Heartworm Genomics: Unprecedented Opportunities for Fundamental Molecular Insights and New Intervention Strategies

Robin B. Gasser and Cinzia Cantacessi

Vector-borne diseases, including canine heartworm disease (CHWD), are of major socioeconomic and canine health importance worldwide. Although many studies have provided insights into CHWD, to date there has been limited study of fundamental molecular aspects of *Dirofilaria immitis* itself, its relationship with the canine host, its vectors, as well as the potential of drug resistance to emerge, using advanced -omic technologies. This article takes a prospective view of the benefits that advanced -omics technologies will have toward understanding *D. immitis* and CHWD. Tackling key biological questions using these technologies will provide a "systems biology" context and could lead to radically new intervention and management strategies against heartworm.

Keywords: canine heartworm disease (CHWD), *Dirofilaria immitis*, heartworm, drug resistance, genomics, genetics, molecular biology, transcriptomics

Many vector-borne diseases have a major adverse impact on animal and human health worldwide.[1-3] A particularly important example of veterinary importance is canine heartworm disease (CHWD), caused by the filarial nematode *Dirofilaria immitis*.[4,5] Despite current knowledge and developments in advanced technologies, there are still major gaps in our knowledge of many areas, including the fundamental molecular biology, molecular epidemiology, ecology and population genetics of the causative agent and its vectors, the disease itself, and also drug resistance development in *D. immitis*.[6,7] Moreover, there are some limitations in diagnosis and intervention, which represent critical obstacles to the effective control of CHWD.[4,8] Although sustained research and funding have contributed significantly to an improved understanding of vector-borne diseases of humans,[9-12] this is not the case for those of veterinary importance, which are presently neglected in terms of research and development.

The revolution in molecular and computer technologies provides substantial prospects for investigating important pathogens and their vectors, providing insights into their epidemiology, ecology, evolution, and cellular processes. However, the relatively high cost and laborious nature of molecular biological research has sometimes been an impediment to progress, particularly in the veterinary field. Revolutionary developments in a range of genomic technologies[13] provide unprecedented opportunities to explore CHWD and other diseases, at a rate and on a scale unimaginable just a couple of years ago, providing enormous prospects to tackle critically important areas of research for the first time. Future research should harness such technologies to address major knowledge gaps in CHWD. Elucidating *D. immitis*, its relationship with its vectors and its definitive hosts, the disease itself, and the epidemiology and ecology of the parasite will have substantial implications for improving the diagnosis, treatment, prevention, and control of CHWD in years to come. This article takes a prospective view of the impact that -omic technologies could have on our knowledge and understanding of *D. immitis* and CHWD on the molecular level. We briefly explain the principles of transcriptomic and genomic sequencing as well as bioinformatic technologies, and describe the exciting implications of these technologies in both fundamental and applied areas.

Background on Genomic and Bioinformatic Technologies

Genomics is the study of the entire complement of genetic material (DNA or genome) in an individual organism (e.g., parasite). Transcriptomics is the molecular science of examining the transcription of all genes at the level of the cell,

From the Faculty of Veterinary Science, The University of Melbourne, Parkville, Victoria 3010, Australia.
Current research in the Gasser Lab is funded mainly through the Australian Research Council (ARC), the National Health & Medical Research Council (NHMRC), and Melbourne Water Corporation.
Address reprint requests to: Robin B. Gasser, DVM, PhD, DVSc, Faculty of Veterinary Science, The University of Melbourne, Parkville, Victoria 3010, Australia. E-mail: robinbg@unimelb.edu.au.

1527-3369/06/0604-0171\.00/0
doi:10.1053/j.tcam.2011.09.003

tissue, and/or whole organism, allowing inferences regarding cellular functions and mechanisms. The ability to explore and measure the transcription of thousands of genes simultaneously has led to major advances in all biomedical fields, from understanding basic functions in model organisms, such as *Caenorhabditis elegans* (a free-living nematode) and *Drosophila melanogaster* (vinegar fly),[14-18] to investigating molecular processes linked to growth, development, and reproduction, to the study of the mechanisms of survival and drug resistance. Until recently, transcriptomes have been characterized qualitatively by sequencing expressed sequence tags using conventional technologies,[19,20] whereas levels of transcription have been assessed using complementary DNA (cDNA) microarrays[21] or real-time polymerase chain reaction (PCR).[22] In the last few years, there has been an increased demand for practical computer tools for the efficient annotation of nucleotide sequence datasets, particularly within the framework of large-scale expressed sequence tag projects.[23]

