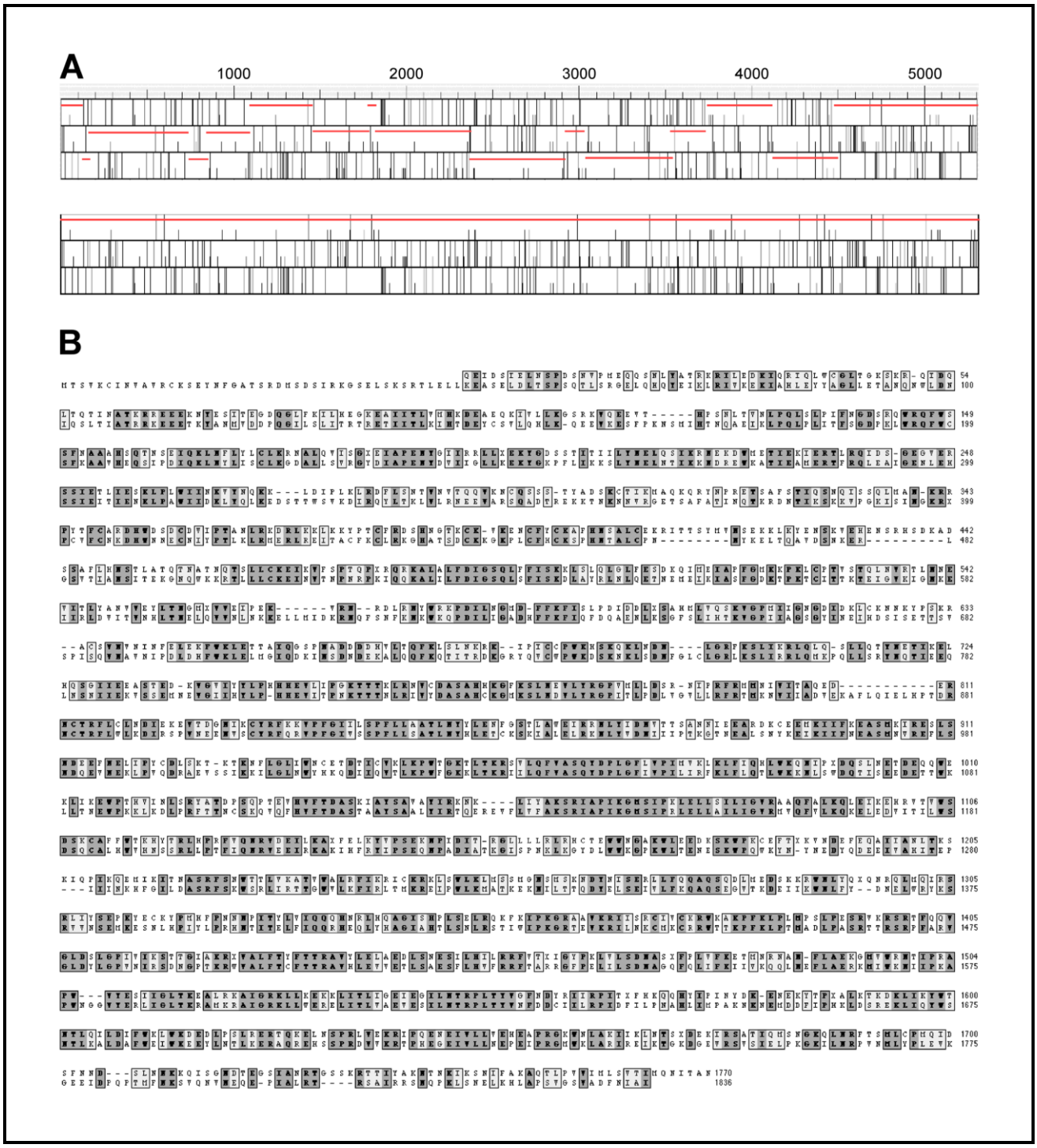


Figure S1. Example of a *D. immitis* PAO pseudogene. Panel A shows the 3 translation frames of the *D. immitis* PAO-290022 element (top), with the ATG and stop codons indicated by short and long vertical bars, respectively. The 5' extremity of this transposable element is truncated. The horizontal red lines correspond the coding sequence matching the *B. malayi* PAO elements. This coding sequence contains 15 frame shifts and numerous stop codons. Below is a "reconstituted" version of the PAO-290022 coding sequence, where all the frame shifts have been removed manually. This artificial gene still contains 15 stop codons. Panel B is an alignment of the reconstituted protein

encoded by PAO-290022 (first lane: 1170 aa) and one full-length (and likely functional) *B. malayi* PAO element (second lane: EDP32991 - 1836 aa).

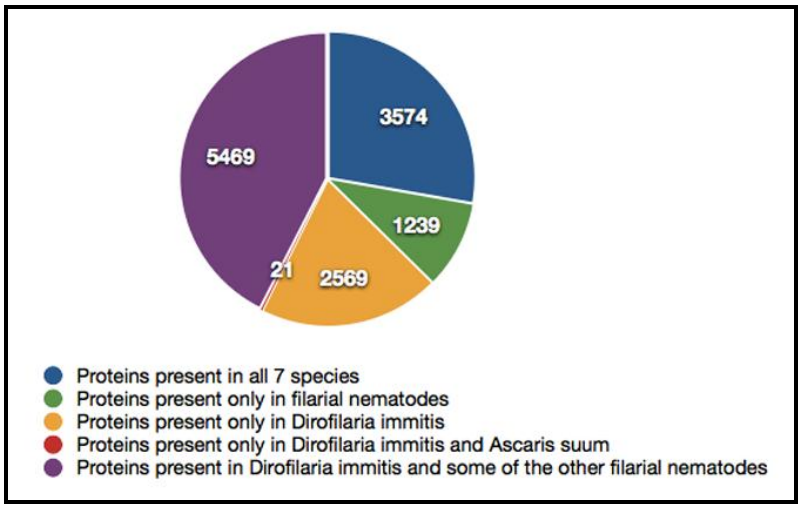


Figure S2. Protein orthology groups. Numbers of proteins in different intersects of the predicted proteomes of filaria (*Dirofilaria immitis*, *Brugia malayi*, *Onchocerca ochengi*, *Litomosoides sigmodontis*, *Loa loa*, *Wuchereria bancrofti*) and *Ascaris suum*, as classified by InParanoid. Of the omnipresent orthology groups (blue sector), 990 had a single member in all the seven species and were therefore used to build the phylogenetic tree shown in Figure 1.

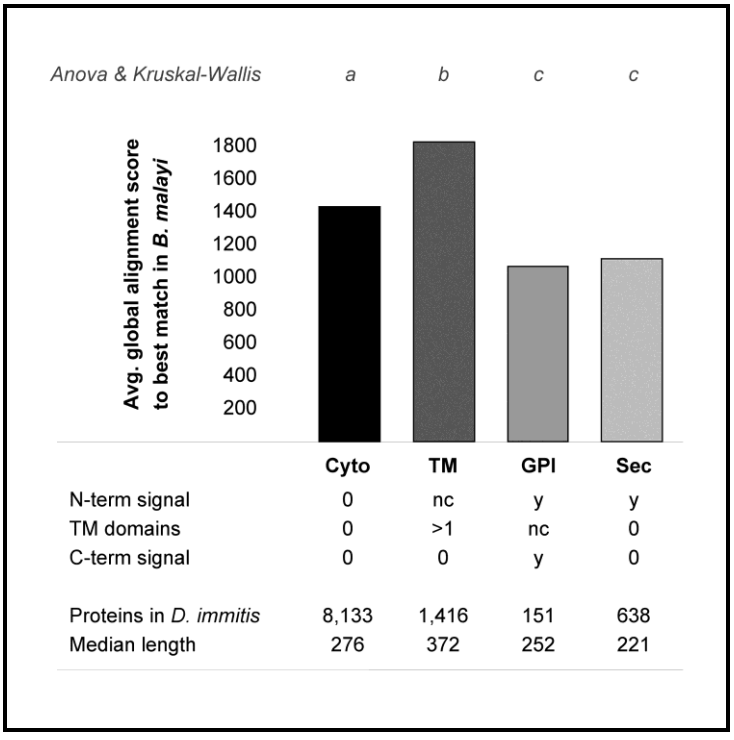


Figure S3. GPI-anchored and secreted proteins are the least conserved. The *D. immitis* proteome was subdivided into the four classes 'cytosolic', 'transmembrane', 'secreted', and 'GPI anchored' based on the predicted presence of transmembrane domains, N-terminal export signals, and C-terminal GPI anchor attachment signals as indicated (nc, not considered). For each set, every protein was aligned to its best match from *B. malayi*. Given that the defined classes of proteins differed in their median length, the Needleman-Wunsch global alignment algorithm was implemented (rather

than the local alignment performed by Blast) to avoid disfavoring of shorter proteins. The resultant scores were significantly lower for the secreted and GPI-anchored proteins as compared to transmembrane or cytosolic proteins ($p<0.01$; 1-way Anova followed by Kruskal-Wallis' post test).

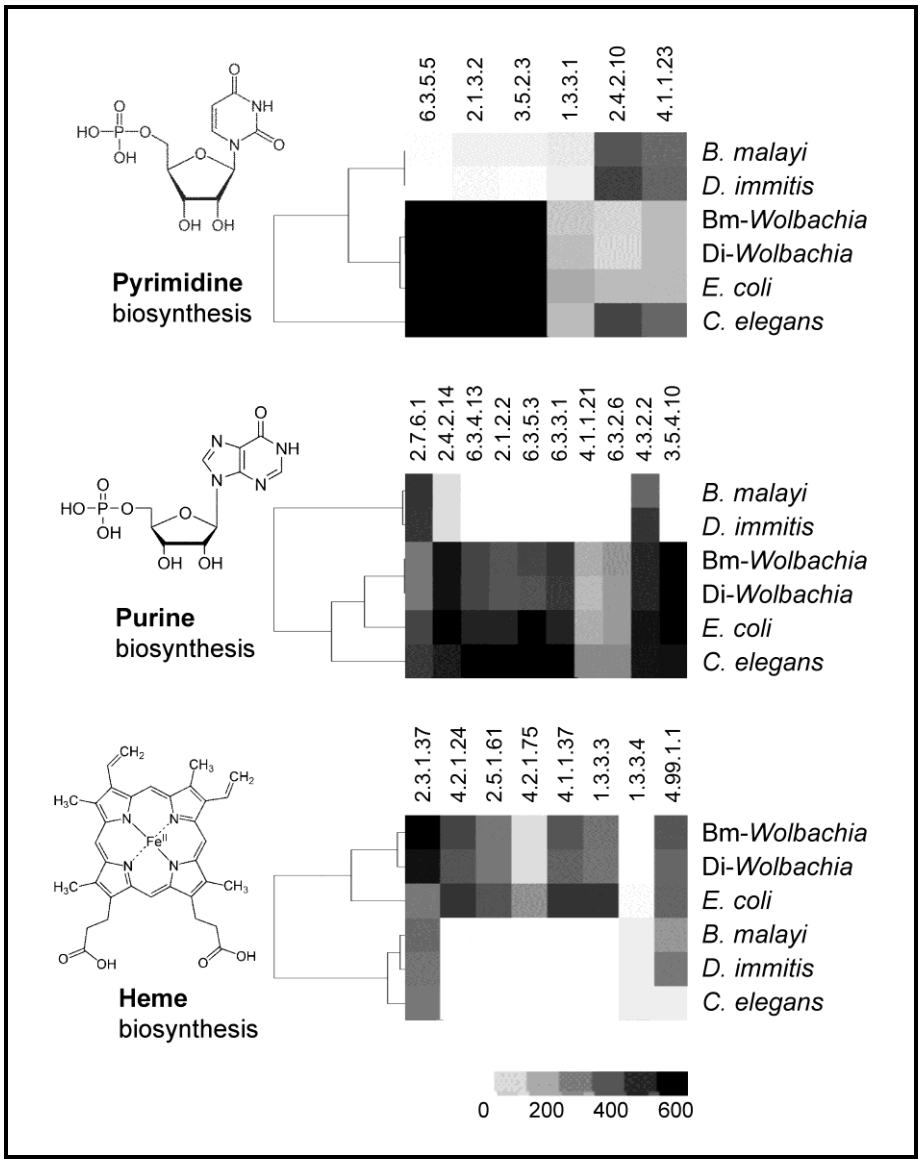


Figure S4. Anabolic pathways in *Wolbachia* and *Dirofilaria*. Three pathways of interest are shown. Hidden Markov model-based profiles were made for all enzymes involved and used to screen predicted proteomes with HMMer. The heat plot represents the best score obtained against each proteome. These scores were clustered by city block distance between the pathway 'vectors'.

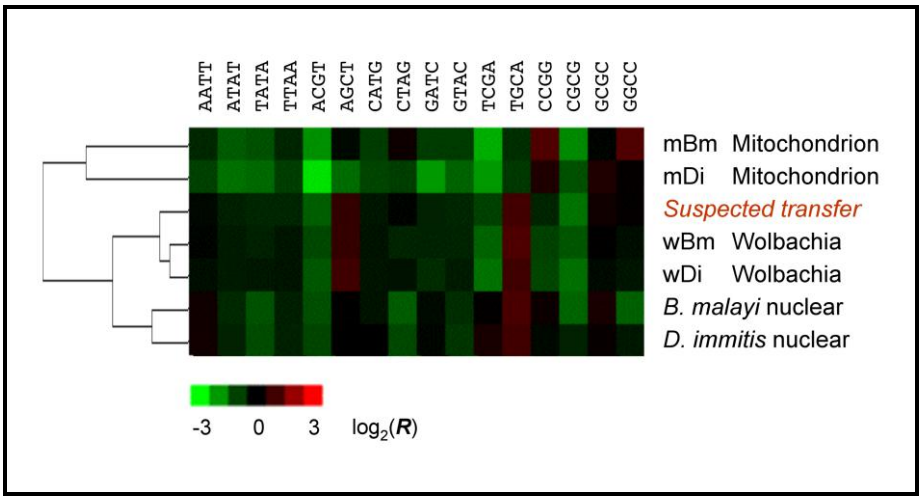


Figure S5. Clustering of DNA based on palindrome frequency. Nuclear genomic DNA sequences of *D. immitis* (this work) and *B. malayi* (GenBank accession DS239371) were compared to *Wolbachia* *w*Di (this work) and *w*Bm (NC_006833), mitochondria mDi (NC_005305) and mBm (NC_004298), and to joined DNA fragments ('horizontal transfer joined') suspected to have been horizontally transferred from wDi to the *D. immitis* nucleus (see main text). For each sequence string, the actual occurrence of the 16 possible palindromes of length 4 (top labels) was divided by the expected occurrence given its GC-content. These ratios were transformed to $\log_2$ and clustered based on city-block distance. The suspected DNA fragments of *D. immitis* clustered with the *Wolbachia* genomes, providing further evidence that they had been acquired by the nucleus through horizontal transfer.

