

RESEARCH ARTICLE

# The *Anisakis* Transcriptome Provides a Resource for Fundamental and Applied Studies on Allergy-Causing Parasites

Fiona J. Baird<sup>1,2\*</sup>, Xiaopei Su<sup>3</sup>, Ibukun Aibinu<sup>4</sup>, Matthew J. Nolan<sup>5</sup>, Hiromu Sugiyama<sup>6</sup>, Domenico Otranto<sup>7</sup>, Andreas L. Lopata<sup>1,2</sup>, Cinzia Cantacessi<sup>3</sup>\*

**1** Centre for Biodiscovery & Molecular Development of Therapeutics, James Cook University, Townsville, Australia, **2** Australian Institute of Tropical Health and Medicine, James Cook University, Townsville, Australia, **3** Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom, **4** School of Applied Sciences, RMIT University, Bundoora, Australia, **5** Department of Pathology and Pathogen Biology, Royal Veterinary College, University of London, Hatfield, United Kingdom, **6** Department of Parasitology, National Institute of Infectious Diseases, Tokyo, Japan, **7** Department of Veterinary Medicine, University of Bari, Valenzano, Italy

☞ These authors contributed equally to this work.

\* [fiona.baird@jcu.edu.au](mailto:fiona.baird@jcu.edu.au) (FJB); [cc779@cam.ac.uk](mailto:cc779@cam.ac.uk) (CC)



**OPEN ACCESS**

**Citation:** Baird FJ, Su X, Aibinu I, Nolan MJ, Sugiyama H, Otranto D, et al. (2016) The *Anisakis* Transcriptome Provides a Resource for Fundamental and Applied Studies on Allergy-Causing Parasites. *PLoS Negl Trop Dis* 10(7): e0004845. doi:10.1371/journal.pntd.0004845

**Editor:** Ricardo Toshio Fujiwara, Universidade Federal de Minas Gerais, BRAZIL

**Received:** February 9, 2016

**Accepted:** June 22, 2016

**Published:** July 29, 2016

**Copyright:** © 2016 Baird et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Raw sequence data analysed in this article are available from the Sequence Read Archive database of NCBI under study number SRP070744. Assembled transcripts, together with corresponding computationally translated amino acid sequences and relevant annotation information, are available from [S2 Table](#).

**Funding:** This work was supported by a 'Collaborations Across Boundaries' grant and a seed grant from the Centre of Biodiscovery and Molecular Development of Therapeutics, James Cook University (FJB and CC). ALL is an Australian

## Abstract

### Background

Food-borne nematodes of the genus *Anisakis* are responsible for a wide range of illnesses (= anisakiasis), from self-limiting gastrointestinal forms to severe systemic allergic reactions, which are often misdiagnosed and under-reported. In order to enhance and refine current diagnostic tools for anisakiasis, knowledge of the whole spectrum of parasite molecules transcribed and expressed by this parasite, including those acting as potential allergens, is necessary.

### Methodology/Principal Findings

In this study, we employ high-throughput (Illumina) sequencing and bioinformatics to characterise the transcriptomes of two *Anisakis* species, *A. simplex* and *A. pegreffii*, and utilize this resource to compile lists of potential allergens from these parasites. A total of ~65,000,000 reads were generated from cDNA libraries for each species, and assembled into ~34,000 transcripts (= Unigenes); ~18,000 peptides were predicted from each cDNA library and classified based on homology searches, protein motifs and gene ontology and biological pathway mapping. Using comparative analyses with sequence data available in public databases, 36 (*A. simplex*) and 29 (*A. pegreffii*) putative allergens were identified, including sequences encoding 'novel' *Anisakis* allergenic proteins (i.e. cyclophilins and ABA-1 domain containing proteins).

Research Council (ARC) Future Fellow and his laboratory is supported by grants from the National Health and Medical Research Council of Australia (NHMRC). Research in the CC laboratory is supported by grants from the Isaac Newton Trust/ Wellcome Trust/University of Cambridge (grant number PNVM/GAAB) and the Royal Society (grant number PNAG/428). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

## Conclusions/Significance

This study represents a first step towards providing the research community with a curated dataset to use as a molecular resource for future investigations of the biology of *Anisakis*, including molecules putatively acting as allergens, using functional genomics, proteomics and immunological tools. Ultimately, an improved knowledge of the biological functions of these molecules in the parasite, as well as of their immunogenic properties, will assist the development of comprehensive, reliable and robust diagnostic tools.

## Author Summary

Nematodes within the genus *Anisakis* (i.e. *A. simplex* and *A. pegreffii*, also known as herring worms) are the causative agents of the fish-borne gastrointestinal illness known as ‘anisakiasis’, with infections resulting in symptoms ranging from mild gastric forms to severe allergic reactions leading to urticaria, gastrointestinal and/or respiratory signs and/or anaphylaxis (‘allergic anisakiasis’). Despite significant advances in knowledge of the pathobiology of allergic anisakiasis, thus far, the exact number and nature of parasite molecules acting as potential allergens are currently unknown; filling this gap is necessary to the development of robust and reliable diagnostics for allergic anisakiasis which, in turn, underpins the implementation of effective therapeutic strategies. Here, we use RNA-Seq and bioinformatics to sequence and annotate the transcriptomes of *A. simplex* and *A. pegreffii*, and, as an example application of these resources, mine this data to identify and characterise putative novel parasite allergens based on comparisons with known allergen sequence data from other parasites and other organisms.

## Introduction

Foodborne diseases include a range of illnesses transmitted *via* the ingestion of foodstuffs contaminated with a variety of chemical compounds and pathogenic microorganisms, including parasites [1–3]. Whilst the global disease burden and costs linked to these illnesses are difficult to estimate, it has been calculated that foodborne infections have cost the Australian economy alone \$1.249 billion in 2006 [4]. Amongst the parasites responsible for foodborne diseases, nematodes within the genus *Anisakis* (i.e. *A. simplex* and *A. pegreffii*, also known as herring worms) are the causative agents of the fish-borne gastrointestinal illness ‘anisakiasis’. Since the 1960s, over 20,000 cases have been reported worldwide [5]; however, this number is most likely severely underestimated. Indeed, in Japan, where the consumption of raw or undercooked fish is common, ~2000–3000 new cases of anisakiasis occur each year [5].

The life cycle of *Anisakis* is indirect, with cetaceans such as dolphins and whales harbouring the dioecious adult nematodes in their gastrointestinal tract; female *Anisakis* release unembryonated eggs that are excreted in the aquatic environment *via* the faeces. Following development into first-, second- and third-stage larvae (L1s, L2s and L3s, respectively), these hatch from the eggs and are ingested by crustacean hosts. When infected crustaceans are ingested by suitable paratenic hosts, such as fish or squid, L3s penetrate the intestine and encapsulate in tissues, particularly those of the liver and the mesentery. Following ingestion of L3-containing paratenic hosts by a suitable cetacean host, L3s develop to fourth-stage larvae (L4s) and subsequently to adult males and females [5]. Humans are accidental hosts for *Anisakis*, with the

infection occurring *via* the ingestion of L3-containing raw or undercooked fish. In humans, the infection is usually self-limiting, and common symptoms range from epigastric pain, nausea, vomiting and low-grade fever (= gastric form), to intermittent or constant abdominal pain with possible complications such as peritonitis and/or ascites (= intestinal form). ‘Ectopic’ anisakiasis occurs when the ingested larvae penetrate the gut wall and undergo a somatic migration to other viscera [5]. However, individuals infected by *Anisakis* spp. can also become sensitised to parasite allergens, leading to the onset of allergic anisakiasis (the most significant form of disease), with symptoms ranging from urticaria, gastrointestinal and/or respiratory signs and/or anaphylaxis [5, 6]. Other than the ingestion of nematodes, allergic reactions can also be elicited by accidental exposure to hidden antigens in processed (including cooked) fish and fish products [7, 8], as well as inhalation of *Anisakis* allergens [9, 10].