Next-generation Sequencing of Transcriptomes and Genomes

There has been a massive expansion in next-generation sequencing (NGS) technologies,[24-27] which now provide unique opportunities to explore de novo the transcriptomes and nuclear genomes of different species, strains and developmental stages of *D. immitis* as well as its vectors and its canine or other definitive hosts. The capacity of NGS techniques to generate millions to hundreds of millions of sequences, in parallel, has put them at the forefront of the scientific research.[28–30]

NGS technologies include 454 (Roche),[24] SOLiD,[27] and Illumina/HiSeq.[25] For example, the 454 technology platform[24] uses a sequencing-by-synthesis approach, by which cDNA is randomly fragmented into 500 to 1000 base pair fragments. During the process of cDNA library construction, an adaptor is ligated to each end of these fragments and then mixed into a population of agarose beads whose surfaces anchor oligonucleotides complementary to the 454-specific adapter sequence, such that each bead is linked to a single fragment. Each of these complexes is transferred into individual oil-water micelles containing amplification reagents and then subjected to an emulsion PCR step, during which ~10 million copies are produced and bound to individual beads. In the sequencing step, the beads anchoring the cDNAs are deposited on a picotitre plate, together with other enzymes required for the pyrophosphate sequencing reaction on a solid support system.[31] The 454 technology (a "long-read" platform) has been used frequently for de novo genomic or transcriptomic studies. However, increasingly, Illumina/HiSeq is being used for this purpose. This latter technology differs significantly from the 454 approach,[25] in that, after fragmentation of cDNA, Illumina-specific adaptors are ligated to each cDNA template and then covalently attached to the surface of a glass slide, allowing individual templates (cDNAs) to form bridge-like structures. During the amplification step (bridge PCR), clonal clusters of ~1000 amplicons are generated and immobilized to a single physical location on the slide. Then the cDNAs are linearized, and the sequencing reagents (including 4 fluorescently labeled nucleotides) are added to the flow cell. After the individual fluorescent bases have been incorporated, the flow cell is interrogated with a laser beam in several locations, allowing several image acquisitions at the end of a single synthesis cycle, and the sequences to be read.[31] This technology is considered ideal for resequencing projects, targeted sequencing, single nucleotide polymorphism analyses, and gene transcription studies.

In the last years, numerous studies have shown the utility of NGS technologies for investigating, for example, aspects of the molecular biology, systematics, and population genetics of parasites.[32-35] In particular, 454 technology has been used recently for the rapid de novo sequencing of the transcriptomes of numerous pathogens of veterinary and human health importance,[33-41] yielding substantial datasets and providing a major step forward in the understanding of the basic molecular biology of these organisms.

Bioinformatic Tools (= Computer Programs) for the Analysis of NGS Datasets

The development of practical and efficient bioinformatic tools has become critical for comprehensive analyses of DNA, RNA and protein sequence data as well as for making biological sense of such datasets. Therefore, there has been relatively rapid progress in the construction of new programs and/or integrated pipelines, some of which are accessible via the Worldwide Web.[42-46] The principles, methods and protocols for the analysis of sequence data, together with currently available bioinformatic tools and pipelines, have been reviewed recently.[47]

In brief, after the acquisition of data, nucleotide sequences are firstly screened for repeats, contaminants and/or adaptor sequences,[47,48] and "clustered" (assembled) into contiguous sequences (of maximum length; called contigs) based on sequence similarity.[47] Long reads (produced by 454 or conventional sequencing technologies) and short reads (e.g., Illumina) are assembled using the algorithms "overlap-layout-consensus"[49] and "de Bruijn graph,"[50,51] respectively. For the former algorithm,[49] pairwise overlaps among reads are computed and stored in a graph; the graphs are used to compute a layout of reads and then a consensus sequence of contigs.[42,52-57] For the de Bruijn graph,[50,51] reads are fragmented into short segments, called "*k*-mers" (number of nucleotides in each segment; overlaps among *k*-mers are captured and stored in graphs, which are then used to produce consensus sequences.[51,56-60]

After assembly, the contigs and single reads (singletons) are compared, using different types of the Basic Local Alignment Software Tool (BLAST),[61] with known sequence data available in public databases, to infer an identity for each query sequence, if significant matches are found.[47] In addition, assembled nucleotide sequences are usually conceptually translated into pre-

dicted proteins with algorithms that identify protein-coding regions from individual contigs.[43,62,63] Protein analyses, including amino acid sequence comparisons with data available in public databases, are then conducted.[44,46,47,64-67] Public databases online represent comprehensive collections of amino acid and nucleotide sequences[68-70]; sequences are stored together with relevant information, including primary references, and predicted and/or experimental data. For transcriptomic datasets, examples of databases include the Sequence Read Archive[71] and UniGene.[72] In addition, there are various specialized collections of gene and protein data on organisms about which much is known. Examples include the databases for *C. elegans* (WormBase at http://www.wormbase.org)[73,74] and *D. melanogaster* (http://flybase.org).[75]