Genome-wide survey for ligand gated ion channels in *D. immitis*

Having the genome in hands, it was possible to make an inventory of the ligand-gated ion channels (LGIC). LGIC are divided into different families; one of them, the nicotinic acetylcholine receptor (nAChR) family, comprises several targets of anthelmintics including the target of the AADs (32). We have identified 24 proteins that contain either the ligand-binding or the transmembrane or both of LGIC (Table 3).

Table 3. Description of the 24 LGIC receptors of *D. immitis*, with homology to *C. elegans.* LBD means ligand-binding domains and "memb" is related to the membrane domains of the LGIC. E (E (expect)-value): parameter describing the number of hits "expected" to see by chance when searching a database with a particular size), Score (number used to assess the biological relevance of a finding; it describes the overall quality of an alignment).

Profile	Hit	E	Score	Homolog/ Best hit in (*C.elegans*)
Neur_chan_LBD	DimmContig1076_DIMM26960	1.5E-33	103.1	lgc-55
Neur_chan_LBD	DimmContig123_DIMM06980	3.6E-50	157.4	lgc-49
Neur_chan_LBD	DimmContig1262_DIMM29165	1.8E-55	174.7	acr-8
Neur_chan_LBD	DimmContig1333_DIMM30000	1.9E-61	194.3	unc-29
Neur_chan_memb	DimmContig14949_DIMM62135	1.2E-19	58.4	unc-49a
Neur_chan_memb	DimmContig160_DIMM08405	3.1E-64	204.4	unc-63
Neur_chan_LBD	DimmContig17_DIMM01370	6.4E-45	140.3	lgc-30
Neur_chan_LBD	DimmContig1990_DIMM36340	2.5E-40	125.2	lgc-26
Neur_chan_LBD	DimmContig206_DIMM09920	9.0E-48	149.6	lgc-46
Neur_chan_LBD	DimmContig226_DIMM10520	4.5E-65	206.1	acr-12
Neur_chan_LBD	DimmContig2623_DIMM41090	8.1E-49	153	lgc-47
Neur_chan_LBD	DimmContig3506_DIMM45965	3.8E-75	239.1	unc-38
Neur_chan_LBD	DimmContig3655_DIMM46615	3.2E-37	115.1	acc-1
Neur_chan_LBD	DimmContig3655_DIMM46620	5.3E-42	130.7	acc-1
Neur_chan_LBD	DimmContig3792_DIMM47160	1.2E-69	221.1	acr-15
Neur_chan_memb	DimmContig3824_DIMM47310	5.5E-35	108.7	lgc-53
Neur_chan_memb	DimmContig3972_DIMM47885	6.3E-23	69.2	
Neur_chan_LBD	DimmContig431_DIMM15880	2.9E-50	157.7	lgc-55
Neur_chan_memb	DimmContig464_DIMM16610	2.0E-76	244.3	apr-14
Neur_chan_memb	DimmContig490_DIMM17190	3.3E-13	37.4	lgc-41
Neur_chan_LBD	DimmContig692_DIMM21120	8.6E-46	143.1	glc-2
Neur_chan_LBD	DimmContig747_DIMM22030	9.7E-48	149.5	glc-4
Neur_chan_LBD	DimmContig8745_DIMM57755	2.9E-19	56.4	unc-38
Neur_chan_LBD	DimmContig953_DIMM25280	3.4E-55	173.8	glc-2

Interestingly, amongst those 24 LGIC receptors no members of the DEG-3 such family of nAChR is found. This contrasts with *B. malayi* which possess two DEG-3 family members.

The number of LGIC varies among classes and species of organisms. An interesting phylogenetic comparison of the number of LGIC was described by Rufener et al.; *C.*

elegans belongs to the clade V in the nematode phylogeny and shows 102 LGIC receptors (126). *D. immitis* is part of the clade III as *B. malayi* and *A. suum*. The number of LGIC between the filariae, *D. immitis* and *B. malayi*, is equivalent to 24 receptors, while the ascaridae shows 98 LGIC receptors (99). In the recent publication of the *Ascaris suum* draft genome, the authors even claim the presence of a homolog to the acr-23 of the free-living nematode. However, the authors failed to perform reciprocal blast searches. A simple tblastn search of the claimed homolog against *C. elegans* indicated that the mentioned sequence has a better hit to des-2 of *C. elegans*. Nevertheless, *A. suum* genome contains DEG-3 family members; while this receptor family is absent in *D. immitis*. The third publication, “Loss of Deg-3-subfamily acetylcholine receptors and lack of stereoselective monepantel sensitivity in *Dirofilaria immitis*”, describes in more detail the search for *D. immitis* LGIC.

Chapter 4. Loss of DEG-3-subfamily acetylcholine receptors and lack of stereoselective monepantel sensitivity in *Dirofilaria immitis*

Publication 3

In submission at Veterinary Parasitology Journal

Loss of DEG-3-subfamily acetylcholine receptors and lack of stereoselective monepantel sensitivity in *Dirofilaria immitis*

Christelle Godel[1-3], Lucien Rufener[3], Daniel Nilsson[4], Sandra Schorderet-Weber[3], Christian Epe[3], Ronald Kaminsky[3], Pascal Mäser[1,2*]

[1]Swiss Tropical and Public Health Institute, CH-4002 Basel, Switzerland

[2]University of Basel, CH-4000 Basel, Switzerland

[3]Centre de Recherche Santé Animale, CH-1566 St. Aubin, Switzerland

[4]Scilife Lab, Karolinska Institutet, Stockholm, Sweden

[*]Corresponding author

In this short communication, I performed in vitro tests with microfilariae of *D. immitis*, coordinated the collection of data, put scientists in relation to continue the tests, draw the conclusions and wrote part of the text.

Loss of DEG-3-subfamily acetylcholine receptors and lack of stereoselective monepantel sensitivity in *Dirofilaria immitis*

Christelle Godel[1-3], Jacques Bouvier[3], Sandra Schorderet-Weber[3], Lucien Rufener[3], Daniel Nilsson[4], Christian Epe[3], Ronald Kaminsky[3], Pascal Mäser[1,2*]

[1]Swiss Tropical and Public Health Institute, CH-4002 Basel, Switzerland

[2]University of Basel, CH-4000 Basel, Switzerland

[3]Centre de Recherche Santé Animale, CH-1566 St. Aubin, Switzerland

[4]Scilife Lab, Karolinska Institutet, Stockholm, Sweden

[*]Corresponding author:

Address	Swiss Tropical and Public Health Institute
	Socinstrasse 57
	CH-4002 Basel
	Switzerland
Tel	+41 61 284 8338
Fax	+41 61 284 8101
E-mail	pascal.maeser@unibas.ch

Keywords heartworm, Zolvix, acetylcholine receptor, pharmacogenomics

The heartworm *Dirofilaria immitis* is a parasitic nematode of mammals that is transmitted by mosquitoes (*Culicidae*). The definite host is the dog, where adult worms reside in the pulmonary arteries and the females shed microfilariae into the bloodstream. Dirofilariasis is a severe disease of dogs (Simon et al., 2009). The recommended preventive (Bowman and Atkins, 2009) is chemoprophylaxis with macrocyclic lactones (ML; ivermectin, milbemycin oxime, moxidectin, or selamectin) which, however, bears the risk of selecting for drug resistant parasites (Bourguinat et al., 2011). In the absence of a vaccine, new drugs are needed for sustainable control of dirofilariasis. Drug development for the heartworm will be of benefit also against the closely related pathogens *Onchocerca volvulus* or *Loa loa*, the causative agents of neglected human filarial diseases. Monepantel (Novartis compound AAD-1566; (Kaminsky et al., 2008)) is a new addition to the anthelmintic armory. Monepantel is the first aminoacetonitrile derivative (AAD) to reach the market, used under the brand name Zolvix® (Novartis 2010) against gastrointestinal nematodes of sheep such as *Haemonchus contortus* or *Trichostrongylus colubriformis*. Monepantel is highly effective against these parasites (Hosking et al., 2009), including ML-resistant isolates (Baker et al., 2011), and well tolerated by the mammalian hosts (Kaminsky et al., 2008). Here we investigate the potential of monepantel against dirofilariasis.

The efficacy of monepantel was tested against the model filarid *Acanthocheilonema viteae* in gerbils and against *D. immitis* in dogs. While there was no efficacy in vivo, the EC_{50} of monepantel against *D. immitis* microfilariae in vitro was 2 µM, even slightly below that of milbemycin oxime (Figure 1). Monepantel is thought to act by allosteric hyperactivation of ligand-gated ion channels, the known targets comprising ACR-23 in *C. elegans*, and MPTL-1 and DES-2 in *H. contortus*. These are putative nicotinic acetylcholine receptor (nAChR) subunits of the DEG-3 subfamily, a clade of nAChR that only occurs in nematodes. The presence of MPTL-1 or ACR-23 orthologues in helminth genomes correlates with monepantel susceptibility (Rufener et al., 2010b), and loss-of-function mutations in MPTL-1 and ACR-23 were associated with loss of monepantel sensitivity in *H. contortus* (Rufener et

al., 2009) and *C. elegans* (Kaminsky et al., 2008), respectively. We have scanned the predicted proteome of *D. immitis* (Godel et al., unpublished/in press) for ligand-gated ion channels with Pfam profiles (Punta et al., 2011) specific for the extracellular ligand-binding domain (PF02931) and for the membrane ion channel domain (PF02932). Using HMMer (Eddy, 2009) with a conservative cut-off E-value of $<10^{-10}$ this yielded 39 hits for the ligand-binding domain, 32 hits for the membrane domain, and an intersect of 24 proteins carrying both domains. The same search in *C. elegans* (Harris et al., 2011) returned 110 proteins with both domains, which is in agreement with published numbers (Jones and Sattelle, 2008). A phylogenetic tree based on the multiple alignment (Thompson et al., 1994) of the pooled hits obtained from *C. elegans* and *D. immitis* revealed that the targets of macrocyclic lactones and levamisole are conserved in the heartworm, but those of the aminoacetonitrile derivatives are not (Figure 2). In fact, DEG-3 subfamily members seem to be completely absent from the *D. immitis* proteome (Figure 2). This was supported by tblastn searches (Altschul et al., 1990) against the *D. immitis* genomic assembly (Godel et al., unpublished/in press) with DEG-3 family sequences, returning but a *des-2* pseudogene which carries several in-frame stop codons, and which is not transcribed in *D. immitis* adult worms (data not shown). The presence of the *des-2* pseudogene in *D. immitis*, and the presence of both DEG-3 and DES-2 orthologues in *Brugia malayi* (Williamson et al., 2007), suggest that the dirofilarial ancestor had lost the functional DEG-3 type genes after separation from the line leading to *Brugia*. Given the absence of DEG-3 subfamily nAChR in *D. immitis*, the activity of monepantel against microfilariae must be directed at another target (or targets).

Aminoacetonitrile derivatives contain a center of chirality. The active enantiomer monepantel is the S-form, the R enantiomer AAD-2224 being 1,000 fold less potent against *H. contortus* or *T. colubriformis* (Figure 1). The same stereoselectivity is reflected by the action of monepantel against ligand-gated ion channels expressed in *Xenopus laevis* oocytes (Rufener et al., 2010a). However, the observed in vitro activity of monepantel against *D. immitis* microfilariae was

not stereoselective, the two enantiomers exhibiting the same potency (Figure 1). This result is in agreement with the absence of the known, stereoselective monepantel targets in *D. immitis* (Figure 2). In summary, we show that monepantel has in vitro activity against *D. immitis* microfilariae but the target, given the absence of DEG-3 genes and the lack of stereoselectivity, is unlikely to be an acetylcholine receptor channel. The molecular nature of the monepantel target in the heartworm remains to be identified.

ACKNOWLEDGEMENTS

We are grateful to the Novartis AH Screening Team and to the Rodent Model Team for technical assistance. CG is supported by a Novartis Animal Health PhD Fellowship.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990, Basic local alignment search tool. J. Mol. Biol. 215, 403-410.

Baker, K.E., George, S.D., Stein, P.A., Seewald, W., Rolfe, P.F., Hosking, B.C., 2011, Efficacy of monepantel and anthelmintic combinations against multiple-resistant Haemonchus contortus in sheep, including characterisation of the nematode isolate. Vet Parasitol.

Bourguinat, C., Keller, K., Bhan, A., Peregrine, A., Geary, T., Prichard, R., 2011, Macrocyclic lactone resistance in Dirofilaria immitis. Vet Parasitol 181, 388-392.

Bowman, D.D., Atkins, C.E., 2009, Heartworm biology, treatment, and control. Vet Clin North Am Small Anim Pract 39, 1127-1158, vii.

Eddy, S.R., 2009, A new generation of homology search tools based on probabilistic inference. Genome Inform 23, 205-211.

Gill, J.H., Redwin, J.M., van Wyk, J.A., Lacey, E., 1995, Avermectin inhibition of larval development in Haemonchus contortus--effects of ivermectin resistance. Int J Parasitol 25, 463-470.

Harris, T.W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W.J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R., Fernandes, J., Han, M., Kishore, R., Lee, R., Muller, H.M., Nakamura, C., Ozersky, P., Petcherski, A., Rangarajan, A., Rogers, A., Schindelman, G., Schwarz, E.M., Tuli, M.A., Van Auken, K., Wang, D., Wang, X., Williams, G., Yook, K., Durbin, R., Stein, L.D., Spieth, J., Sternberg, P.W., 2011, WormBase: a comprehensive resource for nematode research. Nucleic Acids Res 38, D463-467.

Hosking, B.C., Kaminsky, R., Sager, H., Rolfe, P.F., Seewald, W., 2009, A pooled analysis of the efficacy of monepantel, an amino-acetonitrile derivative against gastrointestinal nematodes of sheep. Parasitol Res 106, 529-532.

Jones, A.K., Sattelle, D.B., 2008, The cys-loop ligand-gated ion channel gene superfamily of the nematode, Caenorhabditis elegans. Invert Neurosci 8, 41-47.

Kaminsky, R., Gauvry, N., Schorderet Weber, S., Skripsky, T., Bouvier, J., Wenger, A., Schroeder, F., Desaules, Y., Hotz, R., Goebel, T., Hosking, B.C., Pautrat, F., Wieland-Berghausen, S., Ducray, P., 2008, Identification of the amino-acetonitrile derivative monepantel (AAD 1566) as a new anthelmintic drug development candidate. Parasitol Res 103, 931-939.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C., Pang, N., Forslund, K., Ceric, G., Clements, J., Heger, A., Holm, L., Sonnhammer, E.L., Eddy, S.R., Bateman, A., Finn, R.D., 2011, The Pfam protein families database. Nucleic Acids Res 40, D290-D301.