To date, a total of 14 *A. simplex* allergens have been identified [see ref. 5]. Most of these allergens have been detected in the parasite excretory/secretory (ES) products, with Ani s 1, Ani s 5 and Ani s 7 being recognised by serum antibodies in the majority of individuals affected by allergic anisakiasis [11, 12]. A recent proteomic investigation of L3s of *A. simplex* led to the identification of 17 novel putative allergens, which included structural proteins (e.g. myosin-4), as well as a number of enzymes (e.g. one enolase and one endochitinase) [13]. These antigens, derived from both ES and somatic components (the latter released following death and disintegration of the larvae), act as triggers for the activation of complex immunological and cellular host defences, which often result in allergic sensitisation [14]. However, thus far, the exact number and nature of parasite molecules acting as potential allergens are currently unknown. Next-generation sequencing and bioinformatics now provide rapid and cost-effective opportunities to investigate the fundamental biology of parasites of medical significance [see 15], and to build curated molecular databases for in-depth analyses of specific sets of parasite genes and gene products [16, 17] of biological and/or medical relevance. For instance, the study of the transcriptome (= the complete set of mRNAs transcribed by a cell, tissue or organism at any one time) represents a powerful approach to identify and characterise thousands of parasite transcripts simultaneously, which can then be screened for one or more sequences of interest. In this study, we sequence and characterise the transcriptomes of *A. simplex* and *A. pegreffii* in order to build annotated datasets for fundamental studies of the biology of these parasites. Using these resources, we compile lists of putative parasite allergens based on comparative analyses with known allergen sequence data available in public databases.

## Materials and Methods

### Parasite material

L3s of *A. pegreffii* (AP) and *A. simplex* sensu stricto (AS) were collected between July and October 2013 in Tokyo, Japan. Whole fish (*Scomber japonicas*—chub mackerel; *Scomber australasicus*—blue mackerel and *Trachurus japonicus*—Japanese jack mackerel) were purchased from retail outlets and fish markets; the viscera were removed and inspected for anisakid L3 parasites and, when detected, these were collected and washed three times in saline solution. For each parasite, a segment of the caudal end was sectioned for molecular species identification [18], while the remaining portions were individually stored at -80°C for subsequent RNA extraction. For molecular identification, a region spanning the internal transcribed spacer 1 (ITS1), the 5.8S and the internal transcribed spacer 2 (ITS2) of the ribosomal DNA (rDNA) was amplified using the primer pairs 5'-GTCGAATTTCGTAGGTGAACCTGCGGAAGGATCA-3' and reverse: 3'-GCCGGATCCGAATCCTGGTTAGTTTCTTTTCCT-5' and the thermocycling protocol described by D'Amelio et al. [18]. PCR amplicons were digested using the restriction enzyme *Hinf*I for identification of the diagnostic molecular fingerprints for *A. simplex* s.s.

(~620, 250 and 100 bp) and *A. pegreffii* (~370, 200 and 250 bp), and *HhaI* for differentiation between *A. simplex* s.s. (~550 and 430 bp) and *A. berlandi* (formerly *A. simplex* C) (~550, 300 and 130 bp), respectively [18].

## RNA isolation and Illumina sequencing

RNA was extracted from three samples of each AS (100 L3s per sample) and AP (100 L3s per sample) using the Trizol reagent (Invitrogen, Life Technologies, Carlsbad, USA), and DNase-treated using Turbo DNA-free (Ambion, Austin, USA) according to the manufacturer's instructions. The amounts and integrity of total RNA were determined using a 2100 BioAnalyzer (Agilent Technologies, Santa Clara, California, USA). Polyadenylated (PolyA+) RNA was purified from 10 mg of total RNA from each AS and AP using Sera-Mag oligo(dT) beads, fragmented to a length of 100–500 nucleotides and reverse transcribed to cDNA using random hexamers. The size-fractionated cDNA was end-repaired and adaptor-ligated according to the manufacturer's protocol (Illumina). Ligated products of 200 bp were excised from agarose gels and PCR-amplified (15 cycles) [cf. 19]. Products were cleaned using a MinElute PCR purification kit (Qiagen, Hilden, Germany) and paired-end sequenced on an Illumina HiSeq 2000 [20] according to the manufacturer's protocol.

## Sequence trimming and assembly

Following removal of adapter sequences and sequences with suboptimal read quality (i.e., PHRED score of 32.0) using the filter\_fq script ([https://github.com/greatfireball/filter\\_fq](https://github.com/greatfireball/filter_fq)), the remaining 100-bp paired-read reads generated from the cDNA libraries from AS and AP were each assembled *de novo* using the program Trinity, which combines three independent software modules, i.e. Inchworm, Chrysalis, and Butterfly [21] (<http://trinityrnaseq.sourceforge.net/>). Briefly, a representative set of transcripts was assembled, including full-length transcripts of dominant isoforms and unique portions of alternatively spliced transcripts (Inchworm); next, portions of alternatively spliced transcripts and/or unique portions of paralogous genes were clustered and a de Bruijn graph was constructed for each cluster of transcripts (Chrysalis); finally, de Bruijn graphs were analysed simultaneously and full-length transcripts for alternatively spliced isoforms derived from paralogous genes were reported (Butterfly) [21]. In order to further reduce redundancy and generate comprehensive transcriptome datasets for each AS and AP, the resulting sequences, designated 'Unigenes', were compared across biological replicates of the same species, and 'Clusters' of Unigenes were generated based on sequence similarity (70% similarity cut-off).

## Annotation

The non-redundant, assembled datasets were then compared with transcriptomic and protein sequence data available for *Anisakis* spp. in the EST database of NCBI (<http://www.ncbi.nlm.nih.gov/nucest/?term=anisakis>) and in the WormBase ParaSite database ([http://parasite.wormbase.org/Anisakis\\_simplex\\_prjeb496/Info/Index/](http://parasite.wormbase.org/Anisakis_simplex_prjeb496/Info/Index/)) (e-value cut-off: 1e-05), and annotated using an established approach [cf. 19, 22]. Briefly, assembled Clusters and Unigenes (= contigs) were compared using the BLASTn and BLASTx algorithms to sequences available in WormBase ([www.wormbase.org](http://www.wormbase.org)), in the nucleotide sequence collection (Nt) of NCBI ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), and in the non-redundant (Nr) ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)), SwissProt (<http://expasy.org/>), Kyoto Encyclopedia of Genes and Genomes (KEGG; <http://www.genome.jp/kegg/>) and Clusters of Orthologous Groups (COG; <http://www.ncbi.nlm.nih.gov/COG/>) databases, respectively (e-value cut-off < 0.00001). Putative homologues in other representatives of Clade III nematodes (i.e. *A. suum* and *Toxocara canis*; [23, 24]), other nematodes and

organisms other than nematodes were also identified (e-value cut-off: 1e-05). Gene Ontology terms (GO, <http://www.geneontology.org/>) [25] were assigned to computationally translated AS and AP transcripts based on similarity to peptide sequences in the SwissProt database, according to the categories ‘Biological Process’, ‘Cellular Component’ and ‘Molecular Function’, using the Blast2GO software using default settings [26].

## Identification of putative allergens

In order to compile a list of putative allergens from AS and AP, predicted peptides from each of these species were compared with sequence data currently available in the AllergenOnline database (<http://www.allergenonline.com/>; January 2015 release) using the BLASTp algorithm (e-value cut-off: 1e-05; 70% identity match). At the time of the analyses (December 2015), this database contained ~1,900 peer-reviewed “protein sequence entries categorised into 744 taxonomic protein groups of unique proven or putative allergens (food, airway, venom/salivary and contact)” compiled from the GenBank, RefSeq and TPA nucleotide sequence repositories, as well as from the SwissProt, PIR, PRF and PDB protein sequence databases (<http://www.allergenonline.com/>). Computationally translated amino acid sequences from AS and AP were also compared, by BLASTp, with the sets of putative allergens described by Arcos et al. [27] and Fæste et al. [13] (e-value cut-off: 1e-05).

