

RESEARCH

Open Access

Identification of candidate mimicry proteins involved in parasite-driven phenotypic changes

Francois Olivier Hebert^{1*}, Luke Phelps², Irene Samonte², Mahesh Panchal², Stephan Grambauer³, Iain Barber³, Martin Kalbe², Christian R Landry¹ and Nadia Aubin-Horth¹

Abstract

Background: Endoparasites with complex life cycles are faced with several biological challenges, as they need to occupy various ecological niches throughout their development. Host phenotypes that increase the parasite's transmission rate to the next host have been extensively described, but few mechanistic explanations have been proposed to describe their proximate causes. In this study we explore the possibility that host phenotypic changes are triggered by the production of mimicry proteins from the parasite by using an ecological model system consisting of the infection of the threespine stickleback (*Gasterosteus aculeatus*) by the cestode *Schistocephalus solidus*.

Method: Using RNA-seq data, we assembled 9,093 protein-coding genes from which ORFs were predicted to generate a reference proteome. Based on a previously published method, we built two complementary analysis pipelines to i) establish a general classification of protein similarity among various species (pipeline A) and ii) identify candidate mimicry proteins showing specific host-parasite similarities (pipeline B), a key feature underlying the possibility of molecular mimicry.

Results: Ninety-four tapeworm proteins showed high local sequence homology with stickleback proteins. Four of these candidates correspond to secreted or membrane proteins that could be produced by the parasite and eventually be released in or be in contact with the host to modulate physiological pathways involved in various phenotypes (e.g. behaviors). One of these candidates belongs to the Wnt family, a large group of signaling molecules involved in cell-to-cell interactions and various developmental pathways. The three other candidates are involved in ion transport and post-translational protein modifications. We further confirmed that these four candidates are expressed in three different developmental stages of the cestode by RT-PCR, including the stages found in the host.

Conclusion: In this study, we identified mimicry candidate peptides from a behavior-altering cestode showing specific sequence similarity with host proteins. Despite their potential role in modulating host pathways that could lead to parasite-induced phenotypic changes and despite our confirmation that they are expressed in the developmental stage corresponding to the altered host behavior, further investigations will be needed to confirm their mechanistic role in the molecular cross-talk taking place between *S. solidus* and the threespine stickleback.

Keywords: Parasites, *Schistocephalus solidus*, Cestodes, Ecological genomics, Genomics/proteomics, Molecular mimicry, Host-parasite interactions, Wnt, RT-PCR

* Correspondence: francois-olivier.gagnon-hebert.1@ulaval.ca

¹Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Biologie, Université Laval, Pavillon Charles-Eugènes-Marchand, Québec G1V 0A6, Canada

Full list of author information is available at the end of the article

Background

Interspecific interactions among trophic levels can act as powerful drivers of biodiversity. Among the many possible ecological interactions that modulate energetic transfers through natural ecosystems, host-parasite interactions come up as the most frequent and widespread components of food webs [1]. Several million years of evolution led to the development of extremely diverse parasitic lifestyles, ranging from broad generalists having the potential to infect various species acting as their single host, to highly specialized species seeking the shelter of several specific intermediate hosts [2,3]. Endoparasites with complex life cycles are faced with several biological challenges, as they need to occupy various ecological niches throughout their development. This strategy requires them to keep their current host alive and ultimately find their way into a final host that is indispensable for reproduction [4]. Intermediate hosts involved in these complex life cycles can exhibit drastic parasite-driven phenotypic alterations that enhance the parasite's transmission rate, by making them more vulnerable to predation by the next host for instance [5]. As an example, rats infected with *Toxoplasma gondii* lose their fear of their predator as they become attracted by the smell of feline urine, thus increasing the parasite's chances of transmission to its mammalian definitive host [6,7]. One way of understanding such complex ecological interactions consists of characterizing the molecular cross-talk taking place between the parasite and its hosts [8].