Rufener, L., Baur, R., Kaminsky, R., Mäser, P., Sigel, E., 2010a, Monepantel allosterically activates DEG-3/DES-2 channels of the gastrointestinal nematode Haemonchus contortus. Mol Pharmacol 78, 895-902.

Rufener, L., Keiser, J., Kaminsky, R., Mäser, P., Nilsson, D., 2010b, Phylogenomics of ligand-gated ion channels predicts monepantel effect. PLoS Pathog 6, e1001091.

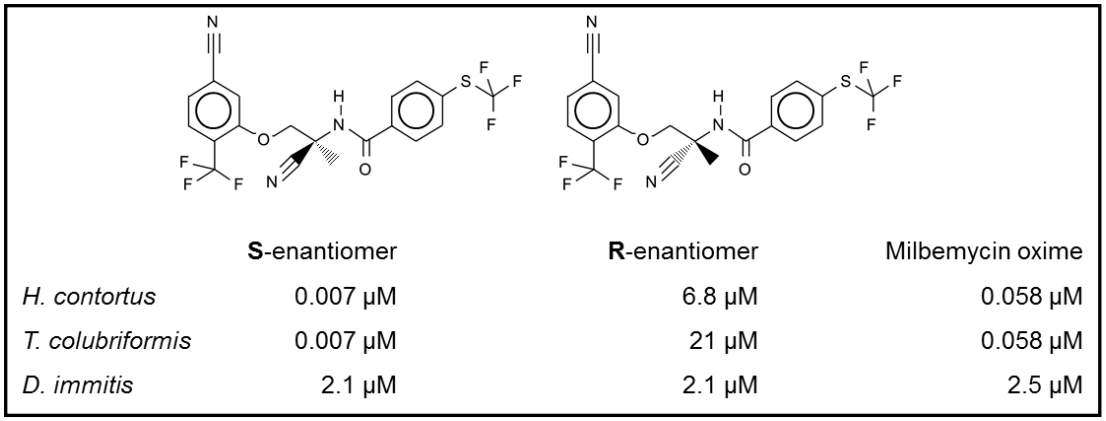
Rufener, L., Maser, P., Roditi, I., Kaminsky, R., 2009, Haemonchus contortus acetylcholine receptors of the DEG-3 subfamily and their role in sensitivity to monepantel. PLoS Pathog 5, e1000380.

Simon, F., Morchon, R., Gonzalez-Miguel, J., Marcos-Atxutegi, C., Siles-Lucas, M., 2009, What is new about animal and human dirofilariosis? Trends Parasitol 25, 404-409.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22, 4673-4680.

Williamson, S.M., Walsh, T.K., Wolstenholme, A.J., 2007, The cys-loop ligand-gated ion channel gene family of Brugia malayi and Trichinella spiralis: a comparison with Caenorhabditis elegans. Invert Neurosci 7, 219-226.

FIGURE LEGENDS



	S-enantiomer	**R**-enantiomer	Milbemycin oxime
H. contortus	0.007 µM	6.8 µM	0.058 µM
T. colubriformis	0.007 µM	21 µM	0.058 µM
D. immitis	2.1 µM	2.1 µM	2.5 µM

Figure 1. Structure and in vitro activity of monepantel (AAD-1566) and its mirror image (AAD-2224). EC_{50} values against *H. contortus* and *T. colubriformis* were determined with a larval development assay (Gill et al., 1995), and against *D. immitis* with a microfilarial viability assay (Godel et al., unpublished). The macrocyclic lactone milbemycin oxime was included as a reference.

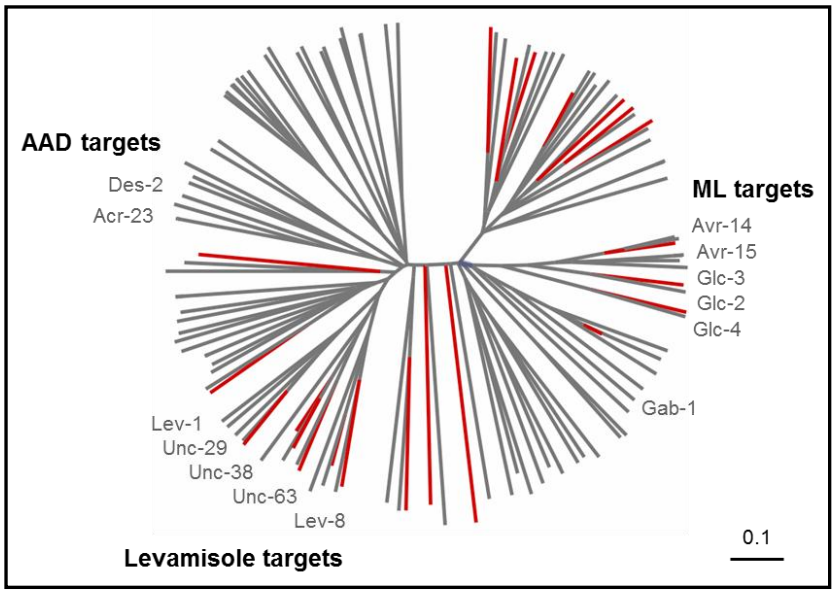


Figure 2. Phylogenetic tree of the predicted ligand-gated ion channels from *C. elegans* (grey) and *D. immitis* (red). All proteins were included that tested double positive for the ligand-binding domain (PF02931) and the membrane domain (PF02932). The scale bar indicates number of changes per residue. The known drug targets of *C. elegans* are labeled (ML, macrocyclic lactones; AAD, aminoacetonitrile derivatives).

Deeper investigations through the DEG-3-subfamily

Additional investigations were performed on this interesting receptor family, DEG-3, in *D. immitis* genome. After blast and reciproqual blast searches, no *acr-23*, *des-2* and *deg-3* homolog were found. Did I miss the receptor(s) to AADs or is/are it/they simply absent?

Another approach was to identify the neighbor genes of *acr-23*, *des-2* and *deg-3* in *C. elegans* and *B. malayi*. A blast search of the neighbor genes in *D. immitis* genome did not help to identify the target receptors of AADs. The neighbor genes are, partly, found in *D. immitis* genome; still no *acr-23*, *des-2* or *deg-3*. The genes might simply not be present. A tblastn search was similarly performed on the adult RNA sequences assembly without *acr-23*, *des-2* or *deg-3* found, leading to the conclusion that none of these receptors are expressed in adults. In the closely related filarial nematode *Onchocerca volvulus* no orthologs to these receptors were so far identified (125).

However, a pseudogene of *des-2* has been identified in the genome. To assess the presence of the *des-2* pseudogene in the genome of *D. immitis*, a RT-PCR (reverse transcriptase) was performed on genomic DNA and complementary DNA with primers specific to the des-2 pseudogene (Figure 16). The *des-2* pseudogene sequence is seen in the genomic DNA and is absent in the complementary DNA; concluding that the *des-2* pseudogene is not active, not translated into protein.

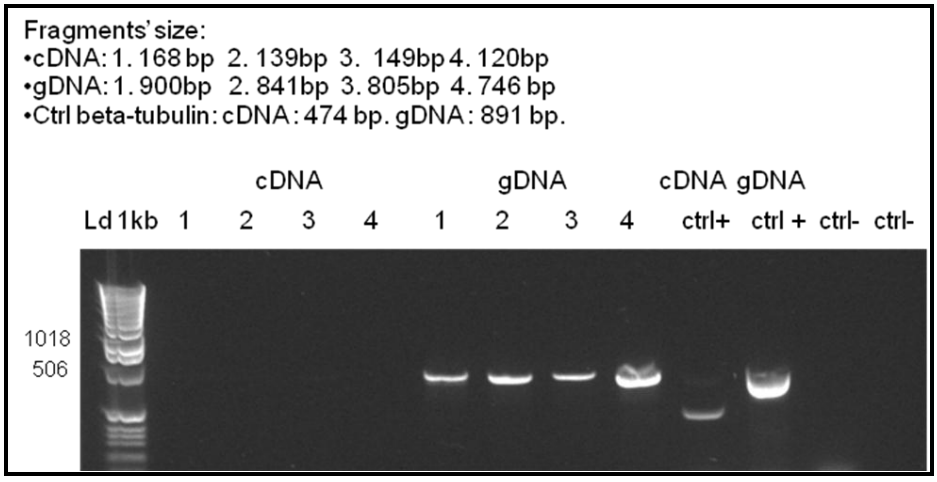


Figure 16. Agarose gel of 1.2% representing the RT-PCR (reverse transcriptase) picture performed on the *des-2* pseudogene sequence of *D. immitis* on complementary DNA and genomic DNA. Positive controls were performed with beta-tubulin specific to *D. immitis*. Negative controls are without primers or without DNA samples.

The aim of the PhD thesis was to characterize the receptor(s) to AADs in the genome of *D. immitis*. However, the absence of the receptor(s) has been proven previously. Having the genome assembled, other analyses are possible. In that way, based on exclusion criteria a closer study of the genome contigs allowed drawing a picture of novel potential drug targets.

In silico novel potential drug targets of *D. immitis*

To obtain a table of the novel potential drug targets by *in silico* filtering methodology, filters were applied on the whole proteome of *D. immitis* using various criteria. A similar methodology was applied on the genome of *B. malayi* by Kumar et al.; obtaining 589 potential targets (126). The filters applied were the absence of a blastp hit in the predicted proteomes of *Homo sapiens* and *Canis lupus familiaris* with an E-value below 10^{-5}, followed by the presence of an ortholog in *C. elegans* which has as an RNA interference (RNAi) phenotype *lethal*, *L_3_arrest*, or *molt_defective* and predicted function as an enzyme or receptor based on the annotation of the *C. elegans* ortholog and blastp searches against SwissProt (essential structural proteins such as innexin, etc., were excluded) (Table 4).

Table 4. Table regrouping the potential drug targets of *D. immitis* obtained after an exclusion criteria study on the whole proteome ((i) presence of an ortholog in *C. elegans* which has as an RNAi phenotype *lethal*, *L_3_arrest*, or *molt_defective*; (ii) absence of a blastp hit in the predicted proteomes of *H. sapiens* and *C. lupus familiaris* with an E-value below 10^{-5}; (iii) predicted function as an enzyme or receptor based on the annotation of the *C. elegans* ortholog and blastp searches against SwissProt).

D.immitis protein	Predicted function	Target for	*H. sapiens* $\log_{10}$ (E)	*C. lupus* $\log_{10}$ (E)	C. elegans RNAi
Nucleic acid synthesis and repair					
DimmContig1352_DIMM30240	RNA-dependent RNA polymerase	Antiviral	>1	0.89	lethal
DimmContig188_DIMM09370	RNA-dependent RNA polymerase		0.28	0.11	lethal
DimmContig829_DIMM23385	Apurinic/apyrimidinic endonuclease	Anticancer	>1	-0.12	lethal
Glycosylation and sugar metabolism					
DimmContig417_DIMM15590	dTDP-4-dehydrorhamnose3,5-epimerase		0.23	0.04	lethal
DimmContig51_DIMM03355	Beta-1,4-mannosyltransferase	Pesticide	0.95	0.94	lethal
DimmContig3217_DIMM44525	UDP-galactopyranose mutase	Antimycobacterial	0.36	0.08	molt defective
DimmContig2055_DIMM36945	Chitin synthase	Insecticide	-3.52	-4.00	lethal
Lipid metabolism					
DimmContig5517_DIMM52545	Lipase		0.08	-1.60	lethal
DimmContig338_DIMM13735	Sterol-C24-methyltransferase (Erg11)	Antifungal	-3.10	-4.00	lethal
DimmContig1191_DIMM28365	Methyltransferase		-0.03	-0.15	lethal
Transport					
DimmContig1379_DIMM30520	Aquaporin		-2.05	-0.52	lethal
Signal transduction					
DimmContig333_DIMM13570	Nuclear hormone receptor		-2.40	-4.52	lethal
DimmContig245_DIMM11140	G-protein coupled receptor	Many indications	-1.06	-0.59	lethal
DimmContig1549_DIMM32395	G-protein coupled receptor	Many indications	-1.26	-1.57	lethal
DimmContig2384_DIMM39455	G-protein coupled receptor	Many indications	-2.70	-1.96	lethal
DimmContig335_DIMM13610	Groundhog protein	Anticancer	>1	0.04	lethal
DimmContig3787_DIMM47150	Warthog protein	Anticancer	>1	0.78	lethal
DimmContig48_DIMM03200	Warthog protein	Anticancer	>1	0.32	lethal
DimmContig253_DIMM11405	Haloacid dehalogenase-like hydrolase		-3.22	-4.00	lethal
DimmContig326_DIMM13435	Apoptosis regulator ced-9	Anticancer	0.77	-1.02	lethal

The filtering criteria are relevant as a method to discover potential new drug targets. However, setting filters can have disadvantages; interesting targets can be missed. For example none of the LGIC receptors came out of the research. This is due to their relative similarity to the human and dog proteome sequences; knowing that the best hits to these proteomes were retrieved. LGIC are nevertheless important anthelmintic targets. In mammals, LGIC are present behind the cerebral barrier and are not reachable for anthelmintic compounds while nematodes LGIC are well exposed to drugs.

Anyway, about 400 potential drug targets came out from this filtering research. Over this large number of targets 298 are functionally unknown proteins. These putative proteins are currently unhelpful to discover new anthelmintics.

Focusing on the potential novel drug targets with known properties interesting conclusions might point out.

Over the 20 novel potential drug targets described (table 4), five main groups emerged; nucleic acid synthesis and repair, glycosylation and sugar metabolism, lipid metabolism, transport, and signal transduction.

Chapter 5. Drug sensitivity tests (unpublished)

While there was no candidate target of monepantel in the predicted proteome of *D. immitis*, we identified by comparative genomics to *C. elegans* and the dog, several other candidate drug targets (see Table 4, page 118). Of these, RNA-dependant RNA polymerase and chitin synthase were studied in more detail.

RNA-dependant RNA polymerase

An RNA-dependant RNA polymerase is encoded in the genome of the heartworm, so there might be activity against *D. immitis* of guanosine analogs, known inhibitors of the viral RNA-dependant RNA polymerase. To test this hypothesis, in vitro tests were performed with different guanosine analogs. 21 compounds, diluted in dimethyl sulfoxide (DMSO), were tested against microfilariae of *D. immitis* in vitro and the efficacy was evaluated after 48 hours of incubation. The maximum concentration value tested was 32 ppm. Positive and negative controls were included in the study. None of the compound showed activity against the parasite tested (Figure 17). Compound 9 was the most active of the guanosine analogs and had an efficacy value of 16%.