## Results

### The *Anisakis* transcriptomes

Paired-end Illumina sequencing of AS and AP cDNA libraries resulted in a total number of 64,065,430 and 65,508,456 raw reads, respectively (Table 1). Raw reads generated in the present study have been deposited in the Sequence Read Archive (SRA) database of NCBI (<http://www.ncbi.nlm.nih.gov/sra>) under study number SRP070744. Following pre-processing of raw reads, assembly and grouping of assembled Unigenes into sequence Clusters, a total of 34,746 (AS) and 33,747 (AP) transcripts were obtained. Conceptual translation of AS and AP transcripts resulted in 18,842 and 17,732 full-length predicted proteins, respectively (Table 1), of which 71.5% (AS) and 71.7% (AP) could be annotated *via* BLAST searches against the Nr, SwissProt, KEGG, COG and/or the GO databases (Table 1). Overall, 96.7%, 47.5% and 33.5% of AS transcripts and 88.0%, 90.5% and 73.8% of predicted proteins matched known nucleotide and amino acid sequences from *A. simplex* ([http://parasite.wormbase.org/Anisakis\\_simplex\\_prjeb496/Info/Index/](http://parasite.wormbase.org/Anisakis_simplex_prjeb496/Info/Index/)), as well as the phylogenetically related ascarid nematodes *A. suum* and *T. canis*, respectively; for AP, 96.6%, 46.1% and 33.2% of transcripts matched *A. simplex*, *A. suum* and *T. canis* sequences, respectively, while comparisons of amino acid sequence data resulted in 88.0%, 89.8% and 75.0% of AP predicted peptides, respectively, matching available peptide sequences from these two nematodes (Table 1).

A total of 5,561 (29.5%) and 5,357 (30.2%) AS and AP predicted proteins, respectively, could be assigned GO terms, while 8,945 (AS, 47.5%) and 8,629 (AP, 48.7%) matched homologous proteins in the KEGG database associated to 126 (AS) and 124 (AP) distinct biological pathways (Table 1). The 7,677 (AS, 40.7%) and 7,387 (AP, 41.7%) predicted proteins with matches in the COG database could be assigned to at least one of 25 functional categories, of which ‘general function prediction’ (AS = 15.7%; AP = 16.0%), ‘replication, recombination and repair’ (AS = 7.8%; AP = 7.7%), and ‘transcription’ (AS = 7.2%; AP = 7.4%), were the most represented (S1 Fig).

**Table 1. Transcriptome sequence data for third-stage larvae of *Anisakis simplex* and *Anisakis pegreffii* prior to and following assembly, and nucleotide and predicted peptide sequence data annotation (Nr = non-redundant; KEGG = Kyoto Encyclopedia of Genes and Genomes; COG = Clusters of Orthologous Groups of Proteins; GO = Gene Ontology).**

|   | <i>Anisakis simplex</i> | <i>Anisakis pegreffii</i> |
|---|-------------------------|---------------------------|
| Sequencing output   |                         |                           |
| No. of raw reads  | 64,065,430              | 65,508,456                |
| No. of clean reads  | 59,110,560              | 60,362,154                |
| Average length  | 100                     | 100                       |
| GC content (%)  | 43.40                   | 43.16                     |
| Transcript Assembly   |                         |                           |
| Unigenes <sup>a</sup>   |                         |                           |
| Total no.   | 62,967                  | 61,258                    |
| Mean Length ± SD (nt)   | 444 ± 692               | 465 ± 732                 |
| N50 <sup>b</sup>  | 1195                    | 1304                      |
| Contigs <sup>a</sup>  |                         |                           |
| No. of clusters   | 12,277                  | 11,394                    |
| No. of singletons (Unigenes)  | 22,469                  | 22,353                    |
| Total no.   | 34,746                  | 33,747                    |
| Mean Length ± SD (nt)   | 1173 ± 1105             | 1190 ± 1123               |
| N50 <sup>b</sup>  | 2027                    | 2082                      |
| Transcript coverage (%) <sup>c</sup>  |                         |                           |
| 100–90  | 32,902                  | 31,893                    |
| 90–80   | 1,620                   | 1,583                     |
| 80–70   | 189                     | 229                       |
| 70–60   | 27                      | 29                        |
| 60–50   | 2                       | 2                         |
| 50–40   | 6                       | 1                         |
| No. of contigs matching <i>Anisakis simplex</i> genes (WormBase ParaSite)               |                         |                           |
| <i>Ascaris suum</i> genes   | 16,500                  | 15,569                    |
| <i>Toxocara canis</i> genes   | 11,625                  | 11,222                    |
| Predicted peptides  |                         |                           |
| No. of predicted peptides by BLAST  | 12,914                  | 12,195                    |
| ESTScan   | 5,928                   | 5,537                     |
| Total   | 18,842                  | 17,732                    |
| No. of predicted peptides matching <i>Anisakis simplex</i> proteins (WormBase ParaSite) |                         |                           |
| <i>Ascaris suum</i> proteins  | 17,045                  | 15,949                    |
| <i>Toxocara canis</i> proteins  | 13,907                  | 13,298                    |
| Annotation  |                         |                           |
| No. of predicted peptides with matches in Nr  |                         |                           |
| Swissprot   | 7,747                   | 7,563                     |
| KEGG (no. of biological pathways)   | 8,945 (126)             | 8,629 (124)               |
| COG   | 7,677                   | 7,387                     |
| GO  | 5,561                   | 5,357                     |
| Total no. of annotated predicted peptides   | 13,472                  | 12,705                    |

<sup>a</sup>Assembled transcripts prior to (= Unigenes) and following sequence clustering (= Contigs) (see [Materials and Methods](#))

<sup>b</sup>Length N (in bp) for which 50% of all nucleotides sequenced are assembled in Contigs of length ≤ N

<sup>c</sup>Percentage of a transcript covered by reads. The value is calculated as a ratio of the number of bases in a transcript covered by unique mapping reads to the number of total bases in that transcript.

doi:10.1371/journal.pntd.0004845.t001

## Putative novel *Anisakis* allergens

Comparative analyses of AS and AP predicted peptides with sequence data available in the AllergenOnline Database (<http://www.allergenonline.com/>) resulted in a total number of 38 (AS) and 31 (AP) matches, respectively, including 22 (AS) and 18 (AP) sequences matching previously known *Anisakis* allergens (S1 Table). Of the 16 (AS) and 13 (AP) remaining predicted peptides, 10 (AS) and 8 (AP) also matched protein sequences described as putative novel allergens by Arcos et al. [28] and Faeste et al. [13] (Table 2). BLAST comparisons between sets of putative allergens identified in AS and AP revealed six sequences unique to AS, which included a heat shock 70 kDa protein (ID: Unigene9825\_AS1A), two fructose-bisphosphate aldolases A, a 60S ribosomal protein and two putative allergens matching sequences previously characterised in *A. suum* and *Ascaris lumbricoides* (Table 2); one sequence encoding a heat shock 70 kDa protein (ID: Unigene16173\_AP1A) was unique to AP (Table 2). All of the putative novel allergens identified matched homologous allergenic proteins in arthropods (AS = 7; AP = 9), fungi (AS = 6; AP = 6), plants (AS = 3; AP = 3), other helminths (AS = 3; AP = 1) and fish (AS = 3; AP = 1) (Table 2). Of the putative allergens characterised in the present study, cyclophilins (AS and AP) and two predicted proteins of unknown function (AS) were identified in *Anisakis* for the first time (Table 2). Summaries of GO annotation and KEGG pathway analysis information for the whole transcriptomes of AS and AP are shown in S1 Fig, whilst the complete lists of AS and AP assembled transcripts, together with corresponding predicted peptides and annotation information, are given in S2 Table.

