

# High throughput sequencing of the *Angiostrongylus cantonensis* genome: a parasite spreading worldwide

ALESSANDRA L. MORASSUTTI<sup>1\*</sup>, ANDREY PERELYGIN<sup>2</sup>,  
MARCOS O. DE CARVALHO<sup>3</sup>, LEANDRO NASCIMENTO LEMOS<sup>4</sup>,  
PAULO MARCOS PINTO<sup>4</sup>, MIKE FRACE<sup>2</sup>, PATRICIA P. WILKINS<sup>2</sup>,  
CARLOS GRAEFF-TEIXEIRA<sup>1</sup> and ALEXANDRE J. DA SILVA<sup>2</sup>

<sup>1</sup> Laboratório de Biologia Parasitária da Faculdade de Biociências e Laboratório de Parasitologia Molecular do Instituto de Pesquisas Biomédicas da Pontifícia Universidade do Rio Grande do Sul (PUCRS), Av Ipiranga 6690, 90690-900 Porto Alegre RS, Brasil

<sup>2</sup> Division of Parasitic Diseases & Malaria, Center for Global Health, Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA, 30329, USA

<sup>3</sup> Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre RS, Brasil

<sup>4</sup> Universidade Federal do Pampa – Campus São Gabriel, Av. Antônio Trilha, 1847 – São Gabriel – RS – CEP: 97300-000

(Received 19 March 2013; revised 9 April 2013; accepted 13 April 2013; first published online 17 July 2013)

## SUMMARY

*Angiostrongylus cantonensis* is a parasitic nematode of rodents and a leading aetiological agent of eosinophilic meningitis in humans. Definitive diagnosis is difficult, often relying on immunodiagnostic methods which utilize crude antigens. New immunodiagnostic methods based on recombinant proteins are being developed, and ideally these methods would be made available worldwide. Identification of diagnostic targets, as well as studies on the biology of the parasite, are limited by a lack of molecular information on *Angiostrongylus* spp. available in databases. In this study we present data collected from DNA random high-throughput sequencing together with proteomic analyses and a cDNA walking methodology to identify and obtain the nucleotide or amino acid sequences of unknown immunoreactive proteins. 28 080 putative ORFs were obtained, of which 3371 had homology to other deposited protein sequences. Using the *A. cantonensis* genomic sequences, 156 putative ORFs, matching peptide sequences obtained from previous proteomic studies, were considered novel, with no homology to existing sequences. Full-length coding sequences of eight antigenic target proteins were obtained. In this study we generated not only the complete nucleotide sequences of the antigenic protein targets but also a large amount of genomic data which may help facilitate future genomic, proteomic, transcriptomic or metabolomic studies on *Angiostrongylus*.

Key words: *Angiostrongylus cantonensis*, genome sequencing, diagnostic targets.

## INTRODUCTION

*Angiostrongylus cantonensis* is a parasitic nematode of the pulmonary arteries in rodents and the leading aetiological agent of eosinophilic meningitis in humans. Humans become infected by ingestion of the L3 present in vegetables, water or through ingestion of raw or undercooked infected molluscs as intermediate host or other animals that may serve as paratenic hosts (Wang *et al.* 2008). To date more than 2000 cases of the disease have been reported, mostly in Southeastern Asia (Wang *et al.* 2008). Cerebral angiostrongyliasis is an emerging global public health problem since the disease has been increasingly observed around the world, including countries where the disease had not been recognized previously, such as in Brazil, the Caribbean, Ecuador and South Africa (Caldeira *et al.* 2007; Diaz, 2008;

Wang *et al.* 2008; Pincay *et al.* 2009; Maldonado *et al.* 2010; Archer *et al.* 2011).

In humans the parasite is unable to complete its life cycle and eventually dies in the meninges. As a consequence the L1 are not found in feces preventing a simple parasitological diagnosis through identification of the parasite in stool samples. Fourth-stage larvae are rarely seen in the cerebrospinal fluid (CSF) and yet this is considered the gold standard for diagnosis (Graeff-Teixeira *et al.* 2009). In general *A. cantonensis* eosinophilic encephalitis can be distinguished from encephalitis caused by viruses, bacteria and protozoan parasites. Nevertheless differential aetiological diagnosis is required since there are other tissue-dwelling parasites that may cause eosinophilic meningitis, such as *Gnathostoma* and *Baylisascaris* (Graeff-Teixeira *et al.* 2009).