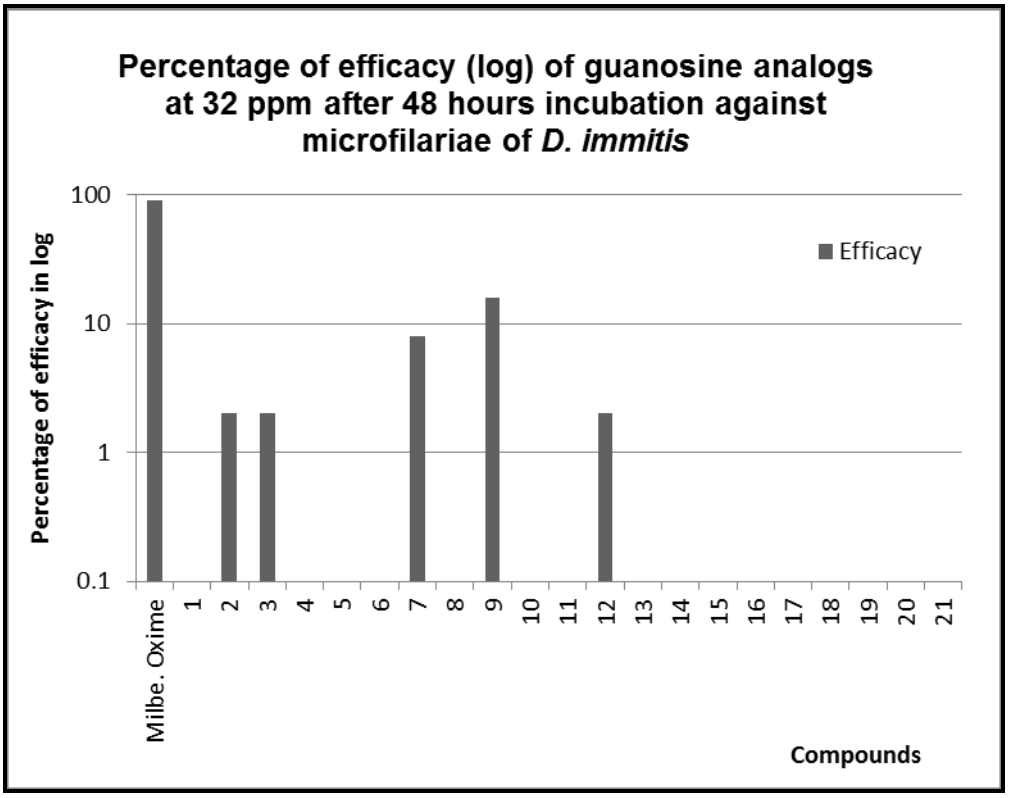


Figure 17. Percentage of efficacy (absence of motility), in logarithm, of 21 guanosine analogs against *D. immitis* microfilariae at 32 ppm after 48 hours of incubation. Test performed in triplicates.

Chitin synthase

Derivatives of lufenuron are commercially available and they are mainly used as insecticides. Seven putative chitin synthase inhibitors were tested in vitro against *D. immitis* microfilariae. Figure 18 illustrates the percentage of efficacy of lufenuron and 6 derivatives against *D. immitis* microfilariae after 48 hours of in vitro incubation. Ivermectin and milbemycin oxime were included as controls. A good efficacy is considered to be between 80 and 100%.

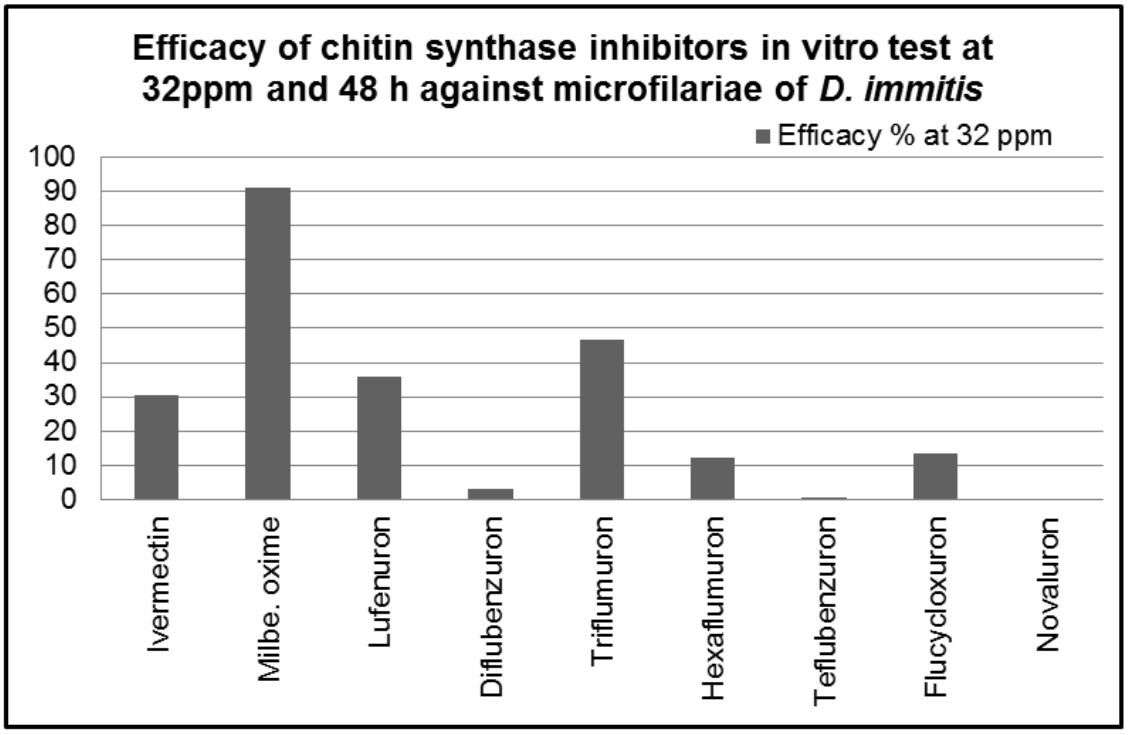


Figure 18. Percentage of efficacy at 32 ppm (particle per million) of in vitro tests with Lufenuron and six derivatives against *D. immitis* microfilariae after 48 hours incubation. Test performed in triplicates.

No significant efficacy was seen in this in vitro test. The control, milbemycin oxime, is good; efficacy reaching 91% after 48 hours at 32 ppm. Ivermectin showed 30.6% of efficacy; however this is a known phenomenon in microfilarial in vitro screening due to a paralysis of the larvae. The lufenuron analogs have a lower percentage of efficacy than lufenuron; except triflumuron which had an efficacy of 46.6% at 32 ppm after 48 hours incubation. The microfilarial in vitro test may not be suitable for efficacy testing of chitin synthase inhibitors. These could be assumed to exhibit maximal activity during molting of the subsequent larval stages and development of eggs in adult females.

Potential novel receptors of *Wolbachia*, endosymbiont of *D. immitis*

An in silico study was performed on the *Wolbachia* genome of *D. immitis* to identify potential novel drug targets. Table 5 summarizes candidate targets with annotation in *Escherichia coli* and corresponding sequences of *Wolbachia*.

Table 5. Potential drug targets of the heartworm endosymbiont, *Wolbachia pipientis*. As identified by the following exclusion criteria: (i) presence of orthologs in *E. coli* which are essential; (ii) absence of a blastp hit in the predicted proteomes of *H. sapiens* and *C. lupus familiaris* with an E-value below 10^{-5}; (iii) predicted function as an enzyme or receptor based on the annotation of the *E. coli* ortholog and blastp searches against SwissProt.

wDi protein	logE (dog)	logE (human)	Ecoli ortholog	Ecoli annotation
fig|82301.4.peg.250	-0.39	-0.46	FTSA	Essential cell division protein FtsA
fig|82301.4.peg.329	-0.74	-1.68	FTSZ	Essential cell division protein FtsZ
fig|82301.4.peg.797	-2.40	-2.22	ACPS	Holo-[acyl-carrier-protein] synthase
fig|82301.4.peg.229	-3.70	-5.05	BIRA	Bifunctional biotin-[acetyl-CoA-carboxylase] ligase and transcriptional repressor
fig|82301.4.peg.765	0.00	0.34	FABZ	3-hydroxy-acyl-[acyl-carrier-protein] dehydratase
fig|82301.4.peg.97	-0.62	0.28	DNAA	Chromosomal replication initiator protein
fig|82301.4.peg.645	-2.05	-1.96	DNAB	Replicative DNA helicase
fig|82301.4.peg.298	-0.46	0.74	RPOA	RpoA is the α subunit of the RNA polymerase core enzyme
fig|82301.4.peg.817	-1.46	>1	TRMD	tRNA (m^1G37)methyltransferase catalyzes the addition of a methyl group to G37 of certain tRNAs
fig|82301.4.peg.753	-5.22	-5.15	MRAY	Phospho-N-acetylmuramoyl-pentapeptide transferase catalyzes the first step of the lipid cycle reactions in the biosynthetic pathway of cell wall peptidoglycans
fig|82301.4.peg.796	-0.27	0.92	MURA	UDP-*N*-acetylglucosamine enolpyruvoyl transferase (MurA) catalyzes the first committed step in the assembly of peptidoglycan
fig|82301.4.peg.490	-0.46	0.23	MURB	*murB* encodes the UDP-*N*-acetylenolpyruvoylglucosamine reductase which catalyzes the second committed step in the biosynthesis of peptidoglycan
fig|82301.4.peg.810	0.20	0.64	MURC	UDP-N-acetylmuramate-alanine ligase is responsible for the addition of the first amino acid of the peptide moiety in the assembly of the monomer unit of peptidoglycan
fig|82301.4.peg.34	-0.89	-0.60	MURD	Biosynthesis of peptidoglycan
fig|82301.4.peg.12	-0.12	0.86	MURE	Biosynthesis of peptidoglycan
fig|82301.4.peg.706	-0.06	0.66	MURG	*murG* gene codes for the N-acetylglucosaminyl transferase responsible for the final intracellular step of peptidoglycan subunit assembly
fig|82301.4.peg.79	-0.80	-0.42	MVIN	MurJ is an essential gene involved in murein synthesis as the flippase responsible for transport of lipid II from the inner face of the inner membrane to the outer face for polymerization
fig|82301.4.peg.556	0.23	0.36	SECA	Inner membrane component of the Sec Protein secretion system
fig|82301.4.peg.488	0.20	0.48	SECD	Component of the Sec Protein secretion system
fig|82301.4.peg.154	-0.23	-0.96	SECF	Component of the Sec Protein secretion system
fig|82301.4.peg.294	-0.19	0.18	SECY	Component of the Sec Protein secretion system
fig|82301.4.peg.499	-0.06	0.46	YFIO	BamD as forming a large protein complex with BamB involved in outer membrane protein (OMP) biogenesis
fig|82301.4.peg.621	0.68	>1	LOLE	Membrane component of the LolCDE lipoprotein transporter, ABC transporter
fig|82301.4.peg.512	0.85	>1	RIBB	RibB catalyzes the synthesis of the 4-carbon compound 3,4-dihydroxy-2-butanone-4-phosphate, a precursor of the terminal intermediate, 6,7-dimethyl-8-ribityllumazine, in the biosynthesis of riboflavin

A selection of antibiotics was tested in the same microfilarial in vitro assay (Figure 19).

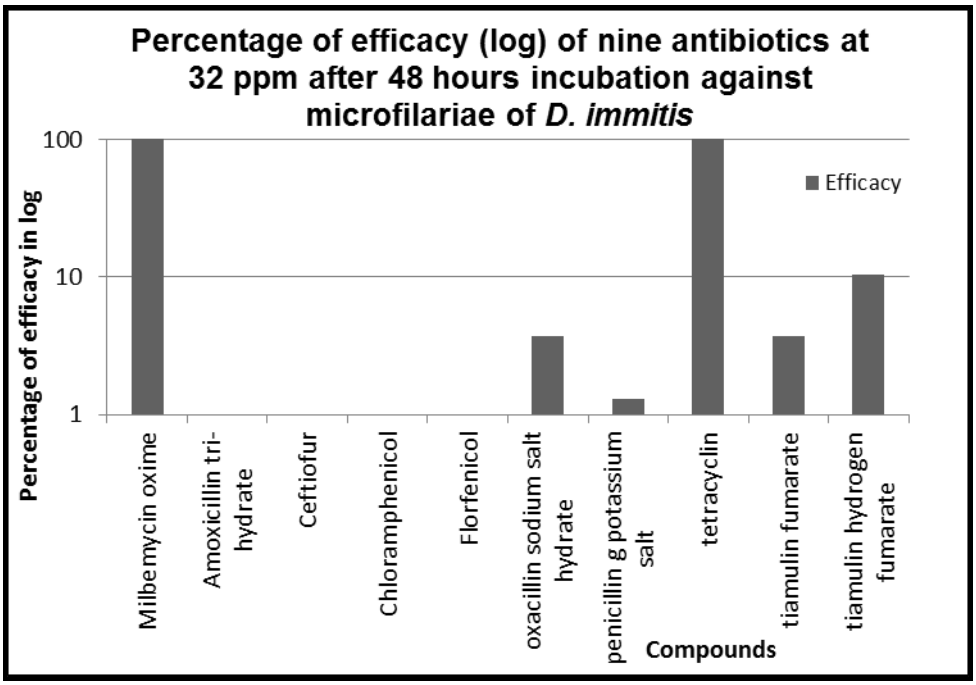


Figure 19. Percentage of efficacy, logarythm scale, of nine antibiotics potentially targeting *Wolbachia* of *D. immitis.* Test performed against *D. immitis* microfilariae at 32 ppm after 48 hours of incubation and in triplicates.

Only tetracyclin exhibited 100% efficacy. Four other compounds (oxacilin sodium salt, penicillin G potassium salt, tiamulin fumarate and tiamulin hydrogen fumarate) were slightly active with, respectively, 3.7, 1.3, 3.7 and 10.3% efficacy.

Discussion

The initial hypothesis of the PhD thesis was that *D. immitis* possesses an MPTL-1 ortholog because it shows susceptibility to AADs in vitro. The aim had been to discover that ortholog, and to identify LGIC in general. For this purpose, we sequenced the *D. immitis* genome. Having the full genome sequence in hands, deeper investigations were performed to identify and characterize LGIC in silico based on phylogenetic comparisons. Based on the obtained results, potential drug targets of *D. immitis* were identified. However, not a single member of the DEG-3 subfamily of acetylcholine receptors, which contains the target of monepantel, was found; except for a pseudogene, corroborating the hypothesis that these receptors were lost during the evolution of *D. immitis*. The most interesting targets were compiled in Table 4 (page 119) and further investigations were performed to correlate drug susceptibility with the presence or absence of target(s). The focus was put on enzymes involved in (1) the nucleic acid synthesis and repair, and (2) sugar metabolism and glycosylation.