The data generated by NGS technologies are often gigabytes to terabytes in size, substantially increasing the demands placed on data transfer, analysis, and storage. For instance, to circumvent this limitation somewhat, we recently constructed an integrated bioinformatic workflow system for the efficient analysis and annotation of large NGS datasets. This platform, which is presently accessible via http://research.vet.unimelb.edu.au/gasserlab/index.html, has proved very effective and time efficient for distilling biologically relevant molecular information from large NGS datasets for pathogens of veterinary importance.

Opportunities and Challenges

Knowledge of the transcriptomes, genomes, and proteomes of *D. immitis*, its vectors, and its definitive host is central to gaining an enhanced understanding of the molecular mechanisms that govern essential biological as well as infection and disease processes and, ultimately, could assist in identifying possible mechanisms of drug resistance and avenues for the development of radically new intervention strategies. Detailed and accurate analyses of nucleic acid and protein sequence data, usually by comparison with reference organisms, are crucial in providing biologically meaningful information on the organisms under study. Until recently, detailed bioinformatic analyses of such data have been restricted largely to specialized laboratories with substantial computer and software capacities. However, increasingly flexible and practical bioinformatic workflow systems are becoming available to assist scientists in their analyses of massive NGS datasets.

Although the draft genome *Brugia malayi* (a filarioid related to *D. immitis*) is available[76] and should provide a useful reference for comparative analysis, transcriptomic and genomic sequence data for *D. immitis* will need to be assembled de novo. This would mean that pooled sequence reads must be assembled without a bias toward known sequences.[51] Because of the amount of RNA required for NGS (~5-10 μg),[77] transcriptomes are usually derived from multiple organisms (in this case, worms), which have the potential to lead to possible polymorphism in the sequence data acquired (linked, for instance, to a biased nucleotide content, single nucleotide polymorphisms[6] and other types of sequence variation within or among worms), and pose challenges for data assembly. In addition, reliable de novo assemblies are dependent on the availability of long reads (> 100 bases) and/or of high-coverage, paired-end sequence data.[78] In previous studies, the combined use of 454 and Illumina sequencing platforms has enabled the assembly of raw reads into large scaffolds, without a need for a reference sequence.[79-81]

The accurate assembly of sequence data is a crucial step in examining coding genes and, subsequently, in addressing biological questions regarding gene and protein functions. Functions are initially predicted by "sequence annotation" (= the process of gathering all available information and relating it to the sequence assembly both by experimental and computational means).[82] Reliable annotation is dependent on the efficiency of the updates and curation of data. Presently, public databases for bioinformatic analyses and annotation of sequence data are accessible via multiple portals, and there is considerable variation in the rate at which public databases are updated and corrected. In addition, some information-management systems incorporate data from large-scale projects, but often the annotation of single records from the literature is slow.[83] Given that, presently, the annotation of sequence data for parasites and vectors relies heavily on the use of bioinformatic approaches and already annotated/curated sequence data for a wide range of organisms, these aspects need to be considered.

The annotation of peptides inferred from a dataset is conducted by assigning predicted biological function(s) by comparisons with information in public databases, including InterPro (www.ebi.ac.uk/interpro/), Gene Ontology (www.geneontology.org/), OrthoMCL (www.orthomcl.org/), and BRENDA (www.brenda-enzymes.org/). Using this approach, together with knowledge of the *B. malayi* genome,[76] inferences can be made regarding the function and essentiality of key groups of molecules involved in biological processes in *D. immitis*. Such groups include molecules linked to reproduction, development, signal transduction, and/or disease processes (e.g., proteases and protease inhibitors, protein kinases, and phosphatases)[36-39,84] as well as the physiology of the nervous system[37] and the formation of the cuticle (nematodes).[37,85]