Evidence from molecular analyses looking at the interaction between *T. gondii* and its murine host suggests that the behavioral change observed in infected rats is partly achieved through the expression of a tyrosine hydroxylase enzyme encoded in the parasite's genome. Interestingly, this protein is homologous to the one found in the host and directly alters dopamine levels in the rodent's brain [9]. Such empirical evidence suggests that one molecular mechanism that can be proposed to explain some of these behavior alterations by parasites involves the use of structural similarities between molecules, a phenomenon coined "molecular mimicry". The term molecular mimicry was first proposed by R. Damian [10] to describe antigen sharing between a parasite and its host. Consistent with this original concept, we use it here to define any molecular structure from the parasite that is similar to a corresponding host molecular structure and can thus potentially give an advantage to the parasite because of their shared similarity [11]. Some parasites use molecular mimicry to subvert host defenses as they express surface molecules similar to their host's antigens, therefore acting as a convenient camouflage [12]. Intracellular parasites can also produce mimicry molecules that interact with specific host proteins allowing them to maximize their cytoadherence

(*Trypanosoma cruzi*: [13,14], *Plasmodium falciparum*: [15,16]). Additionally, molecular mimicry can be a very powerful manipulative tool allowing the parasite to modify or suppress specific pathways in the host (e.g. hormonal messages, see [17-20]). When this strategy is pushed to the extreme, it can lead to serious behavioral changes. For instance, studies suggest that endoparasites like nematode hairworms could induce a water-seeking behavior in their orthopteran host (e.g. crickets, grasshoppers) by expressing mimicry signaling molecules likely to be involved in this unusual suicidal behavior [21,22]. This is one of the rare cases for which empirical evidence has been brought forward to explain the proximate causes of these behavioral changes [23]. Even though the consequences of being infected by "manipulative parasites" have been extensively described, the upstream causes of these phenotypic changes have not received enough attention yet to fully explain why and how infected individuals behave differently [8].

There are many examples of host-parasite interactions involving drastic changes in host phenotype. We chose to study the model system consisting of the infection in the threespine stickleback (*Gasterosteus aculeatus*) by the cestode *Schistocephalus solidus* as it allows us to test several possibilities with regards to molecular mechanisms [24]. *Schistocephalus solidus* is a trophically transmitted tapeworm with a complex life cycle involving two intermediate hosts. The definitive host is generally a piscivorous bird, but it can be any warm-blooded vertebrate [24]. Adult worms use the bird gut to complete the final stages of sexual maturation (i.e. egg production). Eggs released into the water through the bird's feces hatch to produce ciliated coracidia that will be trophically transmitted to any cyclopoid copepod (first intermediate invertebrate host). During the growth phase of the parasite, i.e. before becoming infective, copepods show an increased anti-predator response, which prevents potential premature transfer to the next host [25]. When larvae reach the infective stage (proceroid), copepods exhibit a reduced anti-predator behavior, leading to an increased transmission rate to the next host [26,27]. Infective proceroids will thus eventually find their way into the second obligatory intermediate host, the threespine stickleback (the only species they can successfully infect as second intermediate host, reviewed in [28]). Sticklebacks become infected when they feed on parasitized copepods, and after a few hours in the fish digestive track, proceroids will penetrate the wall of the intestine and migrate into the body cavity of the fish [29]. From there, they will transform into small plerocercoid worms that will grow to very large sizes, sometimes reaching the same mass as their host [30]. Phenotypic effects of parasitism include global physiological changes (e.g. altered reproductive potential, reviewed in [31] and altered immune response, see

[32,33]), change in prey choice [34] and a partial loss of competitive ability [35]. The time when the plerocercoids reach the developmental stage at which they could reproduce in their final bird host coincides with drastic changes in the stickleback's behavior resulting in increased predation rates by the definitive host [36,37]. Behavioral changes in the stickleback include decreased shoaling behavior [38], loss of anti-predator behavior and increased risk-taking behavior [39-42]. Although *S. solidus* infects the body cavity of its host and not the central nervous system, differences in metabolism and concentrations of neuromodulators (i.e. serotonin, epinephrine) are observed between infected and uninfected wild-caught sticklebacks [43].