1. Nucleic acid synthesis and repair receptors

RNA-dependant RNA polymerase

An RNA-dependant RNA polymerase is an enzyme that catalyses the synthesis of an RNA complementary strand from an RNA template (127). The protein is essential in RNA viruses, and all known RNA viruses encode RNA-dependent RNA polymerase. RNA viruses are divided into three major classes; SSRNA positive strand, SSRNA negative strand, and dsRNA genome (128). RNA-dependent RNA polymerases are crucial for the replication of all 3 classes RNA viruses. The *C. elegans* enzyme has a lethal RNAi phenotype (166).

The RNA-dependent RNA polymerase of *D. immitis* may be a novel potential anthelmintic target, since there is no mammalian ortholog.

Unfortunately, however, there are no known inhibitors of metazoan RNA-dependant RNA polymerase. Thus we tested inhibitors of the viral enzymes, guanosine analogs in particular. Acyclovir is an important antiviral low cytotoxicity and a good specificity against herpes zoster (shingles), genital and labial herpes, and chickenpox. Acyclovir molecule after absorption by the host cell is converted to the monophosphate by viral thymidin kinase; followed by conversion to the triphosphate form, acido-GTP, which

is a powerful inhibitor of the viral RNA-dependent RNA polymerase (128). Aciclovir and all other antivirals tested were inactive against *D. immitis* microfilariae (Figure 17 page 122, Chapter 5).

This could be explained by different reasons. RNA-dependent RNA polymerase might not be essential, at least not at this life stage or it might be insensitive to guanosine analogs. In summary, eukaryotic RNA-dependant RNA polymerases might be attractive new antiparasitic drug targets, but proof of principle is lacking as there are no validated inhibitors.

2. Glycosylation and sugar metabolism

Glycosylation and sugar metabolic enzymes represent another encouraging group of potential novel drug targets. Among this category, three predicted enzymerase of particular interest: beta-1,4-mannosyltransferase, UDP-galactopyranose mutase and chitin synthase. The *C.elegans* RNAi phenotype is lethal for two of them while for the UDP-galactopyranose mutase the RNAi phenotype is molt defective (which would suffice for a heartworm preventive).

Beta-1,4-mannosyltransferase

In *C. elegans*, beta-1,4-mannosyltransferase is alternatively named *Bacillus thuringiensis* toxin-resistant, BRE-3. Cry5B toxin is produced by *B. thuringiensis*, NP499087.2 (130).

BRE-3 belongs to the glycosyltransferase-2 family (E.C. 2.4.1 group of hexosyltransferase) (131). Its proposed role is in glycosphingolipid biosynthesis and it is involved in susceptibility to pore-forming toxins in conjunction with BRE-1, BRE-2, BRE-4 and BRE-5 (132).

The Gram-positive *B. thuringiensis* is a soil-dwelling bacterium that was discovered by the Japanese biologist S. Ishiwatari in 1901. In 1911 E. Berliner, a German biologist, rediscovered the bacteria searching for a cause of a disease in caterpillar (132-133). These aerobic bacteria are able of producing endospores. The production of crystal proteins, δ-endotoxins with insecticidal activity, occurs during sporulation. However, many crystal-producing *B. thuringiensis* strains do not have insecticidal properties (134).

Beta-1,4-mannosyltransferase builds the target of the *Bt* toxins (*B. thuringiensis*) and an ortholog of *bre-3*, is encoded in the genome of *D. immitis*, meaning that the target is likely present.

Extracted from bacteria, these *Bt* toxins are mainly used in agriculture as natural insecticides. When the insects ingest the Cry proteins, they insert into the membrane and form a pore in the intestine wall of the insect gut. The pore results in cells lysis and potential death of the insect within 2 to 5 days (135). Bre-3 encodes a protein similar to beta-glycosyltransferases from bacteria that is required for the toxicity of *B. thiuringiensis* Cry5B. *B. thuringiensis* products have a narrow spectrum of activity and are considered non-toxic to humans, other mammals and plants and of less damage than other insecticides (136).

B. thuringiensis expresses crystal protein toxins of various types. The number of known crystal proteins, Cry and Cyt proteins has drastically increased in the last decades to over 600 different δ-endotoxins (137-139). Amongst these, a few seems to have nematicidal properties.

Already in 1987, Bottjer and Bone described morphological changes of eggs and juveniles, and death in the gastro-intestinal nematode *T. colubriformis* after immersion in a crystal-enriched preparation of *B. thuringiensis israelensis* (133). Later on, Edwards et al. from Mycogen, patented novel isolates of *B. thuringiensis* containing toxin active against adult nematodes and larvae (139). Between 1996 and 2010, various authors described the *B. thuringiensis* δ-endotoxins having nematicidal effects on *C. elegans* (140-143), *Ancylostoma ceylanicum* (135) and *Heligmosomoides bakeri* (144). Two other publications (145-146) reported the activity of the *Bt* toxins against diverse nematodes.

Cry proteins must be ingested to intoxicate. In the case of *A. ceylanicum*, it is not surprising that none-feeding stages are not susceptible to Cry5B toxicity; the other stages being susceptible (135). Diverse crystal proteins show nematicidal activity, Cry5, Cry6, Cry12, Cry13, Cry14 and Cry21. The Cry5B crystal protein has been shown to be therapeutic in vivo against the intestinal hookworm, *A. ceylanicum*, in hamsters when delivered daily per os over the course of 3 days (135). Aroian R. et al. (147) showed that Cry5B had significant anthelmintic activity in rodent models against two parasitic nematodes, *H. bakeri* and *Trichuris muris*. The Cry5B is a huge protein with 150kDa and it appears to be rapidly degraded in the stomach with less

than 1% reaching the parasite. There is a quick release of the product in the mammalian body and within two hours the protein is dispersed (147).

In the case of *D. immitis*, very little information is known about the feeding processes. *D. immitis* feed primarily on blood or plasma in and around the heart and lungs. Would the heartworm ingest δ-endotoxins and could they be used as anthelmintics? Currently, we do not know anything on the efficacy of *B. thuringiensis* toxins against *D. immitis*. The best way would be to test Cry5B in vitro against microfilariae and L_3 of *D. immitis*. As a positive control, Cry5B should be tested in the LDA against *Haemonchus contortus,* a blood feeding gastro-intestinal nematode. Before testing the protein, the nematicidal spore-crystal mixture of *B. thuringiensis* needs to be obtained. The preparation of the crystal protein was described in Borgonie et al. and might be feasible (140).

UDP-galactopyranose mutase

The second enzyme of interest in the glycolysation and sugar metabolism group is UDP-galactopyranose mutase (E.C. 5.4.99.9), a known antimycobacterial target.

UDP-galactopyranose mutase is a flavoenzyme involved into bacterial cell wall biosynthesis by catalyzing the conversion of UDP-D-galactopyranose to UDP-D-galacto-1,4-furanose (148). This product is found in bacterial cell walls, including those of many pathogens (149). UDP-D-galacto-1,4-furanose is an essential sugar for the growth of pathogenic bacteria and survival in the host. Importantly, the enzyme has not been found in mammals. The D-galactofuranose is also a component of the cell envelope of eukaryotic parasites, as *Leishmania* spp, *Trypanosoma cruzi*, and of a number of different fungal species including *Penicillium* spp and *Aspergillus* spp (150).

UDP-galactopyranose mutase homologs are found in most of the nematode clades, suggesting an ancient and conserved feature of the phylum. Novelli et al. (151) described a detailed genetic and molecular analysis of the *C. elegans* homolog, *glf-1*. They demonstrated galactofuranose synthesis in the free-living nematode and its importance in the maintenance of the integrity of the coat surface. Additionally, genome-wide RNAi screens showed the roles of the synthesis in nematode survival, in late larval development, and in the completion of embryogenesis (151). UDP-galactopyranose mutase was also proposed as a target against *B. malayi* (126). An

Australian patent describes an UDP-galactopyranose mutase derived from parasitic nematodes and specifically the gastro-intestinal nematode *H. contortus* as potential drug target (152).

A first step forward validation of UDP-galactopyranose mutase as a drug target of *D. immitis* would be in vitro tests with proven UDP-galactopyranose inhibitors. Diverse studies have been performed to identify inhibitors of UDP-galactopyranose mutase. Partha et al. identified three best groups of inhibitors, the substituted thiazolidinones, the aminothiazoles and the pyrazoles (153).

Thiazolidinones are derivatives of thiazolidine. They belong to an important group of heterocyclic compounds with sulfur and nitrogen in a five-membered ring. The different derivatives have various biological activities (154-155), including antibacterial, antitumor, antihistaminic, anti-inflammatory, and anticonvulsant (156). One of the best studied groups is the 4-thiazolidinones; derivatives of thiazolidine with a carboxyl group at the 4 position (155).

2-aminothiazoles are small molecules of heterocyclic amine with antiprion activity. Dykhuizen et al. show that 2-aminothiazoles target the UDP-sugar binding site of UDP-galactopyranose mutase and validate UDP-galactopyranose mutase as an antimycobacterial target (157). 2-aminothiazoles are not yet commercially available.

Last group of potential inhibitors of UDP-galactopyranose mutase are the pyrazoles also known as 1,2-diazoles. The basic structure is represented by simple aromatic, 5-membered ring composed of three carbon and two nitrogen atoms in adjacent positions. Biological properties of pyrazoles are wide; playing a role in inflammation, diabetes, as antibacterials, analgesics, for anti-arrhythmia, as tranquilizers, for muscular soothing and so on (158). Again pyrazoles are not easily commercially available.

In conclusion, UDP-galactopyranose mutase could be an attractive antifilarial target that could be validated by thiazolidinones, 2-aminothiazoles or pyrazoles. However, these would be needed to be synthesized.

Chitin synthase

Chitin synthase (E.C. 2.4.1.16) catalyzes the polymerization of UDP-N-acetyl-β-D-glucosamine to give the insoluble biopolymer chitin, poly-β(1,4)-GlucNAc. This natural amino-polysaccharide is synthesized by certain invertebrates such as fungi,

parasitic nematodes, insects and protozoa. The interest of the chitin synthesis is the fundamentality of the process in many of these organisms while it is totally absent in mammals. Consequently, chitin synthase is an attractive target for drug discovery (159). Currently, chitin synthase is mainly used as insecticide target. Chitin is an important component of the insect exoskeleton and nematode egg shell (160). In the viviparous *D. immitis*, the embryo hatches from the eggshell into the maternal uterus. Microfilariae are shed from the female without extracuticular structures (161). Even with this pattern of embryogenesis the *chs-1*, chitin synthase, gene is present in the genome. One proposed inhibitor of chitin synthase is lufenuron. Lufenuron is the active ingredient of the flea control medication Program®. This compound breaks the flea life cycle by inhibiting the development of the egg, interfering with chitin synthesis and deposition. At therapeutical doses, lufenuron has no effects against adult fleas. Lufenuron is stored in the body fat of the animal and transferred to the adult flea through their bite. The product reaches the eggs through the blood circulation or by the larva feeding on pre-digested blood. Lufenuron is also efficacious in crop protection against lepidoptera, mites and thrips; additionally it is effective as an anti-fungal in plants. It is non toxic to mammals (162). In the medication Program plus®, Lufenuron is combined with milbemycin oxime, a ML, against larval stages of *D. immitis* infections as well as fleas. Program plus® is given in dogs against ecto- and endo-parasites (163).

3. Parallel study performed on *Wolbachia* of *D. immitis*

Potential target were also predicted from the proteome of *D. immitis* endosymbiont *Wolbachia*. Table 5 (page 124, Chapter 5) summarizes the potential drug targets of *Wolbachia* that are essential for *Escherichia coli*.

Wolbachia are alphaproteobacteria showing a mutualistic relationship with filarial nematodes. Related to Rickettsia and Ehrlichia, the obligate endosymbiont is located in most tissues of the nematodes and mainly in hypodermal cords. The endosymbionts' transmission occurs between the filarial adult female and its offspring through oocytes. Tetracycline is effective against *O. volvulus* and raised hopes for further antibiotics active against filariae (164). Without their endosymbiont, filariae show a loss of fertility and viability (165). However, the endosymbiont might be difficult to reach as further membranes would need to be crossed. More importantly,

the prophylactic use of antibiotics in dogs should be discouraged because of the inevitable risk of selecting for drug-resistant bacteria.

No further studies were performed based on the data shown in the Table 5 (page 124, Chapter 5).

Conclusions

Being aware of the presence or absence of the receptors in the target parasite can be very helpful to select novel drugs. For once the mode of action will be known before the effect of the drug.

The genome also brought answers concerning the effect of known anthelmintics as macrocylic lactones, benzimidazoles, imidazothiazoles, cyclodepsipeptides, and amino acetonitriles derivatives by revealing the presence or absence of specific receptors to these anthelmintics.

The unknown target list has a greater number of candidates with unknown functions. Further researches are expected to decrease the number of unknown targets. Simple modifications can be mentioned as changing the filters or add some more, to reduce the number of potential targets in the unknown list; fit new filters as molecular weight >100kDa, transmembrane domains, specific domains, paralogs, mutation-non mutation, and check for known domains in the sequences using interesting databases as target DB and chEMBL,... Eventually, we will end with a reduce list of unknown potential drug targets. Later, a colorimetry assay with *C. elegans* can be set to localize the function of the unknown target. Otherwise, functional screening assays can be performed as cellular assays.

The use of molecular approaches in the race against parasites and anthelmintic resistance might be a great advantage.