## Discussion

In this study, we characterise the transcriptomes of two *Anisakis* species, and provide the scientific community with a resource to explore the biology of these parasites, as well as their allergenic properties, using proteomics and immunological tools. The caudal extremities of the larvae used in this study were removed in order to unequivocally confirm species identification by PCR-coupled RFLP of the of ITS1-5.8S-ITS2 region of the rDNA, thus potentially leading to biases in the sets of *Anisakis* transcripts identified and characterised. Nevertheless, we consider these datasets to represent a comprehensive snapshot of the complements of genes transcribed by the L3s of these parasites. Indeed, prior to this study, only 913 *Anisakis* transcripts were present in the EST database of NCBI (<http://www.ncbi.nlm.nih.gov/nucest/?term=anisakis>); however, a draft genome sequence for this parasite, accompanied by large-scale transcriptomic sequence data to support gene predictions, are currently available from the WormBase ParaSite database as part of the '50 Helminth Genome Initiative' (<http://www.sanger.ac.uk/science/collaboration/50hgp>). Though, given that these resources are as yet unpublished, we opted to assemble the short reads generated in this study *de novo*, and compare the resulting full-length cDNAs (contigs) and, subsequently, the amino acid sequences predicted from these, to *Anisakis* transcripts and predicted proteins (respectively) on the WormBase ParaSite database. Overall, 96.7% and 96.6% of the transcripts assembled in the present study matched *Anisakis* nucleotide sequence data in the latter database, thus providing support to the reliability and robustness of our assembly (cf. Table 1). Conversely, BLAST comparisons of *Anisakis* nucleotide and predicted amino acid sequence data generated in this study with available transcripts and protein sets from other ascarids [23, 24] revealed higher sequence similarities between *Anisakis* and *A. suum* than between the former and *T. canis* (see Table 1). Currently, the relationships amongst Clade III nematodes (also referred to as suborder Spirurina; [28]) are defined according to phylogenetic analyses of the small subunit of the ribosomal RNA (SSU rRNA) [29, 30]. In one of these investigations, SSU rRNA sequences from *Ascaris* spp. and *Toxocara* group together to the exclusion of the *Anisakis* spp. counterparts [30], likely reflecting the

**Table 2. Putative novel allergens identified in the transcriptomes from third-stage larvae of *Anisakis simplex* and *Anisakis pegreffii*, based on homology of predicted amino acid sequences with known allergens in the AllergenOnline database (<http://www.allergenonline.com/about.shtml>). (e-value cut-off: <1e-05, identity cut-off: >70%). (Nr = non-redundant database; GO = Gene Ontology; KEGG = Kyoto Encyclopedia of Genes and Genomes; BP = Biological Process; CC = Cellular Component; MF = Molecular Function).**

| Unigene ID               | Size (bp) | Uniquely mapped reads | Closest match in Nr   | GO annotation   | KEGG pathway annotation   | Closest match in AllergenOnline database | IUIS nomenclature <sup>a</sup> | Taxonomy <sup>b</sup> |
|--------------------------|-----------|-----------------------|---|---|---|--|--------------------------------|-----------------------|
| <i>Anisakis simplex</i>  |           |                       |   |   |   |  |                                |                       |
| CL2584.<br>Contig1_AS1A* | 644       | 335597                | Heat shock 70 kDa protein [ <i>Anisakis pegreffii</i> ]       | Cell, Cell part, Membrane (CC)  | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci, Dfa, Che</i>  |
| Unigene14515_AS1A*       | 438       | 11166                 | Heat shock 70kDa protein 8 [ <i>Danio rerio</i> ]             | Response to stimulus (BP)   | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci, Dfa, Che</i>  |
| Unigene14555_AS1A*       | 213       | 75345                 | Heat shock cognate 71 kDa protein [ <i>Danio rerio</i> ]      | Response to stimulus (BP)   | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci</i>            |
| Unigene7175_AS1A*        | 276       | 21823                 | Heat shock protein 70 b2 [ <i>Ascaris suum</i> ]              | Cell, Cell part, Membrane (CC)  | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci, Dfa, Che</i>  |
| Unigene9825_AS1A*        | 258       | 16179                 | Heat shock 70 kDa protein c [ <i>Ascaris suum</i> ]           | Cellular process, Growth, Multi-organism process, Reproduction, Response to stimulus, Reproductive process, Single-organism process (BP); Cell, Cell part, Membrane, Membrane-enclosed lumen, Organelle part (CC) | Protein processing in endoplasmic reticulum, Protein export           | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci, Dfa, Che</i>  |
| Unigene10469_AS1A        | 171       | 48012                 | Peptidyl-prolyl cis-trans isomerase 3 [ <i>Ascaris suum</i> ] | Cellular component organization or biogenesis, Growth, Rhythmic process, Signaling (BP); Extracellular region, Macromolecular complex, Membrane, Organelle (CC)   | None  | Cyclophilin                              | Unassigned                     | <i>Dca, Bpe, Dfa</i>  |
| Unigene11979_AS1A        | 159       | 1631                  | Peptidyl-prolyl cis-trans isomerase e [ <i>Ascaris suum</i> ] | None  | None  | Cyclophilin                              | Unassigned                     | <i>Dfa</i>            |

(Continued)

**Table 2.** (Continued)

| Unigene ID         | Size (bp) | Uniquely mapped reads | Closest match in Nr   | GO annotation  | KEGG pathway annotation   | Closest match in AllergenOnline database | IUIS nomenclature <sup>a</sup> | Taxonomy <sup>b</sup>               |
|--------------------|-----------|-----------------------|---|--|---|--|--------------------------------|-------------------------------------|
| Unigene15335_AS1A  | 134       | 50809                 | Ppia protein, partial [ <i>Danio rerio</i> ]                    | Membrane (CC)  | None  | Cyclophilin                              | Unassigned                     | <i>Dfa, Cro, Bpe, Dca, Msy, Afu</i> |
| Unigene6668_AS1A   | 172       | 18480                 | Peptidyl-prolyl cis-trans isomerase 3 [ <i>Toxocara canis</i> ] | Extracellular region, Membrane (CC)  | None  | Cyclophilin                              | Unassigned                     | <i>Dca, Bpe, Dfa, Cro</i>           |
| Unigene2527_AS1A** | 347       | 61681                 | Fructose-bisphosphate aldolase 2 [ <i>Ascaris suum</i> ]        | Metabolic process (BP); Organelle part (CC)  | Metabolic pathways, Biosynthesis of secondary metabolites, Glycolysis/ Gluconeogenesis, Fructose and mannose metabolism, Pentose phosphate pathway, Carbon fixation in photosynthetic organisms | Fructose-bisphosphate aldolase A         | Thu a 3.0101                   | <i>Tal</i>                          |
| Unigene8467_AS1A** | 351       | 24201                 | Fructose-bisphosphate aldolase 1 [ <i>Ascaris suum</i> ]        | Cellular process, Metabolic process (BP)   | Metabolic pathways, Biosynthesis of secondary metabolites, Glycolysis/ Gluconeogenesis, Fructose and mannose metabolism, Pentose phosphate pathway, Carbon fixation in photosynthetic organisms | Fructose-bisphosphate aldolase A         | Thu a 3.0101                   | <i>Tal</i>                          |
| Unigene17381_AS1A* | 68        | 174503                | Ribosomal protein L3 [ <i>Danio rerio</i> ]                     | Cell, Cell part, Macromolecular complex, Membrane, Membrane-enclosed lumen, Organelle, Organelle part (CC); Structural molecule activity (MF)        | Ribosome  | 60S ribosomal protein L3                 | Unassigned                     | <i>Afu</i>                          |
| Unigene10513_AS1A* | 431       | 133980                | Enolase [ <i>Anisakis simplex</i> ]                             | Cellular component organization or biogenesis, Cellular process, Metabolic process (BP); Extracellular region, Macromolecular complex, Membrane (CC) | Metabolic pathways, Biosynthesis of secondary metabolites, RNA degradation, Glycolysis/ Gluconeogenesis   | Enolase 3–2                              | Sal s 2.0101                   | <i>Ssa</i>                          |

(Continued)