There is extensive data on the physiological and behavioral impact of *S. solidus* on the stickleback [24,44], but to date, very few molecular mechanisms have been proposed to explain the proximate causes of these changes. Particularly, there is currently no empirical evidence pointing towards the existence or the type of signal that could be released by the worm to affect multiple host phenotypes (whether it is directly or indirectly triggered). Consequently, we investigated the possibility that *S. solidus* could take advantage of molecular mimicry to change its host phenotype (e.g. behavior, immunity, reproduction) using an iterative sequence similarity comparison approach. To do so, we first built a reference transcriptome for *S. solidus* from which we predicted protein sequences. We adapted a previously published method [11] to study molecular mimicry among these predicted tapeworm proteins by building two different pipelines aiming at i) establishing a general classification of protein similarity among various parasite, host and non-host species (pipeline A) and ii) identifying candidate mimicry proteins showing specific host-parasite similarities between *S. solidus* and the stickleback (pipeline B). If *S. solidus* relies on the use of molecular mimicry to change some of its host's phenotypes, being an extracellular parasite, it will have to express at its cell surface or secrete one or multiple types of effector molecules at one point during the infection. We can thus predict that the most plausible protein candidates involved in the development of a molecular signal triggered by the parasite and effective over a distance, either directly or indirectly (i.e. through physiological cascade that ultimately affects the host's central nervous system), will likely be secreted proteins. We confirmed that the candidate genes, selected by their signal peptide (i.e. secretory signal) and high similarity between *S. solidus* and its stickleback host, were expressed in three different developmental stages of the parasite, i.e. non-infective (no host behavioral change), infective (significant host behavioral change) and post-reproduction adult (after egg production in the final host). This first candidate validation serves as a

stepping-stone towards a fully functional characterization of the molecular interaction occurring between *S. solidus* and its second intermediate host.

Methods

Schistocephalus solidus transcriptome assembly

Worms used to produce the transcriptome were collected in two different populations, one in Norway and one in Germany. RNA was extracted using Macherey-Nagel's NucleoSpin® commercial kit (Düren, Germany) according to the manufacturer's protocol. Two different 454 libraries were produced (GS-FLX platform), each containing eight pooled worms collected at three different time points: i) five weeks post-infection (four worms), ii) seven weeks post-infection (two worms) and iii) nine weeks post-infection (two worms) [EMBL:ERS551497, EMBL:ERS551498]. Worms used to produce these 454 libraries covered three developmental stages that can be found within a fish host, i.e. non-infective (parasite mass < 50 mg, no change in host behavior), infective (parasite mass > 50 mg, significant changes in host behavior) and the transition stage from non-infective to infective.

Raw reads were first cleaned using NGS backbone [45] based on quality and length thresholds (PHRED score \geq 20, read length \geq 100 nucleotides). Cleaned reads were subsequently assembled *de novo* using a combination of MIRA 4.0 [46] and RAY 2.3.0 [47]. The MIRA algorithm is an overlap-layout-consensus method, which uses trace signals and additional sequence information whereas the RAY algorithm is a k-mer-based method relying on a de Bruijn graph. To run MIRA, we used the default parameters to perform transcript assembly (job = est, denovo, accurate). Contigs tagged by MIRA as "repetitive", i.e. chimeras generated with highly repetitive reads [48], were discarded after protein ID validation with *blastx* 2.2.29 [49], using different protein databases (swissprot, nr, ftp.ncbi.nlm.nih.gov, accessed on 12/2014). For the second assembly, we took advantage of RAY's "additive Multiple-k" method [50] by pooling contigs obtained with different k-mer values (k = 41, 43, 45, 47, 49). We then used an incremental clustering implemented in the program CD-HIT-EST [51,52] to remove redundancy and to generate the longest and most accurate contigs possible (see Additional file 1 for details and threshold values).

After generating two independent "cleaned datasets", contigs from both assemblies were locally aligned (*blastn*, [49]) against a raw version of the *Schistocephalus solidus* genome (50 Helminth Genomes Initiative, ftp.sanger.ac.uk/pub/pathogens/HGI/). Contigs with either no hit found in the genome or showing low quality blast results were filtered out (e-value threshold = $1e-15$). This procedure was carried out to eliminate potential cDNA contamination from the host fish from which the

parasite worm was extracted, as well as chimeras and false gene sequences. The two datasets were then compared to each other using CD-HIT-EST-2D [53] to identify shared sequences. We applied the same similarity and length coverage thresholds as previously used with CD-HIT-EST. Using custom-made Python scripts (<https://github.com/fohebert/Scripts.git>), we discarded short redundant sequences identified by CD-HIT-EST-2D (thus eliminating redundancy in the reference transcriptome) and excessively long representative sequences (>10,000 nucleotides) more likely to regroup repeated sequences and chimeras, i.e. multiple different contigs aligning on one very long contig [53,54]. Remaining contigs formed the final dataset.