The bioinformatic prediction and prioritization of new drug targets involves "filtering"[86,87] and usually includes inferring targets based on key requirements and principles.[88-92] First, target proteins should have one or more crucial (essential) roles in fundamental biological processes of the pathogen and/or vector, such that the disruption of the molecule or its gene will damage and/or kill both or either and thus disrupt disease transmission or disease itself, but not adversely affect the host.[91,93] Presently, for parasitic nematodes, the inference of drug target candidates is mostly assisted through the use of extensive information on function and essentiality in a range of eukaryotic organisms, including *C. elegans* (WormBase at http://www.wormbase.org). Because most effective drugs achieve their activity by competing with endogenous, small molecules for a binding site on a target protein,

the amino acid sequences predicted from essential genes should be screened for the presence of relatively conserved ligand-binding domains.[94,95] Lists of inhibitors, known based on experimental evidence, to specifically bind to such domains, can be compiled. However, the predictions made are intended to support hypothesis-driven or applied research and thus require extensive experimental validation. The challenge for *D. immitis* is its long, indirect life cycle and the time, cost, and labor involved in producing different stages of the parasite for molecular investigations. Moreover, the parasite cannot be produced or maintained for long periods of time in culture in vitro, which makes the in vitro testing of gene function (as is routinely performed for *C. elegans*)[96] challenging. Nevertheless, assays are available for some developmental stages, such as microfilariae,[97] which might allow functional genomic studies to be conducted in the parasite itself.

Many Implications in Fundamental and Applied Areas

Although various studies have given improved insights into the epidemiology of *D. immitis* using molecular methods,[98-100] there has been very limited study of fundamental molecular biology of this parasite, its vectors, and its relationship with the mammalian host and CHWD as well as possible drug or insecticide resistance with advanced -omic technologies. Using modern genomic (NGS) and bioinformatic technologies to understand the systems biology of *D. immitis* and CHWD should lead to the design of radically new intervention and management strategies against *D. immitis*.

For example, from a fundamental viewpoint, genome-wide sequencing and the definition of a wide range of genetic markers for use in specific and sensitive diagnostic tools could provide a solid foundation for addressing questions regarding the complex network of biological and ecological factors involved in parasite-vector-host-environment interactions and the immunological idiosyncrasies of receptive hosts in endemic regions as well as the role of asymptomatic, chronically infected animals and those infected with multiple pathogens.[2] It would also be very useful to investigate the resistance and susceptibility of, for example, particular breeds, haplotypes, or genotypes of dogs to *D. immitis* and its vectors. For instance, elucidating the relationship between host genotype and phenotype (degree of disease expression) in response to CHWD and/or intervention approach (e.g., treatment or vaccination) would be particularly informative and could provide a profound understanding of the genetic basis of disease. From an epidemiological perspective, changes in the temporal and spatial distribution of *D. immitis*, vectors, and/or their hosts, as a result of climatic change and global warming, might also be monitored with metagenomic approaches.[101] Obviously, another question of major importance is whether *D. immitis* is developing genetic resistance against macrocyclic lactones, which have been routinely used (over many years), particularly in the United States, parts of Europe, and Australia, to prevent CHWD. Although it was originally proposed that *D. immitis* is highly unlikely to develop resistance against such compounds,[102] there is now some recent genetic and experimental evidence[6,7] indicating that some populations of this nematode are less susceptible to macrocyclic lactones than others, which suggests an emergence of genetic resistance against this class of compounds.[97] After the determination of a reference genome for *D. immitis*, it will be possible, using NGS, to genetically compare multiple populations of adult worms indicated to be resistant or susceptible to one or more macrocyclic lactones (via the testing of microfilariae from the blood from infected dogs),[7] and then to establish a clear link between a drug-resistant phenotype and its genotype on a genome-wide scale. Transcriptomic analyses could be used to underpin genomic comparisons by providing evidence of differential transcription of genes (upregulation or downregulation) between resistant and susceptible worms (e.g., after the treatment of microfilariae with sublethal doses of macrocycline lactone). Ultimately, establishing genomic differences would allow the definition of genetic markers that could then be used in a molecular test for the direct and specific molecular detection of drug-resistant worms (via DNA from microfilariae or adult worms) in the blood from infected dogs in endemic regions in which preventative treatment is routinely or commonly used. Clearly, these examples indicate that there are many exciting fundamental areas and questions to tackle using modern genomic tools.

From an applied perspective, the prediction and prioritization of drug targets in *D. immitis* or repellants against vectors is another area of major importance. NGS will provide the efficiency and depth of coverage required to rapidly define de novo the complete genomes of *D. immitis*, vectors and canine and other hosts. Repertoires of drug targets could be inferred on a genome-wide scale. For instance, the *D. immitis* kinome (the complete set of kinase genes in the genome) could represent a unique opportunity for the design of pathogen-selective inhibitors[103] for subsequent validation by high throughput screening.[104-107] Other targets, such as guanosine triphosphatases and protein phosphatases,[103,108,109] might also represent attractive drug target candidates for *D. immitis*, but have not yet been examined in a systematic manner or on a large scale.