Several studies have identified specific and sensitive targets for immunodiagnosis of angiostrongyliasis. Most of the tests described in the literature use crude extract preparations obtained from the nematode. Recombinant protein antigens are needed,

\* Corresponding author. Instituto de Pesquisas Biomédicas da PUCRS, Avenida Ipiranga 6690, 2 andar, Sala 20, CEP: 90690-900 Porto Alegre RS, Brazil. E-mail: almorassutti@gmail.com

but their identification is hindered by the lack of molecular information on *Angiostrongylus* spp. Less than 800 nucleotide sequences, 2631 ESTs, 699 protein sequences and only 43 genes of *Angiostrongylus* were deposited at GenBank as of March 2013, including the mitochondrial genome sequences of *A. costaricensis* and *A. cantonensis* (Lv *et al.* 2012). To date, most *Angiostrongylus* proteomics studies have been performed using tandem mass spectrometry using orthologous sequences of related organisms for peptide mass comparison (Rebello *et al.* 2011; Morassutti *et al.* 2012a). Unfortunately, the value of results from such analyses is limited because the amount of *Angiostrongylus* nucleotide or amino acid data in searchable databases is so small.

Next generation sequencing approaches have allowed rapid sequencing and annotation of genomes from a myriad of organisms. Data collected from such platforms allow researchers to close significant gaps towards the identification of previously unidentified open reading frames that translate into potential diagnostic targets. In this study we combined random high-throughput sequencing of the *A. cantonensis* genome together with proteomics analyses and cDNA walking methodology for novel immunoreactive protein discovery; this strategy is sometimes known as the dirty genome sequencing approach (Greub *et al.* 2009).

## MATERIALS AND METHODS

### Biological materials

Adult *A. cantonensis* worms were recovered from experimentally infected *Rattus norvegicus*. These worms were originally obtained from the Department of Parasitology, Akita Medical School, Japan and have been maintained in our laboratory since 1997. Wistar rats and *Biomphalaria glabrata* were used as definitive and intermediate hosts, respectively. Rats were infected with 104 larvae by gavage inoculation. Animals were euthanized 42 days after infection and female worms were isolated from their lungs and stored at  $-80^{\circ}\text{C}$  in *RNAlater* (Qiagen, Inc., Valencia, CA) to preserve the mRNA, according to manufacturer's instructions until use.

### DNA sequencing

Total genomic DNA of *A. cantonensis* was extracted from female worms using the Genra Puregene Tissue Kit (Qiagen, Valencia, CA, USA), according to the supplementary protocol for purification of archive-quality DNA from nematodes ([www.qiagen.com/literature/render.aspx?id=103616](http://www.qiagen.com/literature/render.aspx?id=103616)). From genomic DNA, random shotgun libraries were produced using standard protocols for nebulization, AMPure XP magnetic bead cleanup (Beckman Coulter, Inc.,

Beverly, MA, USA), end-polishing, adaptor ligation and size selection. Roche-454 bead libraries were amplified using emulsion PCR followed by bead isolation and sequencing using long-read GS-FLX Titanium chemistry. Illumina libraries were flowcell amplified using a cBot single read cluster generation kit. These were then sequenced using v4 cycle sequencing chemistry on a Genome Analyzer IIe sequencer. Roche-454 data was cumulatively assembled using the Roche Newbler *de novo* assembler. Data assemblies which combine Roche-454 reads and the shorter Illumina reads were performed using CLC Genomics Workbench v 4.9 (Cambridge, MA, USA).