Using the EMBOSS function *getorf* [55], we obtained all possible ORFs and predicted amino acid sequences (forward and reverse) for every contig retained in the final dataset. For each contig, the longest ORF was

selected. We finally used *blastx* to retrieve the identity of these sequences according to local databases created with several parasite proteomes (Table 1). This highly filtered dataset specific to *Schistocephalus solidus* was used as a reference proteome for further analyses.

Control and host proteome files

Protein sequences for protein-coding genes from completely sequenced genomes were downloaded from <ftp.ebi.ac.uk> (*Arabidopsis thaliana*, *Schizosaccharomyces pombe*), <ftp.wormbase.org> (*Caenorhabditis elegans*), <http://uniprot.org> (*Brugia malayi*, *Trichomonas vaginalis*, *Trichoplax adhaerens*), <ftp.vector-base.org> (*Anopheles gambiae*), <http://cryptodb.org> (*Cryptosporidium parvum*), <ftp.sanger.ac.uk> (*Echinococcus granulosus*, *Echinococcus multilocularis*, *Hymenolepis microstoma*, *Schistosoma mansoni*), <http://giardiadb.org> (*Giardia lamblia*), <http://tritrypdb.org> (*Leishmania major*, *Trypanosoma cruzi*),

Table 1 Species used as control, host and parasite references for protein identification

Species	Common name	Phylum (class)	Source
Host			
<i>Gasterosteus aculeatus</i>	Threespine stickleback	Chordata (Actinopterygii)	Ensembl
Non-host control fishes			
<i>Xiphophorus maculatus</i>	Platyfish	Chordata (Actinopterygii)	Ensembl
<i>Oryzias latipes</i>	Medaka	Chordata (Actinopterygii)	Ensembl
<i>Lepisosteus oculatus</i>	Spotted gar	Chordata (Actinopterygii)	Ensembl
<i>Takifugu rubripes</i>	Japanese puffer	Chordata (Actinopterygii)	Ensembl
Parasites			
<i>Brugia malayi</i>	Filarial nematode	Nematoda (Onchocercidae)	Uniprot
<i>Cryptosporidium parvum</i>	—	Apicomplexa (Conoidasida)	CryptoDB
<i>Echinococcus granulosus</i>	Dog tapeworm	Plathelminthes (Cestoda)	Sanger
<i>Echinococcus multilocularis</i>	Fox tapeworm	Plathelminthes (Cestoda)	Sanger
<i>Giardia lamblia</i>	—	Metamonada (Eopharyngia)	GiardiaDB
<i>Hymenolepis microstoma</i>	Rodent tapeworm	Plathelminthes (Cestoda)	Sanger
<i>Leishmania major</i>	—	Euglenozoa (Kinetoplastidia)	TritrypDB
<i>Plasmodium falciparum</i>	Malaria	Apicomplexa (Aconoidasida)	PlasmoDB
<i>Schistosoma mansoni</i>	Blood fluke	Plathelminthes (Trematoda)	Sanger
<i>Trypanosoma cruzi</i>	Chagas	Euglenozoa (Kinetoplastidia)	TritrypDB
<i>Taenia solium</i>	Pork tapeworm	Plathelminthes (Cestoda)	GeneDB
<i>Trichomonas vaginalis</i>	—	Metamonada (Parabasalia)	Uniprot
Controls			
<i>Anopheles gambiae</i>	African malaria mosquito	Arthropoda (Culicidae)	Vectorbase
<i>Arabidopsis thaliana</i>	Mouse-ear cress, thale cress	Magnoliophyta (Magnoliopsida)	EBI
<i>Caenorhabditis elegans</i>	"Roundworm"	Nematoda (Chromadorea)	Wormbase
<i>Ciona intestinalis</i>	Vase tunicate, Sea squirt	Chordata (Ascidacea)	JGI
<i>Schizosaccharomyces pombe</i>	Fission yeast	Ascomycota (Schizosaccharomycetes)	EBI
<i>Capitella teleta</i>	—	Annelida (Polychaeta)	Ensembl
<i>Trichoplax adhaerens</i>	—	Placozoa (Tricoplacia)	Uniprot

<http://plasmodb.org> (*Plasmodium falciparum* 3D7), <http://broadinstitute.org> (*Gasterosteus aculeatus*), <ftp://ftp.ensemblgenomes.org> (*Capitella teleta*, *Xiphophorus maculatus*, *Oryzias latipes*, *Lepisosteus oculatus*, *Takifugu rubripes*) and <http://genedb.org> (*Taenia solium*). These proteomes were used either as controls for conserved sequences (various non-parasitic species), non-host fish controls or parasite-specific sequences (parasitic species used throughout the assembly process), while the genome of *G. aculeatus* (<ftp://ftp.ensemblgenomes.org>) was used as the host genome (Table 1).