Table 2. (Continued)

| Unigene ID                | Size (bp) | Uniquely mapped reads | Closest match in Nr   | GO annotation                             | KEGG pathway annotation  | Closest match in AllergenOnline database | IUIS nomenclature <sup>a</sup> | Taxonomy <sup>b</sup>     |
|---------------------------|-----------|-----------------------|---|---|--|--|--------------------------------|---------------------------|
| Unigene7252_AS1A**        | 203       | 15510                 | Sigma class glutathione S-transferase [ <i>Baylisascaris schroederi</i> ] | None                                      | Metabolic pathways, mRNA surveillance pathway, Glutathione metabolism, Arachidonic acid metabolism | Glutathione S-transferase 1              | Asc s 13.0101                  | <i>Asu</i>                |
| CL1712. Contig1_AS1A      | 294       | 8338                  | Allergen, partial [ <i>Ascaris suum</i> ]                                 | None                                      | None   | ABA-1 allergen, partial                  | Unassigned                     | <i>Alu, Asu</i>           |
| CL1712. Contig2_AS1A      | 160       | 2996                  | Polyprotein allergen/ antigen, partial [ <i>Ascaris suum</i> ]            | None                                      | None   | ABA-1 allergen, partial                  | Unassigned                     | <i>Alu, Asu</i>           |
| <i>Anisakis pegreffii</i> |           |                       |   |   |  |  |                                |                           |
| Unigene13566_AP1A*        | 176       | 94213                 | Heat shock cognate 71 kDa protein [ <i>Danio rerio</i> ]                  | Response to stimulus (BP)                 | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis                              | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci</i>                |
| Unigene13814_AP1A*        | 438       | 2533                  | Heat shock 70kDa protein 8 [ <i>Danio rerio</i> ]                         | Response to stimulus (BP)                 | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis                              | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci, Dfa, Che</i>      |
| Unigene16173_AP1A*        | 88        | 22129                 | Inducible heat shock protein 70 [ <i>Tigriopus californicus</i> ]         | Response to stimulus (BP); Organelle (CC) | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis, Protein export              | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci, Dfa, Che</i>      |
| Unigene16415_AP1A*        | 138       | 38735                 | Heat shock protein 70 b2 [ <i>Ascaris suum</i> ]                          | Cell, Cell part, Membrane (CC)            | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis                              | Heat shock 70 kDa protein                | Unassigned                     | <i>Dfa, Che</i>           |
| CL872. Contig1_AP1A*      | 644       | 367002                | Heat shock protein 70 [ <i>Anisakis pegreffii</i> ]                       | Cell, Cell part, Membrane (CC)            | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis                              | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci, Dfa, Che</i>      |
| Unigene21614_AP1A*        | 81        | 9769                  | Heat shock protein 70, partial [ <i>Neobathyscia mancinii</i> ]           | Response to stimulus (BP)                 | Spliceosome, Protein processing in endoplasmic reticulum, Endocytosis                              | Heat shock 70 kDa protein                | Unassigned                     | <i>Pci, Dfa, Che</i>      |
| CL1265. Contig1_AP1A      | 171       | 64890                 | Peptidyl-prolyl cis-trans isomerase 3 [ <i>Ascaris suum</i> ]             | Extracellular region, Membrane (CC)       | None   | Cyclophilin                              | Unassigned                     | <i>Dca, Bpe, Dfa, Cro</i> |
| CL1927. Contig1_AP1A      | 159       | 1062                  | Peptidyl-prolyl cis-trans isomerase e [ <i>Ascaris suum</i> ]             | None                                      | None   | Cyclophilin                              | Unassigned                     | <i>Dfa</i>                |

(Continued)

Table 2. (Continued)

| Unigene ID              | Size (bp) | Uniquely mapped reads | Closest match in Nr   | GO annotation   | KEGG pathway annotation   | Closest match in AllergenOnline database | IUIS nomenclature <sup>a</sup> | Taxonomy <sup>b</sup>     |
|-------------------------|-----------|-----------------------|---|---|---|--|--------------------------------|---------------------------|
| CL1927.<br>Contig2_AP1A | 159       | 3029                  | Peptidyl-prolyl cis-trans isomerase e [ <i>Ascaris suum</i> ]             | None  | None  | Cyclophilin                              | Unassigned                     | <i>Dfa</i>                |
| Unigene16460_AP1A       | 145       | 3867                  | Ppia protein, partial [ <i>Danio rerio</i> ]                              | Membrane (CC)   | None  | Cyclophilin                              | Unassigned                     | <i>Dca, Bpe, Dfa, Cro</i> |
| Unigene10435_AP1A       | 172       | 28438                 | Peptidyl-prolyl cis-trans isomerase 3 [ <i>Toxocara canis</i> ]           | Extracellular region, Membrane (CC)   | None  | Cyclophilin                              | Unassigned                     | <i>Dca, Bpe, Dfa, Cro</i> |
| Unigene2939_AP1A*       | 431       | 93319                 | Enolase [ <i>Anisakis simplex</i> ]                                       | Cellular process, Metabolic process (BP); Extracellular region, Macromolecular complex, Membrane (CC) | Metabolic pathways, Biosynthesis of secondary metabolites, RNA degradation, Glycolysis/ Gluconeogenesis | Enolase 3–2                              | Sal s 2.0101                   | <i>Ssa</i>                |
| Unigene6873_AP1A**      | 203       | 21192                 | Sigma class glutathione S-transferase [ <i>Baylisascaris schroederi</i> ] | None  | Metabolic pathways, mRNA surveillance pathway, Glutathione metabolism, Arachidonic acid metabolism      | Glutathione S-transferase 1              | Asc s 13.0101                  | <i>Asu</i>                |

<sup>a</sup>Systematic allergen nomenclature approved by the World Health Organization and International Union of Immunological Societies (WHO/IUIS) Allergen Nomenclature Sub-committee

<sup>b</sup>Abbreviations of homologous species of matched allergens (alphabetical): *Ascaris lumbricoides* (*Alu*), *Ascaris suum* (*Asu*), *Aspergillus fumigatus* (*Afu*), *Betula pendula* (*Bpe*), *Blattella germanica* (*Bge*), *Catharanthus roseus* (*Cro*), *Cladosporium herbarum* (*Che*), *Daucus carota* (*Dca*), *Dermatophagoides farina* (*Dfa*), *Malassezia sympodialis* (*Msy*), *Penicillium citrinum* (*Pci*), *Salmo salar* (*Ssa*), *Thunnus albacares* (*Tal*), *Tyrophagus putrescentiae* (*Tpu*); Taxonomy: Nematodes: *Alu, Asu*; Fungi: *Afu, Che, Msy, Pci*; Plant: *Bpe, Cro, Dca*; Arthropods: *Bge, Dfa, Tpu*; Fish: *Ssa, Tal*.

\*Also identified by Arcos et al. [28] and Fæste et al. [13].

\*\*Also identified by Fæste et al. [13].

doi:10.1371/journal.pntd.0004845.t002

similarities in morphology and fundamental biology between the former (“terrestrial”) species. While the elucidation of the systematic and phylogenetic relationships between *Anisakis*, *Ascaris* and *Toxocara* species are beyond the scope of the present work, the availability of large-scale transcriptomic and genomic datasets for all of these parasites (<http://www.sanger.ac.uk/science/collaboration/50hgp>; [23, 24]) represents a useful resource for future, comprehensive phylogenomic studies of ascarid nematodes.

In this study, we aimed to sequence and annotate the transcriptomes of two *Anisakis* species, in order to build a molecular resource from which to draw information on the set(s) of molecules responsible for evoking allergic responses in the human host. Two recently published key studies [13, 27] utilised sera from human patients with IgE against *A. simplex* and positive to the skin prick test, coupled with mass spectrometry-based analyses of reactive *Anisakis* proteins to identify and characterise putative novel nematode allergens. Both studies revealed a broad array of potentially allergenic *Anisakis* molecules, with individual sera

binding to multiple protein bands, which resulted in a substantial inter-individual variability of binding patterns [13]. This finding suggests that the whole complement of *Anisakis* allergens is yet to be fully defined, and supports the application of NGS technologies and bioinformatics to assist in this quest. While all of the putative novel allergens identified by Arcos et al. [27] and Fæste et al. [13] matched AS and AP amino acid sequences inferred from cDNAs generated in this study, only subsets of these also matched known allergens in the AllergenOnline database (cf. Table 2). The most likely explanation for this observation is technical and is related to the inevitable incompleteness of the AllergenOnline database which, while regularly updated and manually curated, relies on the collection of sequence data submitted by end-users to public sequence repositories and designated as ‘allerg\*’, or extracted directly from peer-reviewed publications ([www.allergenonline.com](http://www.allergenonline.com)). On the other hand, the absence of some previously identified putative *Anisakis* allergens from the AllergenOnline database is justified by the fact that the allergenic properties of these molecules are yet to be fully elucidated [13, 27].