### Sequence analysis

Contigs were examined by the open access software *AUGUSTUS* version 2.4 (Stanke *et al.* 2008) (<http://bioinf.uni-greifswald.de/augustus>) to predict ORFs using NCBI *Caenorhabditis elegans* sequences as a model. For protein annotation the NCBI database was accessed to identify homology to other organisms associated with the protein sequence generated here using the PyCogent toolkit (Knight *et al.* 2007). The absolute and relative frequencies were calculated in the R project statistical software (Winham and Motsinger-Reif, 2011). Immunogenic protein data were obtained in other related projects (Morassutti *et al.* 2012a, 2012b) that used two-dimensional electrophoresis and tandem mass spectrometry (MS/MS) experiments for antigenic protein identification. From these experiments the generated peptide database was re-searched against the *AUGUSTUS* genome data annotation using MyriMatch software (Tabb *et al.* 2007). Final annotation of each peptide, score ranking and artifact exclusion were done using the IDPicker software (Ma *et al.* 2009).

### RNA isolation and cDNA synthesis

Approximately 30 mg of *A. cantonensis* worms were homogenized in 600  $\mu\text{L}$  of Lysis Buffer RA1 using a T8 homogenizer (IKA WORKS, Inc., Wilmington, NC, USA) and total RNA was isolated using the NucleoSpin RNA II Kit (Machery-Nagel, Inc., Bethlehem, PA, USA), according to the manufacturer's protocol. Concentration of isolated RNA was measured on a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and 5  $\mu\text{g}$  of total RNA were converted into single-stranded cDNA using the SuperScript III First-Strand Synthesis SuperMix (Invitrogen, Carlsbad, CA, USA) with an oligo(dT)20 primer according to manufacturer's recommendations. The cDNA obtained was aliquoted and stored at  $-20^{\circ}\text{C}$ .