Pipeline A - general parasitic protein similarity analysis

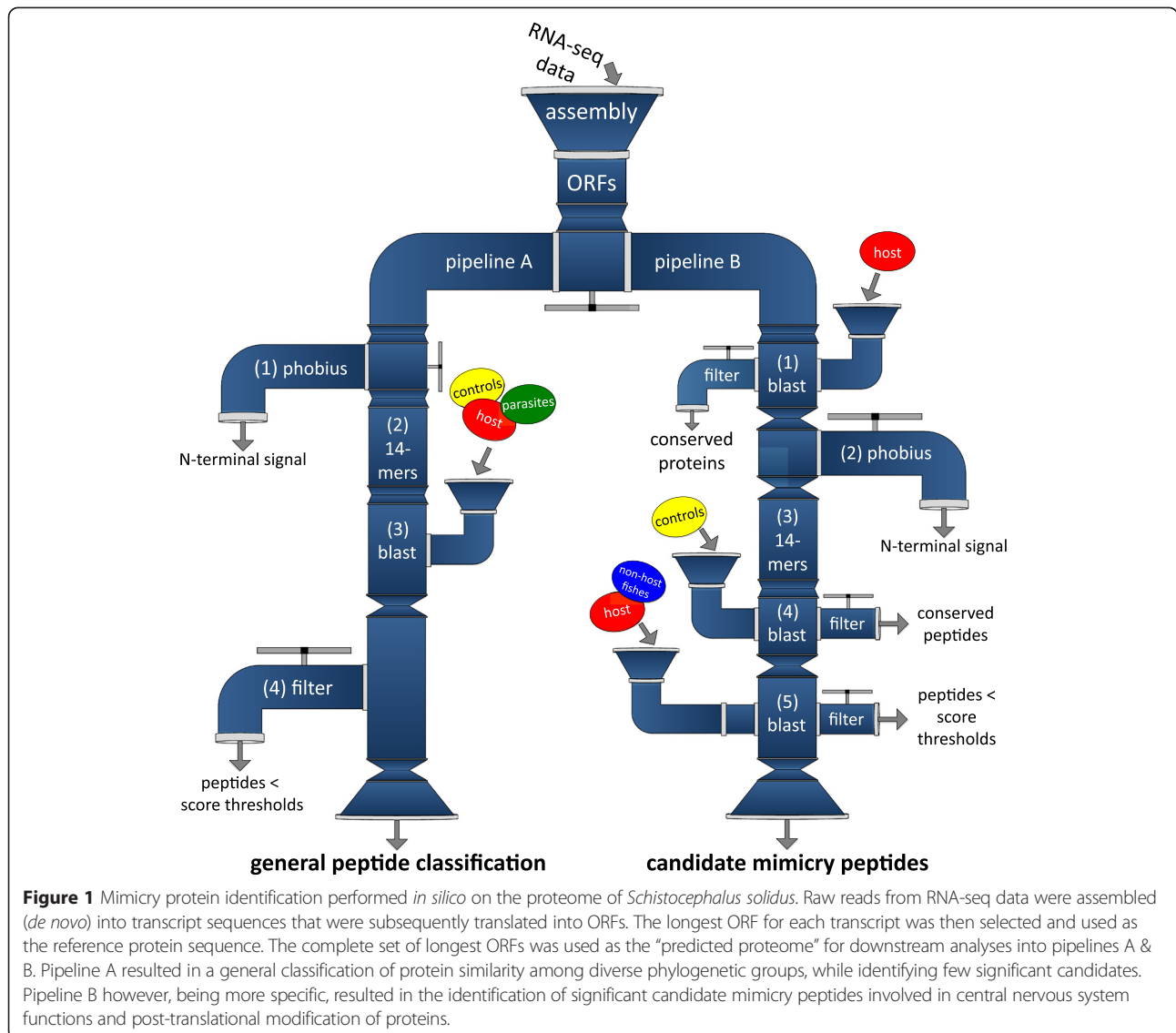
In a first exploratory phase, the proteomes of *S. solidus* as well as six other worms were compared to the stickleback proteome using *blastp*. This procedure was carried out to verify if the stickleback proteome shares a higher similarity with its parasite proteome than various other parasitic and non-parasitic free-living worms that do not have a specific co-evolutionary background with the host. Among these six proteomes, four are from parasite species closely related to *S. solidus* (phylum *Cestoda*: *Echinococcus granulosus*, dog tapeworm; *Echinococcus multilocularis*, fox tapeworm; *Hymenolepis microstoma*, rodent tapeworm; *Taenia solium*, pork tapeworm). The two other proteomes belong to non-parasitic free-living worms, i.e. *Caenorhabditis elegans* (phylum *Nematoda*) and *Capitella teleta* (phylum *Annelida*). *C. teleta* was chosen mainly because it is a non-parasitic marine polychaete that belongs to a phylogenetic group sharing a common ancestor with cestodes that is more recent than the common ancestor between cestodes and nematodes [56]. It thus acts as a solid non-parasitic control that is closely related to *S. solidus*.