Amongst the putative allergens characterised in the present study, sequences encoding heat shock proteins 70 (HSPs 70) and enolases were identified in both AS and AP datasets, and had been previously detected using immune-proteomic approaches [13, 27]. Based on direct comparisons with sequence data in the Nr database of NCBI, these (and other) sequences displayed high sequence similarity with HSPs from *Danio rerio* (cf. Table 2). While this finding is likely to be linked to the overrepresentation of sequences from zebrafish in Nr compared with sequence data from parasitic nematodes, contamination of *Anisakis* mRNA with that from the fish host cannot be excluded. HSPs are a family of highly conserved proteins that play primary roles in maintaining cellular homeostasis but whose increased expression in the presence of conditions of stress such as sudden changes in temperature, injuries and infections, is responsible for the activation of a cascade of immune-molecular events that culminate in inflammatory responses [31]. In particular, in a recent study [32], the expression of an HSP 70 from *A. pegreffii* was increased in L4s compared with L3s, which may be linked to a response of the parasite to the heat stress that immediately follows infection [32]. HSPs 70 from mites, black flies, midges and cockroaches are known allergens and mediators of allergic contact hypersensitivity [33, 34]; in addition, levels of HSPs 70 are increased in the sputum and plasma of asthmatic patients compared with healthy controls [35], while antibodies against these proteins are associated with a number of immunological disorders, such as allergy to metals [36]. Based on this knowledge, it is therefore plausible that antibodies against *Anisakis* HSPs 70, whose expression is increased upon host infection [32], may play a key role in allergic responses to these parasites. Enolases are also recognised as major allergenic proteins in plants, fish, fungi, cockroaches and biting insects [37]. In a previous study, anti-enolase antibodies were detected in sera from mice experimentally infected with *A. simplex* L3s or exposed to parasite crude protein extracts, but not from mouse sera raised against the parasite excretory-secretory antigens [38]; in the same study, anti-enolase antibodies could not be detected in the sera of human patients infected with *A. simplex*, which led the authors to speculate that these molecules do not offer a sufficient antigenic stimulus to act as allergens [38]. Conversely, IgE against *Anisakis* enolases were detected in *Anisakis*-allergic patients by Fæste et al. [13], thus supporting the role of these molecules in allergic anisakiasis. However, it is worthwhile to note that, while sensitisation to *Anisakis* allergens usually follows the ingestion of fish and/or fish products, these products can also contain very similar allergens to those from the parasites ([e.g. 39]; cf. Table 2). Importantly, recent studies [39] have demonstrated the role of fish beta-enolase and fructose-bisphosphate aldolase as allergenic stimuli in patients sensitised to cod, salmon and tuna, thus further complicating the diagnosis of ‘true’ sensitisation to *Anisakis* allergens. In addition, tropomyosin (Ani s 3), one of the most immunogenic proteins known to man, was

also first discovered in *Anisakis* and subsequently identified in over 150 invertebrate species (including shellfish and mites) [see ref. 1]. This muscle protein, which contains a coiled-coiled  $\alpha$ -helical structure, is responsible for exacerbated immune responses in over 30% of the world's population, and may be involved in the known cases of cross-allergenicity between mites and *Anisakis* [40].

Of the molecules inferred as novel putative allergens, sequences encoding peptidyl-prolyl cis-trans isomerases (cyclophilins) were identified in both AS and AP (cf. Table 2). Cyclophilins belong to a family of conserved proteins present in both prokaryotes and eukaryotes, and thought to play key roles in a range of human inflammatory diseases, including rheumatoid arthritis and asthma [41]. Human cyclophilins also act as self-antigens, being recognised by serum IgE from individuals sensitised to environmental cyclophilins, such as those in pollens [see 42]. Cyclophilins have also been identified in a range of parasitic nematode species (e.g. *Angiostrongylus cantonensis*, *Dirofilaria immitis* and *Haemonchus contortus*; [43–45]) and are thought to operate as catalysts and chaperones in cuticle synthesis [46]; in particular, cyclophilins were identified in the excretory-secretory products of *H. contortus* and recognised by sheep hyper-immune sera [47]. Interestingly, cyclophilins were also detected by immunological screening of a cDNA library from the zoonotic cestode *Echinococcus granulosus*, causing cystic echinococcosis (CE), with sera from infected human subjects that had displayed allergic (skin) reactions [48]; sera from these patients did not recognise the homologous human cyclophilins, nor that from the yeast *Malassezia furfur*, thus supporting the hypothesis that the parasite cyclophilin was responsible for the allergic reactions observed in patients with CE [48]. Based on this information, we hypothesise a role of cyclophilins in the array of molecules responsible for allergic anisakiasis, a hypothesis that requires testing.

Amongst the novel putative allergens in AS were two sequences with high sequence similarity to ABA-1 proteins, members of the nematode polyprotein allergens (NPAs) from *A. suum* (cf. Table 2). These proteins are synthesised as repetitive polyproteins, which are subsequently cleaved during post-translational processing into multiple functional units with fatty-acid binding properties [49]. NPAs from *Ascaris* and other parasitic nematodes (e.g. *Brugia malayi*) are associated with hypersensitivity responses in infected individuals [see 50]. In addition, no IgE cross-reactivity could be detected between human sera (from asthmatic patients from an area where ascariasis is endemic and exposure to mites is common) probed with recombinant *Ascaris* ABA-1 and mite fatty-acid binding proteins (FABPs), thus suggesting that parasites are solely responsible for allergic reactions against this protein [51]. Given the high level of sequence similarity (at both nucleotide and protein levels) between *Anisakis* sequences encoding for ABA-1 proteins identified in this study and the *Ascaris* counterparts, it is plausible that these molecules also contribute to the onset of allergic anisakiasis.

While our work represents a step forward in the application of NGS technologies towards building a molecular infrastructure for research on allergic anisakiasis, gaps still exist in our knowledge of the complex host-parasite relationships which culminate in the exacerbated immune reactions observed in individuals infected by *Anisakis*. The completion, curation and refinement of the whole genome sequence of *A. simplex* (<http://www.sanger.ac.uk/science/collaboration/50hgp>) will assist filling these gaps, by providing a solid resource for fundamental functional explorations of these relationships. Ultimately, the discovery of the whole array of parasite molecules responsible for immune hypersensitivity in *Anisakis* (and other parasites) will set the basis of future studies aimed at developing comprehensive, reliable and robust diagnostic tools, which will assist clinicians in choosing appropriate intervention strategies and effectively assessing their outcomes.

## Supporting Information

**S1 Fig. Functional annotation of the *Anisakis* transcriptomes.** Summary of functional annotation information linked to predicted peptides inferred from the transcriptomes of *Anisakis simplex* and *Anisakis pegreffii* third stage larvae as inferred *via* comparisons with sequence data available in the Kyoto Encyclopedia of Genes and Genomes (KEGG, Level 1; expressed as percentage of Unigenes mapping to conserved biological pathways) (A), Gene Ontology (GO; Level 2), according to the categories ‘Biological Process’, ‘Cellular Component’ and ‘Molecular Function’ (B) and Clusters of Orthologous Groups of Proteins (COG) (C) databases. (D) Venn diagram illustrating the number of assembled transcripts shared between *A. simplex* and *A. pegreffii*, and of transcripts unique to each species (e-value cut-off: 1e-15). (PDF)

**S1 Table. Predicted peptides with homology to previously known *Anisakis* allergens.** Predicted peptides inferred from the transcriptomes of third stage larvae of *Anisakis simplex* and *Anisakis pegreffii* with homology to previously known *Anisakis* allergens (e-value cut-off: <1e-5, identity cut-off: >70%) available in the AllergenOnline database (<http://www.allergenonline.com/about.shtml>) (Nr = non-redundant database). (DOCX)

**S2 Table. Lists of assembled transcripts from *Anisakis simplex* and *Anisakis pegreffii*.** Complete lists of assembled transcripts and predicted peptides inferred from the transcriptomes of third stage larvae of *A. simplex* and *A. pegreffii*, and corresponding nucleotide and predicted amino acid sequences and functional annotation. (XLSX)

## Author Contributions

Conceived and designed the experiments: FJB ALL CC. Performed the experiments: FJB XS HS. Analyzed the data: FJB XS IA MJN CC. Contributed reagents/materials/analysis tools: FJB ALL CC. Wrote the paper: FJB ALL CC DO.