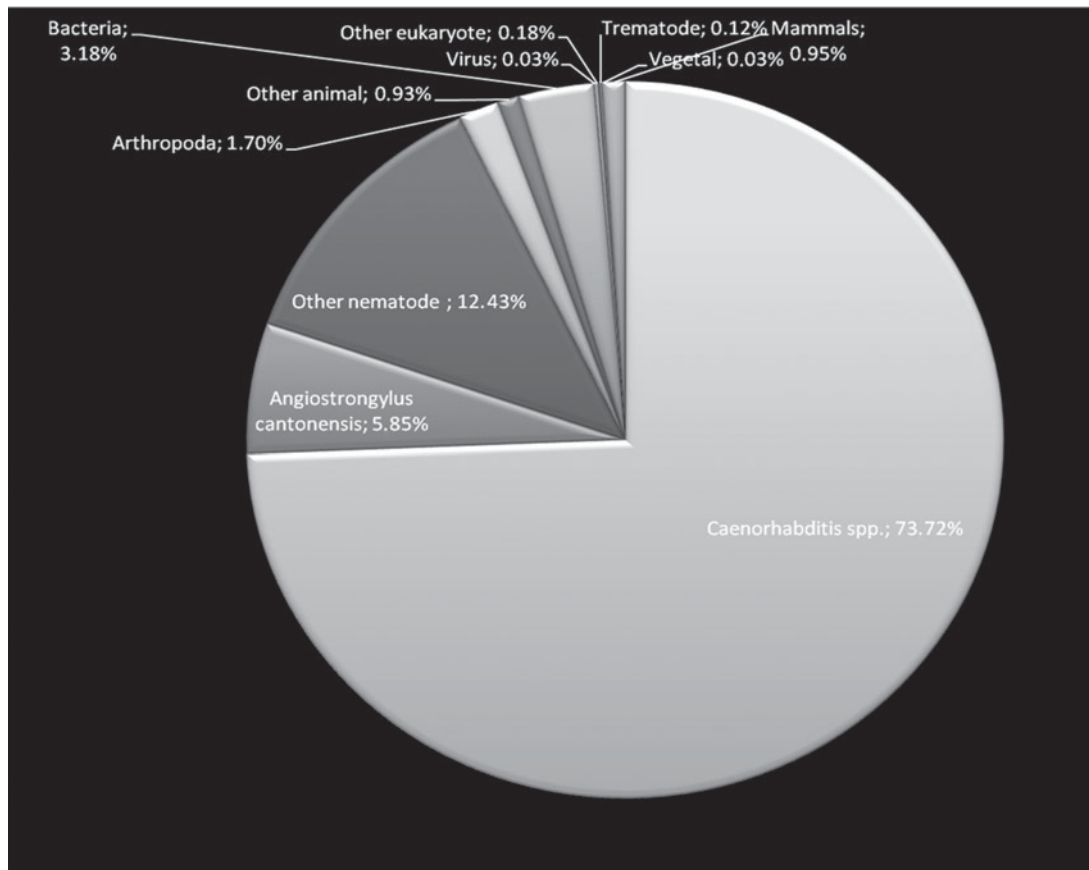


Fig. 1. Annotation of Open Reading Frame Sequence Homologies. Graphic shows the percentage of sequences homologous to different organisms.

#### *cDNA walking*

In order to achieve the complete sequence of each protein of interest, *C. elegans* sequences from GenBank were used to search against the *A. cantonensis* sequences by the tblastn program. Two sets of three gene-specific primers were then designed based on each partial mRNA sequence identified and used to extend the 5' and 3' portions, respectively, using the DNA Walking (DW) SpeedUp Kit (Seegene, Rockville, MD, USA). The multiple sequences obtained from both 5' and 3' walking experiments for each gene were aligned using the MegAlign computer program package (DNASTAR, Inc., Madison, WI, USA) to obtain a full-length consensus cDNA sequence.

Two full-length sequences were obtained directly by *AUGUSTUS* software, corresponding to the 14-3-3 phosphoserine-binding protein and a protein containing a nascent polypeptide-associated complex domain (NAC). Primers were designed to clone and confirm cDNA sequences by sequencing. Primers: 14-3-3: Forward: 5' CAC CAT GAC GGA CAA CAG GGG CGA; Reverse: 5' TCA GTT GGC ACC CTC TCC TTG TTC. NAC: Forward: 5' CAC CAT GGT TGC CGC GGT GGA AGT; Reverse: 5' TTA ACA AAT AAC TGA GAA TCAA. Complete sequences of the amplicons were

achieved by cycle sequencing using BigDye V3.1 chemistry (Applied Biosystems, Foster City, CA). Sequences were assembled, and then submitted to GenBank.

#### *Data access*

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ANFR00000000. The version described in this paper is the first version, ANFR01000000.

## RESULTS

#### *Genomic sequences*

The genomic DNA of *A. cantonensis* was randomly sequenced by pyrosequencing using 454 (Roche) technology. About 2 million reads were assembled into 141 351 contigs with an average length of 0.8 kb each. From these contigs using *AUGUSTUS* software, 28 080 putative ORFs were obtained, of which 3370 were annotated according to their homology with other sequences deposited in GenBank as shown in Fig. 1 (details in supplementary material – in Online version only). More than 90% of the annotated sequences matched previously deposited sequences

Table 1. Partial and full-length *A. cantonensis* galectin sequences

Gene	5'-end <sup>a</sup>	3'-end <sup>a</sup>	GenBank accession number
<i>lec-1</i>	2	3	CV826716; DN190352; DN190496
<i>lec-2</i>	3	3	DN190210
<i>lec-3</i>	7	2	ND
<i>lec-4</i>	1	4	ND
<i>lec-4</i>	8	6	ND
<i>lec-8</i>	2	5	DN190616; DN190836; JF320682
<i>lec-10</i>			CV826714; DQ384534

<sup>a</sup> Numbers of partial galectin sequences detected in our database. ND, not detected.

from nematodes other than *A. cantonensis*. Most of the sequences had homology to *Caenorhabditis* spp. (73%) sequences, but other nematode sequences were also represented (12%), such as *Loa loa*, *Brugia malayi*, *Haemonchus contortus*, *Ancylostoma* spp., *Ascaris suum*, among others (for more detail see supplemental material – in Online version only). Also, homology to other animal sequences, including arthropods, echinoderms, fish, and mammals, and 107 sequences (3.18%) homologous to bacteria organism were observed. The identified genes were separated in categories of their identity to relative homologous organisms (Fig. 1).