References

1. Blaxter M. L. et al. (1998) A molecular evolutionary framework for the phylum Nematoda *Nature* 392; 71-75
2. De Ley P. (2006) A quick tour of nematode diversity and the backbone of nematode phylogeny *WormBook* ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.41.1, http://www.wormbook.org.
3. Chabaud G. A. (1994) The evolutionary expansions of the Spirurida *International Journal for Parasitology* 24: 1179–1201
4. Lok J.B. (1988) Dirofilaria sp: taxonomy and distribution In Boreham, P. F L. and R. B. Atwell, editors. (Editors). Dirofilariasis. CRC Press. Boca Raton, FL.; 1–28
5. Anderson R. C. (1957) The life cycles of Dipetalonematid nematodes (Filaroidea, Dipetalonematidae): The problem of their evolution *Journal of Helminthology* 31(4): 203-224
6. Trotti C. G., Pampiglione S. & Rivasi F. (1997) The species of the genus Dirofilaria, Raillet & Henry, 1911 *Parassitologia* 39 (4); 369-374
7. Godel C. et al. (2012) The genome of the heartworm, *Dirofilaria immitis* Unpublished
8. Grieve R.B et al. (1983) Epidemiology of canine heartworm infection *Epidemiologic reviews* 5; 220-243
9. Knight D. H. (1987) Heartworm infection *Vet. Clinics of North America: small animal practice* 17 (6); 1463-1518
10. American Heartworm Society | PO Box 8266 | Wilmington, DE 19803-8266 | E-mail: info@heartwormsociety.org
11. Bowman D. D. (2010) Presentation at: State of the Heartworm Symposium 2010, Memphis, USA Cities, Countis, Countries, Katrina: The continuing course of Heartworm – Reflections from the Northern US
12. Genchi C. et al. (2010) Changing climate and changing vector-borne disease distribution: the example of *Dirofilaria* in Europe, American Heartworm Symposium Presentation, Memphis.
13. Marinculic A. et al. (2010) The spreading of *Dirofilaria* infections in Croatia", American Heartworm Symposium Presentation, Memphis.
14. Upchurch D. A. et al. (2010) Comparison of radiography, histology, and serology findings for the diagnosis of pulmonary disease: Study of 120 random source cats, American Heartworm Symposium Presentation, Memphis.
15. Venco L. et al. (2010) USA Can Heartworm prevalence in dogs be used as provisional data for assessing the prevalence of the infection in cats? American Heartworm Symposium Presentation, Memphis.
16. Wikipedia the free encyclopedia, "Heartworm", http://en.wikipedia.org/wiki/Heartworm.
17. Kartman L. (1957) The vector of canine filariasis; a review with special reference to factors influencing susceptibility *Rev Bras Malariol Doencas Trop* 9 (5); 3-47

18. Sousby E.J.L. (1965) Textbook of veterinary clinical parasitology, volume I Helminths *Blackwell scientific publications oxford*
19. Anderson R.C.(1992) Nematode parasite of vertebrates, their development and transmission *CAB International*
20. Kotani T. & Powers K. G. (1982) Developmental stages of *Dirofilaria immitis* in the dog *Am J Vet Res* 43;, 2199-2206
21. Kume S. & Itagaki S. (1955) On the life-cycle of *Dirofilaria immitis* in the dog as the final host *Br. Vet. J* 111; 16-24.
22. Grandi C. (2010) Novel Adulticide therapy for Dirofilaria immitis infection: a clinical case in a naturally infected dog, American Heartworm Symposium Presentation, Memphis.
23. Kramer L. (2010) Evaluation of lung pathology in dogs treated with doxycycline or a combination of doxycycline and ivermectin before administration of melarsomine dihydrochloride (Immiticide®), American Heartworm Symposium Presentation, Memphis
24. McCall J. W. (2005) The safety-net story about macrocyclic lactone heartworm preventives: A review, an update, and recommendations *Veterinary Parasitology* 133; 197-206
25. Bourguinat C. et al. (2010) Investigation of genetic changes in Dirofilaria immitis after the use of macrocyclic lactone heartworm preventives, *in Proceedings*. 13th Triennial Heartworm Symp 2010; 28
26. Patton S. et al. (2010) Survey of heartworm prevention practices among dog owners and trainers in North America, American Heartworm Symposium Presentation, Memphis.
27. Bourguinat C. et al (2011) Macrocyclic Lactone resistance in *Dirofilaria immitis Veterinary Parasitology* 181; 388-392
28. Kaminsky R. et al. (2008) A new class of anthelmintics effective against drug-resistant nematodes *Nature* 411; 176-180
29. Rufener L. et al. (2010) Phylogenomics of ligand-gated ion channels predicts monepantel effect *PLoS pathogens* 6 (9):1-11.
30. Mitreva M. et al. (2005) Comparative genomics of nematodes", Trends in Genetics, 21 (10), 573-581
31. Yin Y. et al. (2006) Identification and analysis of genes expressed in the adult filarial parasitic nematode *Dirofilaria immitis International Journal of Parasitology*, 36; 829-839
32. Ghedin E. (2007) Draft genome of the filarial nematode parasite *Brugia malayi Science* 317 (5845):1756-1760
33. Mitreva M. & Jasmer D.P. (2006) Biology and genome of *Trichinella spiralis WormBook* ed. The *C. elegans* Research Community WormBook, doi/10.1895/wormbook.1.124.1, http://www.wormbook.org.
34. Alberts B et al. (2000) Molecular biology of the Cell fourth Edition *GS Garland Science Taylor and Francis Group* 191-234.
35. Obenrader S. The Sanger Method, webpage produced as an assignment for an undergraduate course at Davidson College.

36. Sanger F & Coulson A.R. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase *J. Mol. Biol.* 94; 441-448.
37. Sanger F. et al. (1977) DNA sequencing with chain-terminating inhibitors *Proc. Natl. Acad. Sci. USA* 74 (12);5463-5467.
38. Sanger F. (1980) Determination of nucleotide sequences in DNA, Nobel lecture *Bioscience Reports* 3-18.
39. Maxam A. M.& Gilbert W. (1977) A new method for sequencing DNA *Proc. Natl. Acad. Sci. USA* 74 (29); 560-564.
40. Gilbert W. (1980) DNA sequencing and gene structure, Nobel lecture.
41. Nobel Prize webpage: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1980/
42. Wikipedia,http://en.wikipedia.org/wiki/List_of_sequenced_eukaryotic_genomes
43. Sanger F. (1975) The Croonian Lecture, 1975: Nucleotide sequences in DNA *Proc. R. Soc. Lond. B. Biol. Sci.* 191; 317-333
44. Smith et al. (1986) Fluorescence detection in automated DNA sequencing analysis *Nature* 312; 674-679
45. Smith A. J. H. (1980) DNA sequence analysis by primed synthesis *Methods in Enzymology* 65; 560-580.
46. Rosenblum B. B. et al. (1997) New dye-labeled terminators for improved DNA sequencing patterns *Nucleic Acids Researches* 25 (22); 4500-4504.
47. Hutchinson C. A. (2007) DNA sequencing: bench to bedside and beyond *Nucleic Acids Researches* 35 (18); 6227-6237, doi:10.1093/nar/gkm688
48. Mardis E. R: (2008) Next-generation DNA sequencing methods *Annu.Rev.Genom. Human Genet.* 387-402
49. Shendure J. & Ji H. (2008) Review. Next-generation DNA sequencing *Nature Biotechnology* 26 (10); 1135-1145
50. Venter C. J. et al. (2001) The sequence of the human genome *Science* 291 (5507); 1304-1351
51. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program Available at: www.genome.gov/sequencingcosts
52. Imai K. et al., (1999) Multi-capillary DNA sequencer *Review* 48 (3);107-109
53. Margulies M, Egholm M, Altman WE, *et al* (2005) Genome Sequencing in Open Microfabricated High Density Picoliter Reactors *Nature* 437 (7057); 376–80. Bibcode 2005 Natur.437..376M. doi:10.1038/nature03959
54. Rothberg J. M. & L Leamon J. H. (2008) The development and impact of the 454 sequencing *Nature Biotechnology* 26 (10);1117-1124
55. Ronaghi M. et al. (1998) DNA sequencing: a sequencing method based on real-time pyrophosphate *Science* 281 (5373); 363-365 doi:10.1126/science.281.5375.363
56. Patrick K. L. (2007) Profile. 454 Life Sciences: Illuminating the future of genome sequencing and personalized medicine *Yale Journal Of Biology And Medicine* 80; 191-194

57. 454 sequencing (2011) GS FLX+ System Sanger-like read lengths – the power of next-gen throughput *ROCHE Diagnostics GmbH*
58. Miller W. et al. (2008) Sequencing the nuclear genome of the extinct woolly mammoth *Nature* 456; 387-390 doi:10.1038/nature 07446
59. Dalton R: (2009) Neanderthal genome to be unveil *Nature* 457(7230); 645
60. Pareek C.S. et al. (2011) Sequencing technologies and genome sequencing *J.Appl. Genetics* 52; 413-435
61. Lin B. et al., (2008) Recent patents and advances in the next generation sequencing technologies *Recent Pat Biomed Eng*. 1; 60–67
62. Lister R. et al. (2009) Next is now: new technologies for sequencing of genomes, transcriptomes and beyond *Curr Opin Plant Biol* doi:10.1010/j.pbi.200811.004
63. Bennett S. (2004) Solexa Ltd Company Profile *Pharmacogenomics* 5 (4); 433-438
64. Solexa (2006) application note: DNA sequencing *Solexa Inc.*
65. Li R. et al (2010) The sequence and de novo assembly of the giant panda genome *Nature* 463; 311-317
66. Illumina webpage: http://www.illumina.com/technology/mate_pair_sequencing_assay.ilmn
67. Fasteris PDF
68. Schadt et al. (2010) A window into third-generation sequencing. *Hum. Mol. Genet.* 19; R227-R240
69. Zhang. J et al. (2011) The impact of next generation sequencing on genomics *J. Genet. Genomics* 38 (3); 95-109
70. Braslavsky et al. (2003) Sequence information can be obtained from single DNA molecules *Proc. Natl. Sci. USA* 100; 3960-3964
71. Harris T. D. et al. (2008) Single-molecule DNA sequencing of a viral genome *Science* 320 (5873); 106-109
72. Ozsolak et al. (2010) Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation *Cell* 143; 1018-1029
73. Tucker T. et al. (2009) Massively parallel sequencing: the next big thing in genetic medicine *The American Journal of Human Genetics* 85; 142-147
74. Petterson E. et al. (2009) Generations of sequencing technologies *Genomics* 93; 105-111
75. Rusk N. (2011) Torrents of sequence *Nat Meth.* 8(1); 44-44
76. Life Technologies, Ion Torrent: Technology how is it made?: http://www.iontorrent.com/technology/
77. Rothberg J. M. (2011) An integrated semiconductor device enabling non-optical genome sequencing *Nature* 474; 348-352
78. Marziali A and Akeson M. (2001) New DNA sequencing methods Annu. Rev. Biomed. Eng. 3, 195-223
79. Metzker M. L. (2010)Sequencing technologies – the next generation Nature reviews Genetics 11, 31-46

80. Hutchinson C. A. (2007) DNA sequencing : bench to bedside and beyond Nucleic Acids Research 35 (18) 6227-6237
81. Microsynth Swiss partner of Roche Diagnostics (Switzerland) for 454 Sequencing, Workflow
82. Microsynth – Your Swiss partner for SOLiD Sequencing, Workflow
83. Illumina HiSeq 2000, http://www.illumina.com/systems/hiseq_2000/applications.ilmn
84. Braslavski et al. (2003)
85. Ozsolak and Milos (2011a)
86. Karow J. (2009) At AGBT, Ion Torrent costumers provide first feedback; Life Tech Outlines Platform's Growth. In Sequence.
87. Helicosbio. http://www.helicosbio.com/SequencingServices/tabid/193/Default.aspx
88. Pacific Biosciences, www.pacificbiosciences.com
89. De novo assembly using Illumina reads, Technical note: Illumina sequencing (2009) *Illumina Inc.*
90. Zhang Y. & Waterman M.S. (2005) An Eulerian path approach to local multiple alignment for DNA sequences *PNAS* 2 (5); 1285-1290)
91. Baumgardner J. et al. (2009) Solving a Hamiltonian problem with a bacterial computer *Journal of Biological Engineering* 3; 11
92. Falquet L. (2011) Next generation sequencing: *de novo* genome assembly *Course Lausanne Swiss Institute of Bioinformatics*
93. Zerbino D. R. (2010) Using the Velvet de novo assembler for short-read sequencing technologies *Current Protocols in bioinformatics* Chapter 11; 11.15
94. Opperman C. H. (2008) Sequence and genetic map of Meloidogyne hapla : A compact nematode genome for plant parasitism *PNAS* 105(39): 14802-14807
95. C. elegans Seq Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology *Science* 282: 2012-2018.
96. Hillier L. W. et al. (2008) Whole-genome sequencing and variant discovery in *C. elegans Nature Methods* 5 (2): 183-188.
97. Jex A. R. et al. (2011) *Ascaris suum* draft genome *Nature Letters* doi:10.1038/nature10553.
98. Sanger Institute: http://www.sanger.ac.uk
99. Unnasch T. R. & Williams S. A. (2000) The genomes of Onchocerca volvulus *International Journal for Parasitology* 30:543-552.
100. Lindblad-Toh K, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature. 2005 Dec 8;438:803-19
101. Broad Institute http://www.broadinstitute.org/annotation/genome/filarial_worms/GenomeStats.html
102. Arensburger P. et al. (2010) Sequencing of Culex quiquefasciatus establishes a platform for mosquito comparative genomics *Science* 330:86-87

103. Behura S. K. et al. (2011) Comparative genomic analysis of *Drosophila melanogaster* and vector mosquito developmental genes *PLoS one* 6(7):1-19
104. Holt. R. A. et al. (2002) The genome sequence of the malaria mosquito *Anopheles gambiae Science* 298:129-149
105. Nene V. et al. (2007) Genome sequence of *Aedes aegypti*, a major arbovirus vector *Science* 316 (5832):1718-1723
106. Hellsten U. et al. (2010) The genome of the western clawed frog *Xenopus tropicalis Science* 328 (5978): 633-636
107. Broughton R. E et al. (2001) The complete sequence of the Zebrafish (Danio rerio) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA *Genome Res.* 11 (11): 1958-1967.
108. Ghedin E. (2004) First sequenced genome of a parasitic nematode *Trends in Parasitology* 20 (4): 151-153
109. Pontius J. U. et al. (2007) Initial sequence and comparative analysis of the cat genome *Genome research* 17:1675-1789.
110. International Human Genome Consortium (2001) Initial sequencing and analysis of the human genome *Nature* 409:860-921
111. Severson D. W. et al. (2004) *Aedes aegypti* genomics *Insect biochemistry and molecular biology* 34:715-721
112. Koboldt D. C. et al. (2010) Challenges of sequencing human genomes *Brief Bioinform.*11 (5) : 484-498.
113. Simpson J.T. et al. (2009) ABySS: a parallel assembler for short read sequence data *Genome Res* 19(6); 1117-1123
114. Hucho F. and Weise C. (2001) Ligand-gated ion channels *Angew. Chem. Int. Ed.* 40; 3100-3116
115. Alberts B et al. (2000) Molecular biology of the Cell fourth Edition *GS Garland Science Taylor and Francis Group;* 630-656
116. Gay E. A. & Yakel J. L. (2007) Gating of nicotinic ACh receptors; new insights into structural transitions triggered by agonist binding that induce channel opening *J Physiol* 584.3; 727-733
117. Itier V. & Bertrand D. (2001) Neuronal nicotinic receptors: from protein structure to function *FEBS Letters* 504; 118-125
118. Elgoyhen et al. (2001) *Proc. Natl. Acad. Sci. USA* 98; 3501-3506
119. Sargent P.B. (1993) Next to the cys-loop the nAChR present four specific, constant transmembrane domains (TMI-TMIV) *Annu. Rev. Neurosci.* 16; 403-443
120. Williamson S. et al. (2007) The cys-loop ligand-gated ion channel gene family of Brugia malayi and Trichinella spiralis: a comparison with Caenorhabditis elegans *Invert Neurosci* 7;219-226
121. Rufener L. et al. (2009) *Haemonchus contortus* acetylcholine receptors of the DEG-3 subfamily and their role in sensitivity to monepantel *PLoS pathogen* 5 (4); 1-11