In conclusion, the combined use of genomic, transcriptomic, and proteomic datasets will be central to identifying other groups of molecules essential to the development and survival of *D. immitis* and will pave the way for the design of novel classes of small molecular inhibitors. Clearly, an integrated use of -omic technologies will enable investigations of the systems biology of *D. immitis* and CHWD on a scale never before possible and will provide unprecedented opportunities for the development of entirely new intervention and diagnostic strategies. The ability to explore heartworm in this way will also open up numerous opportunities to investigate key filarioid nematodes of humans and other animals.

References

1. Otranto D, Wall R: New strategies for the control of arthropod vectors of disease in dogs and cats. Med Vet Entomol 22:291-302, 2008

2. Otranto D, Dantas-Torres F, Breitschwerdt EB: Managing canine vector-borne diseases of zoonotic concern: part one. Trends Parasitol 25:157-163, 2009
3. Otranto D, Dantas-Torres F, Breitschwerdt EB: Managing canine vector-borne diseases of zoonotic concern: part two. Trends Parasitol 25:228-235, 2009
4. McCall JW, Genchi C, Kramer LH, et al: Heartworm disease in animals and humans. Adv Parasitol 66:193-285, 2008
5. Lee AC, Montgomery SP, Theis JH, et al: Public health issues concerning the widespread distribution of canine heartworm disease. Trends Parasitol 26:168-173, 2010
6. Bourguinat C, Keller K, Prichard RK, et al: Genetic polymorphism in *Dirofilaria immitis*. Vet Parasitol 176:368-373, 2011
7. Bourguinat C, Keller K, Bhan A, et al: Macrocyclic lactone resistance in *Dirofilaria immitis*. Vet Parasitol 181:388-392, 2011
8. Hoch H, Strickland K: Canine and feline dirofilariasis: life cycle, pathophysiology, and diagnosis. Compend Contin Educ Vet 30:133-140, 2008
9. Sparagano OA, De Luna CJ: From population structure to genetically-engineered vectors: new ways to control vector-borne diseases? Infect Genet Evol 8:520-525, 2008
10. Lefevre T, Thomas F: Behind the scene, something else is pulling the strings: emphasizing parasitic manipulation in vector-borne diseases. Infect Genet Evol 8:504-519, 2008
11. Coutinho-Abreu IV, Zhu KY, Ramalho-Ortigao M: Transgenesis and paratransgenesis to control insect-borne diseases: current status and future challenges. Parasitol Int 59:1-8, 2010
12. Holzmuller P, Grebaut P, Cuny G, et al: Tsetse flies, trypanosomes, humans and animals: what is proteomics revealing about their crosstalks? Expert Rev Proteomics 7:113-126, 2010
13. Walker JM, Rapley R (eds): Molecular Biology and Biotechnology (ed 5). Cambridge, The Royal Society of Chemistry Press, 2009
14. Stathopoulos A, Levine M: Whole-genome expression profiles identify gene batteries in *Drosophila*. Dev Cell 3:464-465, 2002
15. Gupta V, Oliver B: *Drosophila* microarray platforms. Brief Funct Genomic Proteomic 2:97-105, 2003
16. Vibranovski MD, Lopes HF, Karr TL, et al: Stage-specific expression profiling of *Drosophila* spermatogenesis suggests that meiotic sex chromosome inactivation drives genomic relocation of testis-expressed genes. PLoS Genet 5:e1000731, 2009
17. Spencer WC, Zeller G, Watson JD, et al: A spatial and temporal map of *C. elegans* gene expression. Genome Res 21:325-341, 2011
18. Wang Y, Chen J, Wei G, et al: The *Caenorhabditis elegans* intermediate-size transcriptome shows high degree of stage-specific expression. Nucleic Acids Res 39:5203-5214, 2011
19. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74: 5463-5467, 1977
20. Sanger F, Air GM, Barrell BG, et al: Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265:687-695, 1977
21. DeRisi J, Penland L, Brown PO, et al: Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 14:457-460, 1996
22. Wang AM, Doyle MV, Mark DF: Quantitation of mRNA by the polymerase chain reaction. Proc Natl Acad Sci U S A 86:9717-9721, 1989
23. Clifton SW, Mitreva M: Strategies for undertaking expressed sequence tag (EST) projects. Methods Mol Biol 533:13-32, 2009
24. Margulies M, Egholm M, Altman WE, et al: Genome sequencing in microfabricated high-density picolitre reactors. Nature 437:376-380, 2005
25. Bentley DR, Balasubramanian S, Swerdlow HP, et al: Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456:53-59, 2008
26. Harris TD, Buzby PR, Babcock H, et al: Single-molecule DNA sequencing of a viral genome. Science 320:106-109, 2008
27. Pandey V, Nutter RC, Prediger E: Applied Biosystems SOLiD™ System: ligation-based sequencing, in Milton JM (ed): Next Generation Genome Sequencing: Towards Personalized Medicine. Australia, Wiley, 2008, pp 29-41
28. Morozova O, Marra MA: Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255-264, 2008
29. Mardis ER: The impact of next-generation sequencing technology on genetics. Trends Genet 24:133-141, 2008
30. Wang Z, Gerstein M, Snyder M: RNA-seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57-63, 2009
31. Mardis ER: Next-generation DNA sequencing methods. Ann Rev Genom Human Genet 9:387-402, 2008
32. Jex AR, Hu M, Littlewood DT, et al: Using 454 technology for long-PCR based sequencing of the complete mitochondrial genome from single *Haemonchus contortus* (Nematoda). BMC Genomics 9:11, 2008
33. Young ND, Campbell BE, Hall RS, et al: Unlocking the transcriptomes of two carcinogenic parasites, *Clonorchis sinensis* and *Opisthorchis viverrini*. PLoS Negl Trop Dis 4:e719, 2010
34. Young ND, Hall RS, Jex AR, et al: Elucidating the transcriptome of *Fasciola hepatica*—a key to fundamental and biotechnological discoveries for a neglected parasite. Biotechnol Adv 28:222-231, 2010
35. Young ND, Jex AR, Cantacessi C, et al: Progress on the transcriptomics of carcinogenic liver flukes of humans–unique biological and biotechnological prospects. Biotechnol Adv 28: 859-870, 2010
36. Wang Z, Abubucker S, Martin J, et al: Characterizing *Ancylostoma caninum* transcriptome and exploring nematode parasitic adaptation. BMC Genomics 11:307, 2010
37. Cantacessi C, Campbell BE, Young ND, et al: Differences in transcription between free-living and CO_2-activated third-stage larvae of *Haemonchus contortus*. BMC Genomics 11: 266, 2010
38. Cantacessi C, Mitreva M, Jex AR, et al: Massively parallel sequencing and analysis of the *Necator americanus* transcriptome. PLoS Negl Trop Dis 4:e684, 2010
39. Cantacessi C, Jex AR, Hall RS, et al: A practical, bioinformatic workflow system for large data sets generated by next-generation sequencing. Nucleic Acids Res 38:e171, 2010
40. Cantacessi C, Mitreva M, Campbell BE, et al: First transcriptomic analysis of the economically important parasitic nematode, *Trichostrongylus colubriformis*, using a next-generation sequencing approach Infect Genet Evol 10:1199-1207, 2010
41. Cantacessi C, Gasser RB, Strube C, et al: Deep insights into *Dictyocaulus viviparus* transcriptomes provides unique prospects for new drug targets and disease intervention. Biotechnol Adv 29:261-271, 2011