#### Immunoreactive protein identification

Previous experiments using a proteomics approach identified 43 spots considered immunodiagnostic targets (Morassutti *et al.* 2012a, 2012b). From these spots 34 proteins were identified after interrogating GenBank (Morassutti *et al.* 2012a, 2012b). Those peptide sequences were then re-searched using our *A. cantonensis* genomic sequences. 206 putative ORFs identified by the *AUGUSTUS* software matched sequences from previous MS/MS data. From these, only 50 ORFs were identified that shared homology with peptides of other organisms using GenBank data; 156 ORFs matched only peptide sequences from the *A. cantonensis* MS/MS database.

#### Identification and extension of *A. cantonensis* antigenic sequences

After sequence assembly we focused the database analysis to search for sequences of diagnostic interest from previous mass spectrometry analyses of antigenic proteins (Morassutti *et al.* 2012a, b): galectins, the 14-3-3-like proteins, and a protein containing NAC domain were chosen for recombinant expression in the future.

Six galectins were identified in our *A. cantonensis* database. Partial sequences of each individual *lec* gene *lec-1* ( $n=5$ ), *lec-2* ( $n=6$ ), *lec-3* ( $n=9$ ), *lec-4* ( $n=5$ ), *lec-8* ( $n=7$ ) were aligned using the MegAlign program searched with orthologous *C. elegans* sequences. In addition, *A. cantonensis* galectin sequences from the GenBank EST database were used for comparison (Table 1). Based on these alignments, the full-length *lec-1*, *lec-2* and *lec-8* gene sequences were predicted. Alignment of the partial *lec-8* sequence revealed a 118 bp deletion within GenBank EST DN190616. To verify the actual predicted proteins, three pairs of primers were designed and used to amplify corresponding full-length cDNAs. The partial gene sequences of *A. cantonensis* LEC-3, LEC-4 and LEC-5 were extended to full-length sequences by cDNA walking. Each was PCR amplified using 5' and 3' terminal gene-specific primer pairs. The complete sequences of the amplicons were obtained by individual sequencing and then submitted to GenBank under accession numbers: *lec-1* (JN133961), *lec-2* (JN133962), *lec-3* (JN133963), *lec-4* (JN133964), *lec-5* (JN133965) and *lec-8* (JN133966). The other two targets protein 14-3-3 and NAC full-length ORFs were directly obtained by *AUGUSTUS* prediction. Primers were designed for cDNA amplification of each predicted ORF and amplicons were sequenced for confirmation of the sequence and finally deposited in GenBank under the accession numbers JN133968 and JN133967, respectively.

#### DISCUSSION

Recently Greub and co-workers applied a combined strategy of genome sequencing and proteomics to discover immunogenic proteins of the emerging pathogen *Parachlamydia acanthamoebae* and named the approach the 'dirty genome approach'. This strategy demonstrated that incomplete genome sequence information could be used as a starting point to discover protein targets, especially for diagnostic target discovery (Greub *et al.* 2009).

As our previous mass spectrometry analyses of antigenic proteins have identified targets for angiostrongyliasis diagnosis (Morassutti *et al.* 2012a, 2012b), galectins, the 14-3-3 and a protein containing NAC domain were chosen for future recombinant expression because of their promising immunodiagnostic features (Schechtman *et al.* 2001; Siles-Lucas *et al.* 2008; Wang *et al.* 2009). In our previous studies, these proteins were identified based on homology with orthologous sequences even though GeneBank-deposited DNA sequences of *Angiostrongylus* spp. have been increasing after 2000 EST sequences deposited by Wang in 2011 and the whole mitochondrial genome of both *A. cantonensis* and *costaricensis* were published (Lv *et al.* 2012) including 12 protein coding genes, 22 transfer RNAs and 2 ribosomal

RNAs. However, none of those *Angiostrongylus* deposited sequences matched with our search and actual protein sequence was needed to produce recombinant proteins for diagnostics. Also, in our original studies we sequenced some peptides that did not match anything in the databases.