## References

1. Lopata AL, Lehrer SB. (2009) New insights into seafood allergy. *Curr Opin Allergy Clin Immunol* 9: 270–277. doi: [10.1097/ACI.0b013e32832b3e6f](https://doi.org/10.1097/ACI.0b013e32832b3e6f) PMID: [19398906](https://pubmed.ncbi.nlm.nih.gov/19398906/)
2. World Health Organization. (2014) Advancing food safety initiatives: strategic plan for food safety including foodborne zoonoses 2013–2022. Geneva, Switzerland, World Health Organization: 32.
3. Fukuda K. (2015) Food safety in a globalized world. *Bull World Health Organ* 93: 212. doi: [10.2471/BLT.15.154831](https://doi.org/10.2471/BLT.15.154831) PMID: [26229182](https://pubmed.ncbi.nlm.nih.gov/26229182/)
4. Abelson P, Potter-Forbes M, Hall GV. (2006) The Annual Cost of Foodborne Illness in Australia. Australian Government Department of Health & Ageing, 2006.
5. Baird FJ, Gasser RB, Jabbar A, Lopata AL. (2014) Foodborne anisakiasis and allergy. *Mol Cell Probes* 28, 167–174. doi: [10.1016/j.mcp.2014.02.003](https://doi.org/10.1016/j.mcp.2014.02.003) PMID: [24583228](https://pubmed.ncbi.nlm.nih.gov/24583228/)
6. López-Serrano MC, Gomez AA, Daschner A, Moreno-Ancillo A, Parga D, Suarez JM, et al. (2000) Gastroallergic anisakiasis: findings in 22 patients. *J Gastroenterol Hepatol* 15, 503–506. PMID: [10847436](https://pubmed.ncbi.nlm.nih.gov/10847436/)
7. Moneo I, Caballero ML, González-Muñoz M, Rodríguez-Mahillo AI, Rodríguez-Perez R, Silva A. (2005) Isolation of a heat-resistant allergen from the fish parasite *Anisakis simplex*. *Parasitol Res* 96, 285–289 PMID: [15895253](https://pubmed.ncbi.nlm.nih.gov/15895253/)
8. Anibarro B, Seoane FJ, Mugica MV. (2007) Involvement of hidden allergens in food allergic reactions. *J Investig Allergol Clin Immunol* 17, 168–172. PMID: [17583104](https://pubmed.ncbi.nlm.nih.gov/17583104/)

9. Nieuwenhuizen N, Lopata AL, Jeebhay MF, De'Broski RH, Robins T, Brombacher F. (2006) Exposure to the fish parasite *Anisakis* causes allergic airway hyperreactivity and dermatitis. *J Allergy Clin Immunol* 117, 1098–1105. PMID: [16675338](#)
10. Kirstein F, Horsnell WGC, Nieuwenhuizen N, Ryffel B, Lopata AL, Brombacher F. (2010) *Anisakis*-induced airway hyperresponsiveness is mediated by IFN- $\mu$  in the absence of IL-4R- $\alpha$  responsiveness. *Infect Immun* 78, 4077–4086. doi: [10.1128/IAI.01131-09](#) PMID: [20605987](#)
11. Kobayashi Y, Shimakura K, Ishizaki S, Nagashima Y, Shiomi K. (2007) Purification and cDNA cloning of a new heat-stable allergen from *Anisakis simplex*. *Mol Biochem Parasitol* 155, 138–145. PMID: [17689675](#)
12. Anadón AM, Romarís F, Escalante M, Rodríguez E, Gárate T, Cuéllar C, et al. (2009) The *Anisakis simplex* Ani s 7 major allergen as an indicator of true *Anisakis* infections. *Clin Exp Immunol* 156, 471–478. doi: [10.1111/j.1365-2249.2009.03919.x](#) PMID: [19438600](#)
13. Fæste CK, Jonscher KR, Dooper MM, Egge-Jacobsen W, Moen A, Daschner A, et al. (2014) Characterisation of potential novel allergens in the fish parasite *Anisakis simplex*. *EuPA Open Proteom* 4, 140–155. PMID: [27110489](#)
14. Nieuwenhuizen NE, Lopata AL. (2013) *Anisakis*—A food-borne parasite that triggers allergic host defences. *Int J Parasitol* 43, 1047–1057. doi: [10.1016/j.ijpara.2013.08.001](#) PMID: [23994604](#)
15. Cantacessi C, Campbell BE, Gasser RB. (2012) Key stronglylid nematodes of animals—Impact of next-generation transcriptomics on systems biology and biotechnology. *Biotechnol Adv* 30, 469–488. doi: [10.1016/j.biotechadv.2011.08.016](#) PMID: [21889976](#)
16. Cantacessi C, Hofmann A, Pickering D, Navarro S, Mitreva M, Loukas A. (2013) TIMPs of parasitic helminths—a large-scale analysis of high-throughput sequence datasets. *Parasit Vectors* 6, 1.
17. Cantacessi C, Prasopdee S, Sotillo J, Mulvenna J, Tesana S, Loukas A. (2013) Coming out of the shell: building the molecular infrastructure for research on parasite-harboring snails. *PLoS Negl Trop Dis* 7, e2284. doi: [10.1371/journal.pntd.0002284](#) PMID: [24069465](#)
18. D'Amelio S, Mathiopoulos KD, Santos CP, Pugachev ON, Webb SC, Picanco M, Paggi L. (2000) Genetic markers in ribosomal DNA for the identification of members of the genus *Anisakis* (Nematoda: Ascaridoidea) defined by polymerase-chain-reaction-based restriction fragment length polymorphism. *Int J Parasitol* 30, 223–226. PMID: [10704605](#)
19. Prasopdee S, Sotillo J, Tesana S, Laha T, Kulsantiwong J, Nolan MJ, Loukas A, et al. (2014) RNA-Seq reveals infection-induced gene expression changes in the snail intermediate host of the carcinogenic liver fluke, *Opisthorchis viverrini*. *PLoS Negl Trop Dis* 8, e2765. doi: [10.1371/journal.pntd.0002765](#) PMID: [24676090](#)
20. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–59. doi: [10.1038/nature07517](#) PMID: [18987734](#)
21. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29, 644–652. doi: [10.1038/nbt.1883](#) PMID: [21572440](#)
22. Cantacessi C, Jex AR, Hall RS, Young ND, Campbell BE, Joachim A, et al. (2010) A practical, bioinformatic workflow system for large data sets generated by next-generation sequencing. *Nucleic Acids Res* 38, e171. doi: [10.1093/nar/gkq667](#) PMID: [20682560](#)
23. Jex AR, Liu S, Li B, Young ND, Hall RS, Li Y, et al. (2011) *Ascaris suum* draft genome. *Nature* 479, 529–533. doi: [10.1038/nature10553](#) PMID: [22031327](#)
24. Zhu XQ, Korhonen PK, Cai H, Young ND, Nejsum P, von Samson-Himmelstjerna G, et al. (2015) Genetic blueprint of the zoonotic pathogen *Toxocara canis*. *Nat Commun* 6, 6145. doi: [10.1038/ncomms7145](#) PMID: [25649139](#)
25. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. (2000) Gene Ontology: tool for the unification of biology: The Gene Ontology Consortium. *Nat Genet* 25, 25–29. PMID: [10802651](#)
26. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36, 3420–3435. doi: [10.1093/nar/gkn176](#) PMID: [18445632](#)
27. Arcos SC, Ciordia S, Roberston L, Zapico I, Jiménez-Ruiz Y, Gonzalez-Muñoz M, et al. (2014) Proteomic profiling and characterization of differential allergens in the nematodes *Anisakis simplex* sensu stricto and *A. pegreffii*. *Proteomics* 14, 1547–1568. doi: [10.1002/pmic.201300529](#) PMID: [24723494](#)
28. De Ley P, Blaxter M. (2002) Systematic position and phylogeny. In: Lee DL, editor. *The Biology of Nematodes*. Taylor and Francis, London, pp. 1–30.
29. Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, Vierstraete A, et al. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* 392, 71–75. PMID: [9510248](#)