122. Perret C. (2008) Development of an in vitro test with model parasite of gerbils, *Acanthocheilonema viteae*, and target parasite *Dirofilaria immitis* affecting dogs *Master Thesis.*
123. Gill J. H. et al. (1995) Avermectin inhibition of larval development in Haemonchus contortus – Effects of ivermectin resistance *Int J Parasitol* 25; 463-470
124. Jones A. K. & Sattelle D.B. (2008) The cys-loop ligand-gated ion channel gene superfamily of the nematode, Caenorhabditis elegans *Invert Neurosci* 8; 41-47
125. Bourguinat C. & Prichard R., unpublished data, McGill University
126. Kumar S. et al. (2007) Mining predicted essential genes of *Brugia malayi* for nematode drug targets *PLoS One* 11; 1-9
127. Wikipedia: http://en.wikipedia.org/wiki/RNA-dependent_RNA_polymerase
128. Ahlquist P. (2002) RNA-dependent RNA polymerases, viruses, and RNA silent *Science* 296; 1270-1273
129. Drugbank: www.drugbank.ca/drugs/DB00787
130. Griffits J. L. et al. (2005) Glycolipids as receptors for Bacillus thuringiensis crytal toxin *Science* 307; 922-925
131. Uniprot: http://www.uniprot.org/uniprot/Q03562)
132. Cheng T. C. (1984) Pathogens of invertebrates: application in biological control and transmission mechanisms, *Society for Invertebrate Pathology Meeting* 7; 159
133. Zakharyan R.A *et. el.* (1979). "Plasmid DNA from Bacillus thuringiensis". Microbiologiya 48 (2); 226–229. ISSN 0026-3656
134. Bottjer K. P. & Bone L. W. (1987) Changes in Morphology of *Trichostrongylus colubriformis* eggs and juveniles caused by *Bacillus thuringiensis israelensis Journal of Nematology* 19 (3); 282-286
135. Capello M. et al. (2006) A purified *Bacillus thiringiensis* crystal protein with therapeutic activity against the hookworm parasite *Ancylostoma ceylanicum PNAS* 13 (41); 15154-15159
136. Ibrahim M. A.et al. (2010) *Bacillus thuringiensis* A genomics and proteomics perspective *Bioengineered Bugs* 1(1); 31-50
137. Sussex University website: http://www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/
138. Schnepf E. et al. (1998) Bacillus thuringiensis and its pesticidal crystal proteins *Microbiology and Molecular Biology Reviews* 62 (3); 775-806
139. Edwards et al. (1990) United States Patent 4,948,734
140. Borgonie G. et al. (1996) Effect of nematicidal *Bacillus thuringiensis* strains on free-living nematodes. 1. Light microscopic observations, species and biological stage specificity and identification of resistant mutant of *Caenorhabditis elegans Fundam. Appl. Nematol.* 19 (4); 391-398

141. Borgonie G. et al. (1996) Effect of nematicidal *Bacillus thuringiensis* strains on free-living nematodes 2. Ultrastructural analysis of the intoxication process in *Caenorhabditis elegans Fundam. Appl. Nematol*. 19 (5); 407-414
142. Borgonie G. et al. (1996) Effect of a nematicidal *Bacillus thuringiensis* strain on free-living nematodes. 3. Characterization of the intoxication process *Fundam. Appl. Nematol.* 19 (6); 523-528
143. Marroquin L. D. et al. (2000) B.t. Toxin susceptibility and isolation of resistance mutants in the nematode *C. elegans Genetics Society of America* 155; 1693-1699
144. Hu Y. et al. (2010) *Bacillus thuringiensis* Cry5B protein is highly efficacious as a single-dose therapy against an intestinal roundworm infection in mice *PLoS neglected tropical diseases* 4 (3); 1-7
145. Guo et al. (2008) New strategy for isolating novel nematicidal crystal protein genes from *B. thuringiensis* strain YBT-1518 *Applied and environmental microbiology* 74 (22); 6997-7001
146. Wei J-Z. et al (2003) B.t. crystal proteins that target nematodes *PNAS* 100 (5); 2760-2765
147. Aroian R. et al. (2011) Bt crystal pore-forming proteins as anthelmintics Presentation at the Symposium “Membrane ion-channels in helminth parasites: resistance and sites of action for anthelmintics”, Philadelphia
148. Trejo G. et al. (1970) Uridine Diphosphate-α-D-Galactofuranose, an intermediate in the biosynthesis of galactofuranosyl residues *Biochem. J.* 117; 637-639
149. Borrelli S. et al. (2010) Short communication: Antimycobacterial activity of UDP-galactopyranose mutase inhibitors *International Journal of Antimicrobial Agents* 36; 364-368
150. Sanders D. A. R. et al. (2001) UDP-galactopyranose mutase has a novel structure and mechanism *Nature Structural Biology* 8 (10); 858-863
151. Novelli J. et al. (2009) Characterization of the *Caenorhabditis elegans* UDP-galactopyranose mutasae homolog glf-1 reveals an essential role for galactofuranose metabolism in nematode surface coat synthesis *Developmental Biology* 335; 340-355
152. Behm C. et al. (2010) UDP-galactopyranose mutase form *Haemonchus contortus,* and RNAi agents that inhibit it *PCT/AU2010/001317*
153. Partha S. K. et al. (2011) Identification of novel inhibitors of UDP-galactopyranose mutase by structure-based virtual screening *Molecular Informatics* 30; 873-883
154. Cunico W. et al. (2008) Chemistry and biological activities of 1,3-thiazolidin-4-ones *Mini Reviews in Organic Chemistry* 5; 336-344
155. Abhinit M. et al. (2009) Exploring potential of 4-thiazolidinone: a brief review *Intern. Journal of Pharmacy and Pharmaceutical Sciences* 1 (1); 47-64

156. Ravichandran V. et al. (2011) Design, synthesis, and evaluation of thiazolidinone derivatives as antimicrobial and anti-viral agents *Chem Biol Drug Des* 78; 464-470
157. Dykhuizen E. C. et al. (2008) Inhibitors of UDP-galactopyranose mutase thwart mycobacterial growth *J. Am. Chem. Soc. Communications* 130; 6706-6707
158. Christodoulou M. S. et al. (2010) Novel pyrazole derivatives: Synthesis and evaluation of anti-angiogenic activity *Bioorganic and Medicinal Chemistry* 18; 4338-4350
159. Foster J. M. et al. (2005) Parasitic nematodes have two distinct chitin synthases *Molecular & Biochemical Parasitology* 142; 126-132
160. Spindler K.-D. et al. (1990) Chitin metabolism: a target for drugs against parasites *Parasitology Res* 76; 283-288
161. Harris M. T. & Fuhrman J. A. (2002) Structure and expression of chitin synthase in the parasitic nematode *Dirofilaria immitis Molecular & Biochemical Parasitology* 122; 231-234
162. Wilson T. G. & Cryan J. R. (1997) Lufenuron, a chitin-synthesis inhibitor, interrupts development of Drosophila melanogaster *The Journal of Experimental Zoology* 278; 37-44
163. Summary of the product characteristics, Program Plus, Novartis, PDF: http://www.ms.public.lu/fr/activites/pharmacie-medicament/mise-marche-medic-veterin/rcp/p/progplus5komma75mg115mgcpr.pdf
164. Fenn K. & Blaxter M. (2004) Are filarial nematode Wolbachia obligate mutualist symbionts? *TRENDS in Ecology and Evolution* 19 (4); 163-166
165. Slatko B. E. et al. (2010) The *Wolbachia* endosymbiont as an anti-filarial nematode target *Symbiosis* 51; 55-65
166. Webpage:http://www.wormbase.org/species/c_elegans/gene/WBGene00004510#04-9ea-10

Appendices

List of abbreviations

Abbreviation	Translation
3'-OH	Link at 3' end with oxygen-hydrogen
5-HT	Serotonin
18S ribosomal RNA	S : Svedberg units
A	Adenine
AAD / AADs	Amino Acetonitrile Derivatives
ABC-transporters	ATP-binding cassettes transporters
ABySS	Assembly by short sequences
ACh	Acetylcholine
A. ceylanicum	*Ancylostoma ceylanicum*
AcpS	Acyl carrier protein synthase
***acr-23* / ACR-23**	Acetylcholin receptors
AH	Animal Health
A. suum	*Ascaris suum*
ATG	Signals start of translation
ATP	Adenosine triphosphate
A. viteae	*Acanthocheilonema viteae*
blastp	Basic local alignment search tool protein-protein
blastn	Basic local alignment search tool nucleotide-nucleotide
blastx	Basic lical alignment search tool using a translated nucleotide query
B. malayi	*Brugia malayi*
bp	base pair
***bre-3* (*-1, -2, -4, -5*)**	Bt toxins resistant
Bt	*Bacillus thuringiensis*
B. thuringiensis	*Bacillus thuringiensis*
BZT	Benzimithiazoles
C	Cytosine
Ca^{2+}	Calcium
CAP	Companion Animal Parasiticides
ccmB	Heme exporter protein B
CD4	Cluster of differentiation 4, glycoprotein
CD-Hit	Cluster database at high identity with tolerance
CEGMA	Core eukaryotic gene mapping approach
C. elegans	*Caenorhabditis elegans*
chs-1	Chitin synthase gene 1
Cl^-	Chloride
C. lupus familiaris	*Canis lupus familiaris*
CoA	Coenzyme A
Cry proteins (Cry-5 (-6, -12, -13, -14, -21)	Crystal proteins

Cry5B	Crystal proteins 5B
Ctrl+ / Ctrl-	Control + / Control -
Cyto	Cytosolic
Cyt proteins	Cytolitic proteins
DB	Database
dbEST	EST database
ddATP, ddCTP, ddGTP, ddTTP	Dideoxyribonucleotide (A, C, G, T)
***deg-3* / DEG-3**	Degeneration of certain neurons
***des-2* / DES-2**	Degeneration suppressors
DHFR	Dihydrofolate reductase
D. immitis	*Dirofilaria immitis*
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
DnaB	Mobile replication promoter signal for primase
dNTP	Mix of the four dideoxyribonucleosides
DOXP	1-dioxy-D-xylulose 5-phosphate
dsRNA	Double stranded RNA
dTDP	Thymidine diphosphate
EC 90	Efficacy concentration at 90%
EC 50	Efficacy concentration at 50%
E.C.	Enzyme commission
E. coli	*Escherichia coli*
ED 50	Effective dose that cures 50%
EMBL-EBI	European Molecular Biology Laboratory – European Bioinformatics Institute
EMBOSS	European Molecular Biology Open Software Suite
ENT-1	Equilibrated nucleoside transporter
ePCR	Emulsion PCR
erg-11	Ergosterol biosynthesis
ESTs	Expressed sequence tags
E-value	Expected value
FabZ	Encodes for beta-hydroxyacyl-acyl carrier protein dehydratases in *E. coli*
FAD	Flacin adenine dinucleotide
FDA	US Food and Drug Administration
FtsZ	Filamenting temperature-sensitive mutant Z
G	Guanine
g	Gram
GA	Genome analyzer
GA, Athens	Georgia
$GABA_A$ / $GABA_B$	Gamma aminobutyric acid A or B

Gb	Giga base
G. duodenalis	*Giardia duodenalis*
GI	Gastro-intestinal
glf-1	Galactofuranose synthesis (UDp-galactopyranose mutase)
GluR	Glutamate receptors
GMP	Guanosine monophosphate
GPCR	G protein-coupled receptors
GPI	Gylcophosphatidylinositol
GTP	Guanosine triphosphate
GTR+	Generalised time-reversible
HARD	Heartworm associated respiratory disease
H. bakeri	*Heligmosomoides bakeri*
HBSS	Hank's Balanced Salt Solution
H. contortus	*Haemonchus contortus*
hemG	Hemin
hemJ	Transformation in heme synthesis
hemY	Protoheme IX synthesis protein
HMM	Hidden markov model (statistic)
HMMer	Sequence analyzer
H. sapiens	*Homo sapiens*
IC 50	Inhibitory concentration of 50%
IL-4, -16	Interleukine 4, -16
IM	Immune modulator
K^+	Potassium
K / k-mers	Quantity or number of amino acid sequences
Kb	Kilo base
kDa	Kilo Dalton
L_3	Third stage larvae
L_3_arrest	L_3 developmental stage arrest
LAP	Liver-enriched transcriptional activator protein
LBD	Ligand binding domain
Ld	Ladder
LD 50	Lethal dose that kills 50%
LDA	Larval development assay
LGIC	Ligand-gated ion channels
L. Loa	*Loa loa*
Log	Logarithm
L. sigmodontis	*Litomosoides sigmodontis*
LTR	Long terminal repeat
MAKER	Genome annotation pipeline
Mb	Mega base
mBm	Mitochondria *B. malayi*

mDi	Mitochondria *D.immitis*
mdr-1	Multi drug resistant gene 1
Memb	Membrane domain
mf	Microfilariae
MIT	Massachusetts Institute Technology
ml	Millilitres
ML / MLs	Macrocyclic lactones
Molt_defective	Defective molting stage
***mptl-1* / MPTL-1**	Monepantel
mg/kg	Miligrams per kilograms
MIF	Macrophage migration inhibiton factor
mRNA	Mitochondrial RNA
MUMMER	Program to align genome
µg	Micrograms
Ns	Number of N in contigs (gaps)
N50	Statistical measure of average length of a set of sequences
Na^+	Sodium
NADH	Nicotinamide adenine dinucleotide
nAChR	Nicotinic acetylcholine receptors
NHGRI	National Human Genome Research Institute
nm	Nanometers
NUCMER	Nucleotide Mumer
***NTR* gene**	Nitrogenase and nitrate reductase gene
O. ochengi	*Onchocerca ochengi*
O. volvulus	*Onchocerca volvulus*
P2	Aminopurine permease
P2X	ATP-gated purine receptors
PAO /BEL	Subclass of LTR retrotransposon
PCR	Polymerase chain reaction
PDZ-domain	Acronym combining the first letters of three proteins — post synaptic density protein (PSD95), Drosophila disc large tumor suppressor (Dlg1), and zonula occludens-1 protein (zo-1)
Per os	Oral administration
PFAM	Large collection of protein families, represented by multiple sequence alignments and hidden Markov models (HMMs)
P-glycoproteins	Permeability glycoproteins
PhD	Doctor of Phylosophy
PHRAP	DNA sequence assembler
P. falciparum	*Plasmodium falciparum*
ppm	Particles per millions

PRPP	5-phosphoribosyl pyrophosphate
RAxML	Randomized Axelerated Maximum Likelihood
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-Seq	Whole transcriptome shotgun sequencing
R. norvegicus	*Rattus norvegicus*
RT-PCR	Reverse-transcriptase PCR
Sec	Secretory pathway
Sec	Secreted
SEM	Scanning electron microscopy
seq	Sequencing
SH2 domain	Src Homology 2
SLO-1	Encodes a voltage and calcium activated potassium channel
SMRT	Single molecule real time
SNAP	Gene finder based on HMM
SNPs	Single nucleotide polymorphism
SOAP de novo	Short oligonucleotide analysis package
SOCS	Suppressors of cytokine signalling
SOLID	Sequencing by oligo ligation / detection
Spp.	Species
SSRNA	Simple strand RNA chain (virus)
Src	Family of proto-oncogenic tyrosine kinases
T	Thymine
TbAT1	Gene encoding the aminopurine permease
tblastn	Basic local alignment search tool protein-nucleotide translation
T. b. rhodensiense	*Trypanosoma brucei rhodensiense*
T. colubriformis	*Trichostrongylus colubriformis*
T. cruzi	*Trypanosoma cruzi*
TH1	T helper cell 1
TMI-TMIV / M1-M4	Transmembrane domains 1-4
tRNA	Transfer RNA
T. vaginalis	*Trichomonas vaginalis*
UDP	Uridine diphosphate
US	United States
VC	Vaccine candidates
***w*Bi**	*Wolbachia* of *B. malayi*
***w*Di**	*Wolbachia* of *D. immitis*
***w*Mel**	*Wolbachia* of *Drosophila melanogaster*
W. bancrofti	*Wuchereria bancrofti Wolbachia* of *Drosophila melanogaster*
WO bacteriophage	Mobile element in the genome of Wolbachia

List of figures

Figure 1. Third stage larvae of *D. immitis*. SEM picture showing the worm cuticle details.