After contig assembly, the peptide sequences obtained from mass spectrometry analysis were compared with our *A. cantonensis* genomic database. *AUGUSTUS* software predicted 3370 proteins and 206 of those had homology with the peptide sequences we originally obtained by mass spectrometry (Morassutti *et al.* 2012a, 2012b). This is in stark comparison to the 26 proteins we identified using the existing GenBank entries (for annotation details, please see supplementary material – in Online version only). Thus, the remaining 156 protein sequences that matched the peptides generated by the MS/MS experiments were found only as a result of having the *A. cantonensis* genomic sequence database. This result indicates these 156 proteins are different from all published sequences in the GenBank and may be either exclusive to *Angiostrongylus* species or may be the first deposited sequences of a previously unrecognized relative. Although there is no known identity or function for these proteins, our proteomic data supports the *AUGUSTUS* prediction, especially since some of these were found from the native protein source, further demonstrating the value of this approach.

Galectins are a family of sugar-binding proteins with affinity for N-acetyl lactosamines, an interaction mediated via a conserved carbohydrate-recognition domain (CRD). It has a particular interest in parasite–host relationships because helminth galectins such as those from *B. malayi* and *Onchocerca volvulus* have been implicated as potential immune modulators (Klion and Donelson, 1994; Hewitson *et al.* 2008). Also, a *C. elegans* galectin-8 has been shown to be involved in host defence during bacterial infection (Ideo *et al.* 2009). In addition, galectins have also been identified as targets for vaccination as observed for the sheep gastrointestinal nematode *Trichostrongylus colubriformis* (Kiel *et al.* 2007) as well as targets for diagnosis. The LEC-10 protein of *A. cantonensis* has been shown to be immunoreactive (Hao *et al.* 2007) and we identified a Lec5 as a potential target for immunodiagnosis (Morassutti *et al.* 2012a).

In this paper we identified and extended the coding sequence of at least six galectins from *A. cantonensis*. Potentially more galectin sequences may be identified in the future taking into account that 11 members of the galectin family were previously described in the *C. elegans* genome (Nemoto-Sasaki *et al.* 2008). Interestingly both 5' and 3' terminal segments between ESTs DN190616 (previously deposited) and DN190836 are identical, while internal sequences are different. Most likely that these two

ESTs originated by alternatively spliced transcripts of the *lec-8* gene since EST DN190616 is missing an internal exon as compared with EST DN190836. Alternative splicing in an invertebrate galectin gene had been reported for the cnidarian organism *Hydra*, which generates two nemato-galectin transcripts and these are mutually exclusively expressed (Hwang *et al.* 2010). In addition, we were able to amplify only a single *lec-8* transcript corresponding to EST DN190836, which may suggest it could be expressed in a stage-dependent manner.

The *A. cantonensis* predicted genes matched a number of bacterial sequences (3.18%), some attaining about 80% similarity to the genus *Burkholderia* spp. and *Neisseria* spp. bacteria (see supplementary material – in Online version only). Interestingly, although previous research failed to demonstrate the presence of the endosymbiont organism *Wolbachia* spp. in *A. cantonensis* (Foster *et al.* 2008), another study isolated a gram-positive bacteria from the intestines of *A. cantonensis*, indicating worm microbiota is reduced in number and diversity (Graeff-Teixeira personal communication). However the identity of this bacillus was not identified. The observation of sequences with homology to bacterial sequences from genomic preparations of the whole worm DNA might be indicative of DNA co-isolation either because it is endosymbiont or contamination from the host. Further studies are in progress to amplify and identify this possible prokaryotic organism.

In this study we generated not only the completed sequences of some possible antigenic protein targets of *A. cantonensis*, but also greatly increased the amount of genomic sequence data available in *A. cantonensis* databases. This will enhance future genomic, proteomic, transcriptomic or metabolomic studies for *Angiostrongylus* and will facilitate the search for diagnostic and possibly therapeutic targets for angiostrongyliasis.

#### FINANCIAL SUPPORT

Financial support was provided by Brazilian agencies: CNPq (scholarship 201760/2009-6) and US Food Safety Initiative, CDC and APHL-USA. C. Graeff-Teixeira is a recipient of a CNPq PQ 1D fellowship and of grants 300456/2007-7 and 477260/2007-1.