42. Huang X, Madan A: CAP3: A DNA sequence assembly program. Genome Res 9:868-877, 1999
43. Iseli C, Jongeneel CV, Bucher P: ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc Int Conf Intell Syst Mol Biol 1:138-148, 1999
44. Conesa A, Götz S, García-Gómez JM, et al: Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676, 2005
45. Nagaraj SH, Deshpande N, Gasser RB, et al: ESTExplorer: an expressed sequence tag (EST) assembly and annotation platform. Nucleic Acids Res 35:W143-W147, 2007
46. Hunter S, Apweiler R, Attwood TK, et al: InterPro: the integrative protein signature database. Nucleic Acids Res 37: D211-D215, 2009
47. Nagaraj SH, Gasser RB, Ranganathan S: A hitchhiker's guide to expressed sequence tag (EST) analysis. Brief Bioinfomatics 8:6-21, 2007
48. Falgueras J, Lara AJ, Fernandez-Poso N, et al: SeqTrim: a high throughput pipeline for pre-processing any type of sequence read. BMC Bioinformatics 11:38, 2010
49. Myers EW: Toward simplifying and accurately formulating fragment assembly. J Comput Biol 2:275-290, 1995
50. Idury RM, Waterman MS: A new algorithm for DNA sequence assembly. J Comput Biol 2:291-306, 1995
51. Zerbino DR, Birney E: Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821-829, 2008
52. Sutton GG, White O, Adams MD, et al: TIGR assembler: a new tool for assembling large shotgun sequencing projects. Genome Sci Technol 1:9-19, 1995
53. Green P: Documentation for PHRAP. Genome Center, University of Washington, Seattle. Available from: http://www.phrap.org/phredphrap/phrap.html. Accessed September 24, 2011.
54. Huang X, Wang J, Aluru S, et al: PCAP: a whole genome assembly program. Genome Res 13:2164-2170, 2003
55. Chevreux B: MIRA: an automated genome and EST assembler. PhD Thesis, German Cancer Research Center Heidelberg, Duisburg, Germany, 2005
56. Scheibye-Alsing K, Hoffmann S, Frankel A, et al: Sequence assembly. Comput Biol Chem 33:121-136, 2009
57. Miller JR, Koren S, Sutton G: Assembly algorithms for next-generation sequencing data. Genomics 95:315-327, 2010
58. Warren RL, Sutton GC, Jones SJM, et al: Assembling millions of short DNA sequences using SSAKE. Bioinformatics 23:500-501, 2007
59. Hernandez D, François P, Farinelli L, et al: De novo bacterial genome sequencing: millions of very short reads assembled on a desktop computer. Genome Res 18:802-809, 2008
60. Li R, Li Y, Kristiansen K, et al: SOAP: short oligonucleotide alignment program. Bioinformatics 24:713-714, 2008
61. Altschul SF, Gish W, Miller W, et al: Basic Local Alignment Search Tool. J Mol Biol 215:403-410, 1990
62. Fukunishi Y, Hayashizaki Y: Amino acid translation program for full-length cDNA sequences with frameshift errors. Physiol Genomics 5:81-87, 2001
63. Min XJ, Butler G, Storms R, et al: OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res 33:W677-W680, 2005
64. Hofmann K, Bucher P, Falquet L, et al: The Prosite Database, its status in 1999. Nucleic Acids Res 27:215-219, 1999
65. Ashburner M, Ball CA, Blake JA, et al: Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. 25:25-29, 2000
66. Attwood TK, Croning MD, Flower DR, et al: Prints-S: the database formerly known as prints. Nucleic Acids Res 28:225-227, 2000
67. Bateman A, Birney E, Durbin R, et al: The Pfam protein families database. Nucleic Acids Res 28:263-266, 2000
68. Benson DA, Karsch-Mizrachi I, Lipman DJ, et al: GenBank Nucleic Acids Res 30:17-20, 2002
69. Stoesser G, Baker W, van den Broek A, et al: The EMBL nucleotide sequence database. Nucleic Acids Res 30:21-26, 2002
70. Tateno Y, Imanishi S, Miyazaki S, et al: DNA Data Bank of Japan (DDBJ) for genome scale research in life sciences. Nucleic Acids Res 30:27-30, 2002
71. Shumway M, Cochrane G, Sugawara H: Archiving next generation sequencing data. Nucleic Acids Res 38:D870-D871, 2010
72. Wheeler DL, Church DM, Lash AE, et al: Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 29:11-16, 2001
73. Harris TW, Antoshechkin I, Bieri T, et al: WormBase: a comprehensive resource for nematode research. Nucleic Acids Res 38:D463-D467, 2010
74. Schwarz EM, Antoshechkin I, Bastiani C, et al: WormBase: better software, richer content. Nucleic Acids Res 34:D475-D478, 2006
75. Tweedie S, Ashburner M, Falls K, et al: FlyBase: enhancing *Drosophila* gene ontology annotations. Nucleic Acids Res 37: D555-D559, 2009
76. Ghedin E, Wang S, Spiro D, et al: Draft genome of the filarial nematode parasite *Brugia malayi*. Science 317:1756-1760, 2007
77. Jarvie T, Harkins T: Transcriptome sequencing with the Genome Sequencer FLX system. Nat Methods 5:6-8, 2008
78. Pepke S, Wold B, Mortazavi A: Computational approaches to the analysis of ChIP-seq and RNA-seq data. Nat Methods 6:S22-S32, 2009
79. Reinhardt JA, Baltrus DA, Nishimura MT, et al: De novo assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*. Genome Res 19:294-305, 2009
80. Nagarajan H, Butler JE, Klimes A, et al: De novo assembly of the complete genome of an enhanced electricity-producing variant of *Geobacter sulfurreducens* using only short reads. PLoS One 5:e10922, 2010
81. Tsai IJ, Otto TD, Berriman M: Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol 11:R41, 2010
82. Gregory SG, Barlow KF, McLay KE, et al: The DNA sequence and biological annotation of human chromosome 1. Nature 441:315-321, 2006
83. Benitez-Paez A: Considerations to improve functional annotations in biological databases. OMICS 13:527-535, 2009
84. Huang CQ, Gasser RB, Cantacessi C, et al: Genomic-bioinformatic analysis of transcripts enriched in the third-stage larva of the parasitic nematode *Ascaris suum*. PLoS Negl Trop Dis 2:e246, 2008
85. Rodriguez-Valle M, Lew-Tabor A, Gondro C, et al: Comparative microarray analysis of *Rhipicephalus (Boophilus) microplus* expression profiles of larvae pre-attachment and feeding