Several virulence factors previously identified in pathogens show strong sequence and structural homology to host proteins [57]. However, other pathogenic effectors show no apparent sequence similarity to any host protein, but display mimicry for short fragments of the protein only (e.g. tyrosine phosphatases in *Yersinia* spp, [58,59] and *Salmonella* spp., [60,61]). Additional sequence similarity analyses based on fragmented proteins between *S. solidus* and its host were thus required to detect such cases of more cryptic molecular mimicry. This task was performed through pipelines A and B. First, pipeline A was used with fragmented *S. solidus* proteins and *blastp* searches against the proteomes of i) control, ii) parasite and iii) host species (Table 1) to test for general protein similarity between the parasite and various phyla (Figure 1, pipeline A). This procedure enabled us to label fragmented proteins according to their species specificity, i.e. with which species they share a certain degree of similarity (conserved vs. species-specific proteins/peptides). The following steps describing the details of pipeline A are

inspired from the method presented in [11] to address the issue of sequence similarity at the protein fragments level, but was adapted to our study system. Most of the pipeline parameters found in [11] are re-used here, with minor changes in the peptide selection process, as we wanted to generate a general classification of protein similarity first. To do so, we used additional proteomes and we did not discard any peptide, but labeled them instead, according to their blast score. More specifically, a hidden Markov Model (trained on all swissprot entries) implemented in the program PHOBIUS [62] was used to predict N-terminus signals in *S. solidus* proteins. We trimmed out these short sequences using custom-made Python scripts (<https://github.com/fohebert/Scripts.git>) based on PHOBIUS output to eliminate potential false positives, as N-terminus signals can be shared among various types of proteins [62]. Proteins were then cleaved into short overlapping fragments of 14 residues (14-mers) with an incremental sliding window of one. General protein similarity was obtained by performing sequential *blastp* searches of *S. solidus* 14-mers against control, parasite and host species. Queries with ungapped *blastp* identities above pre-defined thresholds (Additional file 2, as defined in [11]) were considered highly similar to their hit sequence. The choice of k-mer length was made based on the method developed by Ludin et al. [11], as similarity thresholds for this particular length were tested and empirically validated (see Additional file 2). Peptides of 14 residues also represent a fair compromise between specificity and sensitivity, since short sequences are prone to align everywhere on the proteome (thus less specific), whereas long peptides can potentially return a high rate of false negatives (being too stringent). Results were then used to build a preliminary classification of protein specificity and to get an overview of the proportion of potential host-specific proteins encoded in the genome of *S. solidus*.

Pipeline B - host-parasite specific protein similarity analysis

Identification of candidate mimicry proteins was achieved by the use of an iterative sequence similarity approach designed to identify highly similar peptides between a parasite and its host (pipeline B in Figure 1, adapted from the method described in [11]). We used custom Python scripts (<https://github.com/fohebert/Scripts.git>) to adapt the original idea from Ludin et al. [11] to our study system, as presented in Figure 2 (pipeline B). First, sequence similarity between the *S. solidus* proteome and several control proteomes (Table 1) was assessed by using ungapped *blastp* searches. Proteins returning an e-value $\leq 1e-15$ were discarded, as they were considered generally conserved among eukaryotes. Proteins exhibiting low similarity with control proteomes were then analyzed using PHOBIUS to identify potential N-terminus signals. Again, these short and conserved sequences were