#### REFERENCES

- Archer, C. E., Appleton, C. C., Mukaratirwa, S. and Hope, K. J. (2011). The rat lung-worm *Angiostrongylus cantonensis*: a first report in South Africa. *South African Medical Journal* **101**, 174–175.
- Caldeira, R. L., Mendonça, C. L., Goveia, C. O., Lenzi, H. L., Graeff-Teixeira, C., Lima, W. S., Mota, E. M., Pecora, I. L., Medeiros, A. M. and Carvalho, O. S. (2007). First record of molluscs naturally infected with *Angiostrongylus cantonensis* (Chen, 1935) (Nematoda: Metastrongylidae) in Brazil. *Memorias do Instituto Oswaldo Cruz* **102**, 887–889.

- Diaz, J. H. (2008). Helminthic eosinophilic meningitis: emerging zoonotic diseases in the South. *Journal of the Louisiana State Medical Society* **160**, 333–342.
- Foster, J. M., Kumar, S., Ford, L., Johnston, K. L., Ben, R., Graeff-Teixeira, C. and Taylor, M. J. (2008). Absence of *Wolbachia endobacteria* in the non-filarid nematodes *Angiostrongylus cantonensis* and *A. costaricensis*. *Parasites and Vectors* **1**, 31.
- Graeff-Teixeira, C., da Silva, A. C. and Yoshimura, K. (2009). Update on eosinophilic meningoencephalitis and its clinical relevance. *Clinical Microbiology Reviews* **22**, 322–348.
- Greub, G., Kebbi-Beghdadi, C., Bertelli, C., Collyn, F., Riederer, B. M., Yersin, C., Croxatto, A. and Raoult, D. (2009). High throughput sequencing and proteomics to identify immunogenic proteins of a new pathogen: the dirty genome approach. *PLoS ONE* **4**, e8423.
- Hao, L., Wu, K., Chen, X. G. and Wang, Q. (2007). Cloning, prokaryotic expression and immunoreactivity evaluation of *Angiostrongylus cantonensis* galectin. *Nan Fang Yi Ke Da Xue Xue Bao* **27**, 584–587.
- Hewitson, J. P., Harcus, Y. M., Curwen, R. S., Dowie, A. A., Atmadja, A. K., Ashton, P. D., Wilson, A. and Maizels, R. M. (2008). The secretome of the filarial parasite, *Brugia malayi*: proteomic profile of adult excretory-secretory products. *Molecular and Biochemical Parasitology* **160**, 8–21.
- Hwang, J. S., Takaku, Y., Momose, T., Adamczyk, P., Özbek, S., Ikeo, K., Khalturin, K., Hemmrich, G., Bosch, T. C., Holstein, T. W. et al. (2010). Nematogalectin, a nematocyst protein with GlyXY and galectin domains, demonstrates nematocyte-specific alternative splicing in *Hydra*. *Proceedings of the National Academy of Sciences USA* **107**, 18539–18544.
- Ideo, H., Fukushima, K., Gengyo-Ando, K., Mitani, S., Dejima, K., Nomura, K. and Yamashita, K. (2009). A *Caenorhabditis elegans* glycolipid-binding galectin functions in host defense against bacterial infection. *Journal of Biological Chemistry* **284**, 26493–26501.
- Kiel, M., Josh, P., Jones, A., Windon, R., Hunt, P. and Kongsuwan, K. (2007). Identification of immuno-reactive proteins from a sheep gastrointestinal nematode, *Trichostrongylus colubriformis*, using two-dimensional electrophoresis and mass spectrometry. *International Journal of Parasitology* **37**, 1419–1429.
- Klion, A. D. and Donelson, J. E. (1944). OvGalBP, a filarial antigen with homology to vertebrate galactoside-binding proteins. *Molecular and Biochemical Parasitology* **65**, 305–315.
- Knight, R., Maxwell, P., Birmingham, A., Carnes, J., Caporaso, J. G., Easton, B. C., Eaton, M., Hamady, M., Lindsay, H., Liu, Z. et al. (2007). PyCogent: a toolkit for making sense from sequence. *Genome Biology* **8**, R171.