30. Nadler SA, Carreno RA, Mejía-Madrid H, Ullberg J, Pagan C, Houston R, et al. (2007) Molecular phylogeny of clade III nematodes reveals multiple origins of tissue parasitism. *Parasitology* 134, 1421–1422. PMID: [17506928](#)
31. Srivastava P. (2002) Roles of heat-shock proteins in innate and adaptive immunity. *Nat Rev Immunol* 2, 185–194. PMID: [11913069](#)
32. Chen HY, Cheng YS, Shih HH. (2015) Expression patterns and structural modelling of Hsp70 and Hsp90 in a fish-borne zoonotic nematode *Anisakis pegreffii*. *Vet Parasitol* 212, 281–291. doi: [10.1016/j.vetpar.2015.07.006](#) PMID: [26215928](#)
33. Radauer C, Bublin M, Wagner S, Mari A, Breiteneder H. (2008) Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J Allergy Clin Immunol Pract* 121, 847–852.
34. Yusuf N, Nasti TH, Huang CM, Huber BS, Jaleel T, Lin HY, et al. (2009) Heat shock proteins HSP27 and HSP70 are present in the skin and are important mediators of allergic contact hypersensitivity. *J Immunol* 182, 675–683. PMID: [19109201](#)
35. Changchun H, Haijin Z, Wenjun L, Zhenyu L, Dan Z, Laiyu L, et al. (2011) Increased heat shock protein 70 levels in induced sputum and plasma correlate with severity of asthma patients. *Cell Stress Chaperones* 16, 663–671. doi: [10.1007/s12192-011-0271-9](#) PMID: [21643870](#)
36. Jin GB, Nakayama H, Shmyhlo M, Inoue S, Kondo M, Ikezawa Z, et al. (2003) High positive frequency of antibodies to metallothionein and heat shock protein 70 in sera of patients with metal allergy. *Clin Exp Immunol* 131, 275–279. PMID: [12562388](#)
37. Sookrung N, Wong-din-dam S, Tungtrongchitr A, Reamtong O, Indrawattana N, Sakolvaree Y, et al. (2014) Proteome and allergenome of Asian wasp, *Vespa affinis*, venom and IgE reactivity of the venom components. *J Proteome Res* 13, 1336–1344. doi: [10.1021/pr4009139](#) PMID: [24437991](#)
38. Rodríguez E, Romarís F, Lorenzo S, Moreno J, Bonay P, Ubeira FM, et al. (2006) A recombinant enolase from *Anisakis simplex* is differentially recognized in natural human and mouse experimental infections. *Med Microbiol Immunol* 195, 1–10. PMID: [16049725](#)
39. Kuehn A, Hilger C, Lehnert-Weber C, Codreanu-Morel F, Morisset M, Metz-Favre C, et al. (2013) Identification of enolases and aldolases as important fish allergens in cod, salmon and tuna: component resolved diagnosis using parvalbumin and the new allergens. *Clin Exp Allergy* 43, 811–822. doi: [10.1111/cea.12117](#) PMID: [23786287](#)
40. Bernardini R, Mistrello G, Novembre E, Roncarolo D, Zanotta S, Lombardi E, et al. (2005) Cross-reactivity between IgE-binding proteins from *Anisakis simplex* and *Dermatophagoides pteronyssinus*. *Int J Immunopathol Pharmacol* 18, 671–675. PMID: [16388714](#)
41. Nigro P, Pompilio G, Capogrossi MC. (2013) Cyclophilin A: a key player for human disease. *Cell Death Dis* 4, e888. doi: [10.1038/cddis.2013.410](#) PMID: [24176846](#)
42. Glaser A, Limacher A, Fluckiger S, Scheynius A, Scapozza L, Cramer R. (2006) Analysis of the cross-reactivity and of the 1.5 A crystal structure of the *Malassezia sympodialis* Mala s 6 allergen, a member of the cyclophilin pan-allergen family. *Biochem J* 396, 41–49. PMID: [16483252](#)
43. Valle C, Troiani AR, Lazzaretti P, Bouvier J, Cioli D, Klinkert MQ. (2005) Molecular and biochemical characterization of a protein cyclophilin from the nematode *Haemonchus contortus*. *Parasitol Res* 96, 199–205. PMID: [15830208](#)
44. Chen KY, Cheng CJ, Yen CM, Tang P, Wang LC. (2014) Comparative studies on the proteomic expression patterns in the third- and fifth-stage larvae of *Angiostrongylus cantonensis*. *Parasitol Res* 113, 3591–3600. doi: [10.1007/s00436-014-4024-4](#) PMID: [25028210](#)
45. González-Miguel J, Morchón R, Carretón E, Montoya-Alonso JA, Simón F. (2013) Surface associated antigens of *Dirofilaria immitis* adult worms activate the host fibrinolytic system. *Vet Parasitol* 196, 235–240. doi: [10.1016/j.vetpar.2013.01.028](#) PMID: [23433649](#)
46. Page AP. (2001) The nematode cuticle: synthesis, modification and mutants. In: Kennedy MW, Harnett W, editors. *Parasitic nematodes: molecular biology, biochemistry and immunology*. CABI Publishing, Wallingford, UK, pp. 167–193.
47. Yatsuda AP, Krijgsveld J, Cornelissen AW, Heck AJ, de Vries E. (2003) Comprehensive analysis of the secreted proteins of the parasite *Haemonchus contortus* reveals extensive sequence variation and differential immune recognition. *J Biol Chem* 278, 16941–16951. PMID: [12576473](#)
48. Ortona E, Vaccari S, Margutti P, Delunardo F, Rigano R, Profumo E, et al. (2002) Immunological characterization of *Echinococcus granulosus* cyclophilin, an allergen reactive with IgE and IgG4 from patients with cystic echinococcosis. *Clin Exp Immunol* 128, 124–130. PMID: [11982600](#)
49. Kennedy M. (2000) The nematode polyprotein allergens/antigens. *Parasitol Today* 16, 373–380. PMID: [10951596](#)

50. Meenan NA, Ball G, Bromek K, Uhrin D, Cooper A, Kennedy MW, et al. (2011) Solution structure of a repeated unit of the ABA-1 nematode polyprotein allergen of *Ascaris* reveals a novel fold and two discrete lipid-binding sites. PLoS Negl Trop Dis 5, e1040. doi: [10.1371/journal.pntd.0001040](https://doi.org/10.1371/journal.pntd.0001040) PMID: [21526216](https://pubmed.ncbi.nlm.nih.gov/21526216/)
51. Acevedo N, Sánchez J, Eler A, Mercado D, Briza P, Kennedy M, et al. (2009) IgE cross-reactivity between *Ascaris* and domestic mite allergens: the role of tropomyosin and the nematode polyprotein ABA-1. Allergy 64, 1635–1643. doi: [10.1111/j.1398-9995.2009.02084.x](https://doi.org/10.1111/j.1398-9995.2009.02084.x) PMID: [19624559](https://pubmed.ncbi.nlm.nih.gov/19624559/)



Minerva Access is the Institutional Repository of The University of Melbourne

**Author/s:**

Baird, FJ; Su, X; Aibinu, I; Nolan, MJ; Sugiyama, H; Otranto, D; Lopata, AL; Cantacessi, C

**Title:**

The Anisakis Transcriptome Provides a Resource for Fundamental and Applied Studies on Allergy-Causing Parasites

**Date:**

2016-07-01

**Citation:**

Baird, F. J., Su, X., Aibinu, I., Nolan, M. J., Sugiyama, H., Otranto, D., Lopata, A. L. & Cantacessi, C. (2016). The Anisakis Transcriptome Provides a Resource for Fundamental and Applied Studies on Allergy-Causing Parasites. PLOS NEGLECTED TROPICAL DISEASES, 10 (7), <https://doi.org/10.1371/journal.pntd.0004845>.

**Persistent Link:**

<http://hdl.handle.net/11343/257199>

**File Description:**

published version

**License:**

CC BY