adult female stage on *Bos indicus* and *Bos taurus* cattle. BMC Genomics 11:437, 2010
86. Geary TG, Thompson DP, Klein RD: Mechanism-based screening: discovery of the next generation of anthelmintics depends upon more basic research. Int J Parasitol 29:105-112, 1999
87. McCarter JP: Genomic filtering: an approach to discovering novel antiparasitics. Trends Parasitol 20:462-468, 2004
88. Krasky A, Rohwer A, Schroeder J, et al: A combined bioinformatics and chemoinformatics approach for the development of new antiparasitic drugs. Genomics 89:36-43, 2007
89. Caffrey CR, Rohwer A, Oellien F, et al: A comparative chemogenomics strategy to predict potential drug targets in the metazoan pathogen, *Schistosoma mansoni*. PLoS One 4:e4413, 2009
90. Doyle MA, Gasser RB, Woodcroft BJ, et al: Drug target prediction and prioritization: using orthology to predict essentiality in parasite genomes. BMC Genomics 11:222, 2010
91. Seib KL, Dougan G, Rappuoli R: The key role of genomics in modern vaccine and drug design for emerging infectious diseases. PLoS Genet 5:e1000612, 2009
92. Woods DJ, Knauer CS: Discovery of veterinary antiparasitic agents in the 21st century: a view from industry. Int J Parasitol 40:1177-1181, 2010
93. Gilleard JS, Woods DJ, Dow JAT: Model-organism genomics in veterinary parasite drug-discovery. Trends Parasitol 21: 302-305, 2005
94. Hopkins AL, Groom CR: The druggable genome. Nature 1:727-730, 2002
95. Chang A, Scheer M, Grote A, et al: BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucleic Acids Res 37:D588-D592, 2009
96. Hope A (ed): *C. elegans*: A Practical Approach, Oxford University Press, Oxford, 1999
97. Bourguinat C, Keller K, Blagburn B, et al: Correlation between loss of efficacy of macrocyclic lactone heartworm anthelmintics and P-glycoprotein genotype. Vet Parasitol 176:374-381, 2011
98. Yates DM, Wolstenholme AJ: An ivermectin-sensitive glutamate-gated chloride channel subunit from *Dirofilaria immitis*. Int J Parasitol 34:1075-1081, 2004
99. Yates DM, Wolstenholme AJ: *Dirofilaria immitis*: identification of a novel ligand-gated ion channel-related polypeptide Exp Parasitol 108:182-185, 2004
100. Yin Y, Martin J, McCarter JP, et al: Identification and analysis of genes expressed in the adult filarial parasitic nematode *Dirofilaria immitis*. Int J Parasitol 36:829-839, 2006
101. Cortinas MR, Guerra MA, Jones CJ, et al: Detection, characterization, and prediction of tick-borne disease foci. Int J Med Microbiol 291:11-20, 2002
102. Prichard RK: Is anthelmintic resistance a concern for heartworm control? What can we learn from the human filariasis control programs? Vet Parasitol 133:243-253, 2005
103. Liotta F, Siekierka JJ: Apicomplexa, trypanosoma and parasitic nematode protein kinases as antiparasitic therapeutic targets. Curr Opin Invest Drugs 11:147-156, 2010
104. Woods D, Butler C, Williams T, et al: Receptor-based discovery strategies for insecticides and parasiticides: a review. Adv Exp Med Biol 692:1-9, 2010
105. Smout MJ, Kotze AC, McCarthy JS, et al: A novel high throughput assay for anthelmintic drug screening and resistance diagnosis by real-time monitoring of parasite motility. PLoS Negl Trop Dis 4:e885, 2010
106. Engel JC, Ang KK, Chen S, et al: Image-based high-throughput drug screening targeting the intracellular stage of *Trypanosoma cruzi*, the agent of Chagas' disease. Antimicrob Agents Chemother 54:3326-3334, 2010
107. Siqueira-Neto JL, Song OR, Oh H, et al: Antileishmanial high-throughput drug screening reveals drug candidates with new scaffolds. PLoS Negl Trop Dis 4:e675, 2010
108. Bajsa J, Duke SO, Tekwani BL: *Plasmodium falciparum* serine/threonine phosphoprotein phosphatases (PPP): from housekeeper to the 'holy grail.' Curr Drug Targets 9:997-1012, 2008
109. Price HP, Peltan A, Stark M, et al: The small GTPase ARL2 is required for cytokinesis in *Trypanosoma brucei*. Mol Biochem Parasitol 173:123-131, 2010