- Lv, S., Zhang, Y., Zhang, L., Liu, Q., Liu, H. X., Hu, L., Wei, F. R., Steinmann, P., Graeff-Teixeira, C., Zhou, X. N. and Utzinger, J. (2012). The complete mitochondrial genome of the rodent intra-arterial nematodes *Angiostrongylus cantonensis* and *Angiostrongylus costaricensis*. *Parasitology Research* **111**, 115–123.
- Ma, Z. Q., Dasari, S., Chambers, M. C., Litton, M. D., Sobecki, S. M., Zimmerman, L. J., Halvey, P. J., Schilling, B., Drake, P. M., Gibson, B. W. and Tabb, D. L. (2009). IDPicker 2.0: improved protein assembly with high discrimination peptide identification filtering. *Journal of Proteome Research* **8**, 3872–3881.
- Maldonado, A., Jr., Simões, R. O., Oliveira, A. P., Motta, E. M., Fernandez, M. A., Pereira, Z. M., Monteiro, S. S., Torres, E. J. and Thiengo, S. C. (2010). First report of *Angiostrongylus cantonensis* (Nematoda: Metastrongylidae) in *Achatina fulica* (Mollusca: Gastropoda) from Southeast and South Brazil. *Memorias do Instituto Oswaldo Cruz* **105**, 938–941.
- Morassutti, A. L., Levert, K., Pinto, P. M., da Silva, A. J., Wilkins, P. and Graeff-Teixeira, C. (2012a). Characterization of *Angiostrongylus cantonensis* excretory-secretory proteins as potential diagnostic targets. *Experimental Parasitology* **130**, 26–31.
- Morassutti, A. L., Levert, K., Perelygin, A., da Silva, A. J., Wilkins, P. and Graeff-Teixeira, C. (2012b). The 31-kDa antigen of *Angiostrongylus cantonensis* comprises distinct antigenic glycoproteins. *Vector Borne and Zoonotic Diseases* **12**, 961–968.
- Nemoto-Sasaki, Y., Hayama, K., Ohya, H., Arata, Y., Kaneko, M. K., Saitou, N., Hirabayashi, J. and Kasai, K. (2008). *Caenorhabditis elegans* galectins LEC-1-LEC-11: structural features and sugar-binding properties. *Biochimica et Biophysica Acta* **1780**, 1131–1142.
- Pincay, T., García, L., Narváez, E., Decker, O., Martini, L. and Moreira, J. M. (2009). Angiostrongyliasis due to *Angiostrongylus cantonensis* in Ecuador. First report in South America. *Tropical Medicine and International Health* **14**, 37.
- Rebello, K. M., Barros, J. S., Mota, E. M., Carvalho, P. C., Perales, J., Lenzi, H. L. and Neves-Ferreira, A. G. (2011). Comprehensive proteomic profiling of adult *Angiostrongylus costaricensis*, a human parasitic nematode. *Journal of Proteomics* **74**, 1545–1559.
- Schechtman, D., Winnen, R., Tarrab-Hazdai, R., Ram, D., Shinder, V., Grevelid, C. G., Kunz, W. and Arnon, R. (2001). Expression and immunolocalization of the 14-3-3 protein of *Schistosoma mansoni*. *Parasitology* **123**, 573–582.
- Siles-Lucas, M., Merli, M. and Gottstein, B. (2008). 14-3-3 proteins in *Echinococcus*: their role and potential as protective antigens. *Experimental Parasitology* **119**, 516–523.
- Stanke, M., Diekhans, M., Baertsch, R. and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* **24**, 637–644.
- Tabb, D. L., Fernando, C. G. and Chambers, M. C. (2007). MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *Journal of Proteome Research* **6**, 654–661.
- Wang, Q. P., Lai, D. H., Zhu, X. Q., Chen, X. G. and Lun, Z. R. (2008). Human angiostrongyliasis. *Lancet Infectious Diseases* **8**, 621–630.
- Wang, Y., Cheng, Z., Lu, X. and Tang, C. (2009). *Echinococcus multilocularis*: proteomic analysis of the protoscoleces by two-dimensional electrophoresis and mass spectrometry. *Experimental Parasitology* **123**, 162–167.
- Winham, S. J. and Motsinger-Reif, A. A. (2011). An R package implementation of multifactor dimensionality reduction. *BioData Mining* **4**, 24.