Figure 2. Phylogeny of nematodes based on 18S ribosomal RNA (1).

Figure 3. US prevalence of antigen positive tests to *D. immitis* in dogs per county. Counties in grey did not reported results. In white, no dog reported as positive (0%). Remaining counties were coded as follows: 0.1–2.0% (taupe), 2.1–4.0% (salmon), 4.1–6.0% (red), 6.1–20.0% (brick red

Figure 4. Repartition of *D. immitis* in Europe.

Figure 5. *D. immitis* life cycle.

Figure 6. Susceptible and non-suscepticle ages of heartworm to Macrocyclic lactones and melarsomine based upon the age of the worm in days.

Figure 7. Illustration of the Sanger dideoxy method for DNA sequencing. Addition of primers, DNA polymerase and labeled dideoxyribonucleotides, lacking the 3' hydroxyl group, to a mix of DNA simple strand. Without the 3' hydroxyl group, reactions are stopped producing a group of DNA strands of various sizes. These products are separated by size on electrophoresis gels. The labeled nucleotides allow the reading of the sequences over the four columns of the gel. See text for explanation.

Figure 8. Decreasing costs to sequence a human genome over 10 years.

Figure 9. Illumina Hiseq. Sequencing workflow of the Genome Analyzer. A. Generation of sequencing libraries by random fragmentation of DNA strand and ligation of adapters at both ends of the fragment. B. Addition of fragment to the flow cell coated with complementary oligonucleotides. It is followed by the formation of bridges by hybridization and the amplification of the 3' to 5' in order to form clusters. C. A denaturation step occurs followed by the addition of primers, polymerase and labeled nucleotides. The surface is imaged and the process can be repeated.

Figure 10. Illustration representing the mode of action and localization of numerous different anthelmintics at the neuromuscular junction. The different class of drug represented have specific mode of action, targeting for example tubulins for the benzimidazoles, calcium channels for the cyclodepsipeptides and potassium-natrium channels for avermectins. Related to each class, some drug names are given as example.

Figure 11. Schematic representation of nicotinic acetylcholine receptors protein with the four transmembrane domains (M1-M4), the extracellular chain with the N-terminal and glycolysation sites and the intracellular domain with the cys-loop.

Figure 12. Pentameric nicotinic acetylcholine receptor. The neurotransmitter binds to the receptor and allows the cation flow through the channel causing muscular contractions

Figure 13. Monepantel (AAD 2225) activity in four in vitro tests and one in vivo test. Concerning the in vitro tests, the efficacy against microfilariae or L_3 is evaluated after 48 hours of incubation with the compounds. In the LDA, egg to L_3, the incubation time is longer with duration of seven days. The results are represented in ppm (particle per million). In vivo,

compounds are applied per os at 40 mg/kg. The efficacy is evaluated at necropsy, by count of the adult parasites. Efficacy is defined by comparison to positive control, milbemycin oxime, and a low concentration in ppm.

Figure 14. In vitro efficacies of AADs against *D. immitis* and *A.viteae* microfilariae; the tests were performed in triplicates. Results are represented in ppm (particle per million). Efficacy is defined by comparison to positive control, milbemycin oxime, and a low concentration in ppm.

Figure 15. In vitro efficacies of AADs against gastro-intestinal nematodes in the LDA, from egg to L3 were evaluated after seven days of incubation. In vitro efficacies of AADs were evaluated against *D. immitis* and *A.viteae* microfilariae after 72 hours of incubation all tests were performed in triplicates. Results are represented in ppm (particle per million).

Figure 16. Agarose gel of 1.2% representing the RT-PCR (reverse transcriptase) picture performed on the *des-2* pseudogene sequence of *D. immitis* on complementary DNA and genomic DNA. Positive controls were performed with beta-tubulin specific to *D. immitis*. Negative controls are without primers or without DNA samples.

Figure 17. Percentage of efficacy, in logarithm, of 21 guanosine analogs against *D. immitis* microfilariae at 32 ppm after 48 hours of incubation. Test performed in triplicates.

Figure 18. Percentage of efficacy at 32 ppm (particle per million) of in vitro tests with Lufenuron and six derivatives against *D. immitis* microfilariae after 48 hours incubation. Test performed in triplicates.

Figure 19. Percentage of efficacy, logarythm scale, of nine antibiotics potentially targeting *Wolbachia* of *D. immitis.* Test performed against *D. immitis* microfilariae at 32 ppm after 48 hours of incubation and in triplicates.

List of Tables

Table 1. Summary of different methods of high-throughput sequencing and their specific characteristics.

Table 2. Comparison between various animal species of the genome size, number of ligand-gated ion channels and the sequencing method for the genome studies.

Table 3. Description of the 24 LGIC receptors of *D. immitis*, with homology to *C. elegans.* LBD means ligand-binding domains and “memb” is related to the membrane domains of the LGIC. E (E (expect)-value): parameter describing the number of hits “expected” to see by chance when searching a database with a particular size), Score (number used to assess the biological relevance of a finding; it describes the overall quality of an alignment).

Table 4. Table regrouping the potential drug targets of *D. immitis* obtained after an exclusion criteria study on the whole proteome ((i) presence of an ortholog in *C. elegans* which has as an RNAi phenotype *lethal*, *L_3_arrest*, or *molt_defective*; (ii) absence of a blastp hit in the predicted proteomes of *H. sapiens* and *C. lupus familiaris* with an E-value below 10^{-5}; (iii) predicted function as an enzyme or receptor based on the annotation of the *C. elegans* ortholog and blastp searches against SwissProt).

Table 5. Potential drug targets of the heartworm endosymbiont, *Wolbachia pipientis*. As identified by the following exclusion criteria: (i) presence of orthologs in *E. coli* which are essential; (ii) absence of a blastp hit in the predicted proteomes of *H. sapiens* and *C. lupus familiaris* with an E-value below 10^{-5}; (iii) predicted function as an enzyme or receptor based on the annotation of the *E. coli* ortholog and blastp searches against SwissProt.

Curriculum vitae

Godel Christelle
Pré-Mermoud 2
1580 Avenches
Phone +41 79 405 00 77
Email csgodel@hotmail.com

29 ans
Married, 1 child
Swiss

KEY COMPETENCES

Ambitious, methodical, meticulous, implicated and consciencious in my job.

FORMATION

PhD ès Science in Parasite Molecular Biology 2009-2012
PhD carried out at the Swiss Tropical and Parasitical Health, Basel, with a fellowship and project from Novartis Animal Health, St-Aubin FR.
"Identification and characterization of potential AAD receptors in *Dirofilaria immitis*"
Side-project developed at the Ecole Polytechnique Fédérale de Lausanne (EPFL), ausanne, Development of an *in vitro* screening test in liquid culture involving the worm *C. elegans*.

Master ès Science in Parasites Biology 2006-2008
University Studies in Parasites Biology carried out at the University of Neuchâtel
Master Thesis performed at Novartis Animal Health, St-Aubin FR :
"Development of an *in vitro* test with model parasite of gerbils, *Acanthocheilonema viteae*, and target parasite, *Dirofilaria immitis*, affecting dogs".

- Optional courses in "Pharmaceutical Law and foods" and "Law and Medical Progress"

Bachelor ès Science in Biology 2003-2006
University Studies in Biology carried out at the University of Neuchâtel

Veterinary Studies 2002-2003
Universität Zurich, first year

PROFESSIONAL EXPERIENCES

Laboratory technician 2009
Temporary post in the "Rodent Model" team , Dr. Sandra Schorderet-Weber, Novartis Animal Health SA, St-Aubin FR

- *In vivo* tests following on Gerbils (*Meriones unguiculatus*), Rabbits and Mice; dissections, blood takings
- Data treatment and results analysis

Pratical Training in Molecular Biology 2008
Laboratory of Dr. Alessandro Puoti, University of Fribourg

- Use of molecular methods applicated to the nematode *Caenorhabditis elegans*, DNA extraction, RNA extraction, PCR, RT-PCR, Electrophoresis,…
- Splicing assays, Backcrosses,…

Module I of Animal Experimentation 2008
"LTK Modul I: Einführung in die Labortierkunde", Institut für Labortierkunde, Universität Zürich

Master Thesis in Parasite Biology 2007-2008
Screening laboratory, Dr. Jacques Bouvier, Novartis Animal Health SA, St-Aubin FR

- Development of *in vitro* tests with filarial nematodes, *Acanthocheilonema viteae* (microfilariae and L_3) and *Dirofilaria immitis* (microfilariae).
- Use of molecular methods applicated to filariae, *A. viteae* and *D. immitis,* RNA extraction, DNA extraction, PCR, RT-PCR, Electrophoresis,…

Practical Training in Parasitology 2007
Screening laboratory, Dr. Jacques Bouvier, Novartis Animal Health SA, St-Aubin FR

- Development of *in vitro* tests with the nematode *Trichostrongylus colubriformis* (L_3 and L5)

Student Assistant in Parasitology 2007
Laboratory of Molecular Parasitology, Prof. Bruno Betschart, University of Neuchâtel

Student Assistant in Eco-Ethology 2006-2007
Laboratory of Eco-Ethology, Prof. Redouan Bshary, University of Neuchâtel

Practical Training in Eco-Ethology 2006
Laboratory of Eco-Ethology, Dr. Peter Neuhaus, University of Neuchâtel

- Ethologic study of American Bisons, Elk Island National Park, Alberta, Canada

SYMPOSIUMS, CONFERENCES

SSTMP
"Annual meeting for PhD student in Parasitology", Münchenwiller, Prof. H.-P. Beck 2008

Swiss Tissue Culture Society
"Nanoparticles-Biomedical Applications and Challenges", Basel, Prof. P. Mäser 2009

Roche
"High-Definition Genomics-From Genome to Gene Function", Zurich, Dr. L. Roggo 2009

SSTMP
"Annual meeting for PhD student in Parasitology", Basel, Prof. H.-P. Beck 2009

SSTMP
"Diagnostics: From rapid tests to molecular diagnostics", Basel, Prof. C. Lengeler 2009

American Heartworm Society
American Heartworm Society's 13th Triennial Symposium, Memphis (USA), Lisa S. Scott 2010

3e Cycle Symposium
"Phylogenomics: New Perpectives for Phylogenetic Studies", Geneva, Dr. J. Pawlowski 2010

Interuniversity Doctoral Program in Organismal Biology (Ecology and Evolution)
"Avoiding Tragedies of the Commons: An evolutionary Approach to Human Cooperation", Neuchâtel, Dr. Christiane Bobillier 2010

De Novo Genome Assembly Workshop
University of Lausanne, Dr. Laurent Falquet 2011

WAAVP
23rd International Conferenceof the World Association for the Advancement of Veterinary Parasitology
Dr. G. Mauricio Bulman, Buenos Aires, Argentina 2011

Symposium on Membrane Ion-channels in Helminth Parasite, Resistance and Sites of Action for Anthelmintics
Presentation on the genome of the heartworm *Dirofilaria immitis*.
Dr. R. Martin and Dr. A. Wolhstenholme, Philadelphia, USA 2011

COURSES

Course of Anesthesiology, Isabelle Iff, Novartis Animal Health Research Center, St-Aubin FR 2012

Data Management and Data Integrity, Compliance and Ethics training (v12)
Novartis Animal Health Research Center, St-Aubin FR 2011

Introduction to GLP (OCDE), Laboratory and Preclinical QA 2011

Raw Data, Metadata and Processed Data in GLP Environment
Laboratory and Preclinical QA, Advance Quality Program 2011

LANGUAGES

French	Mother tongue
English	First Certificate of English, University of Cambridge ESOL examinations 3 months of intensive language course, Calgary, Alberta, Canada
German	Intermediate, B1 (european standard of languages) 2 months of intensive language course, Regensburg, Germany
Spanish	Elementary, A2 (european standard of languages)

PARALLELE ACTIVITIES

Member of the SSTMP 2009
(Swiss Society of Tropical and Medical Parasitology)

Institut Council 2006-2008
Representant of the University Master of Science, option Parasites Biology at the Biology Institut Council of the University of Neuchâtel

INTERESTS

Annexe Formations
SSI and CMAS diving diploma, Youth diploma of rescue, J&S “Sport de camp/Trekking” 1 and 2, Chef of scout group, Athletism instructor, Youth formation and animation, Course of first aid technician, Equestrian diploma

Sports

Running, Horse riding, Cycling, Skiing, Snowboarding, Tennis playing

Passions
Animals, Nature, Cultral and sportive travels, Languages, Photography, Reading and Informatic

JOBS IN PARALLELE TO STUDIES

English Teaching Replacement, Secondary school of Avenches, Vaud	2011-2012
Scholar support, at Secondary, High School and University grades	1999-2007
Work as Equestrian, Rafour stable, Montmollin	2006-2007
Housework, for Private people	2007
Maintenance and Cleaning work, Public Maintenance of the town of Neuchâtel (summer)	2003-2007
Housework in an elder house, Le Chalet, Bevaix	2002-2003
Secretary, Zurich Insurance, Neuchâtel (summer)	2002
Scholar support, in science at ESRN, Peseux	1995-1998