trimmed out using our custom-made Python scripts designed for pipeline A. Trimmed sequences were subsequently fragmented in overlapping 14-mers, following the same procedure used in the first pipeline (Figure 1). These *S. solidus* 14-mers were blasted against the complete array of control proteomes and sequences above pre-defined identity thresholds (as empirically determined and validated in [11], see red solid line, Additional file 2) were discarded. Remaining *S. solidus*-specific peptides (i.e. 14-mers that share very low or no similarity with control proteomes) were screened against the host proteome with an ungapped *blastp* search. These peptides were also screened against five non-host fish proteomes (*Danio rerio*, *Xiphophorus maculatus*, *Oryzias latipes*, *Lepisosteus oculatus* and *Takifugu rubripes*, see Table 1), all sharing a distant common ancestor with the stickleback [63-65] and for which no infection by *S. solidus* has ever been reported

due to the high specificity of the infection [28]. This extra step was performed in order to assess the proportion of candidate 14-mers expected by chance only when investigating any fish proteome, although we cannot exclude that mimicry proteins might resemble proteins generally conserved among fish taxa. Queries returning hits with sufficient similarity were labeled as molecular mimicry candidates. GoMiner [66] was ultimately used to perform an enrichment analysis for gene ontology terms on these final sequences to identify potential biological functions over-represented among candidates. Final candidates were also screened for the presence of a secretory signal peptide using SignalIP [67] to label proteins as secreted or non-secreted. As an additional control to assess the statistical significance of the method, we performed the same analysis (i.e. pipeline B) on a randomized version

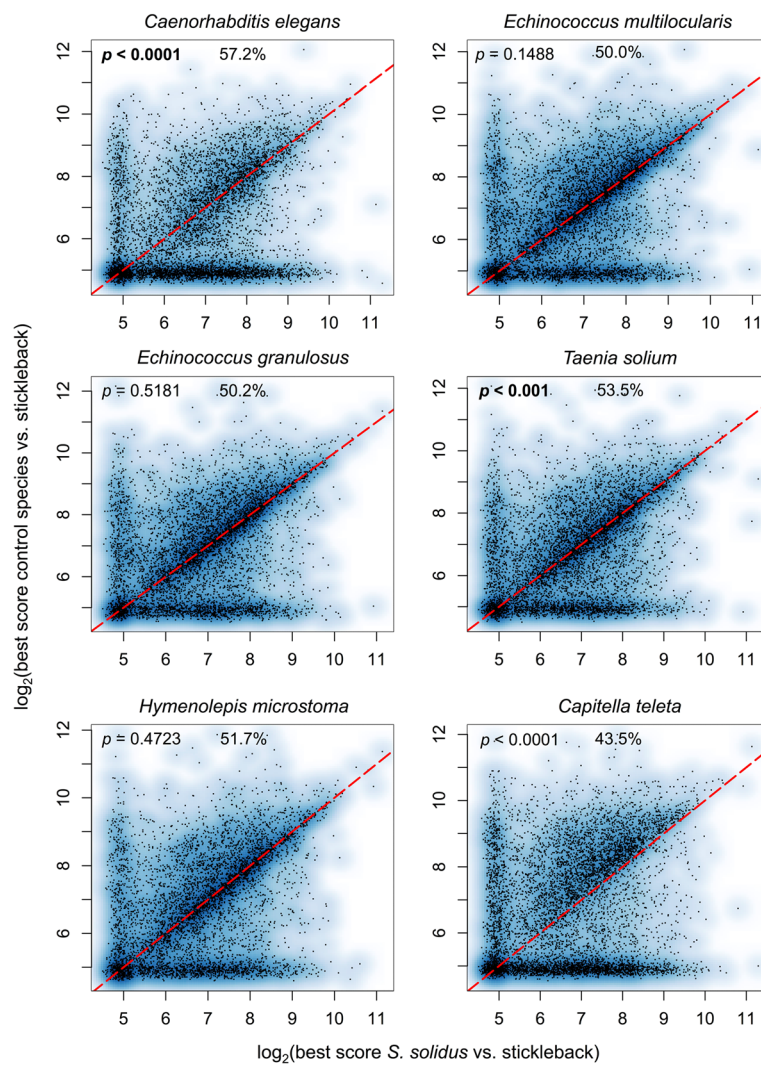


Figure 2 Protein similarity comparisons using full-length *blastp* searches against various parasitic and non-parasitic control worm species. Scatter plots of the best *blastp* scores of *S. solidus* proteome (x axis) and six control worm species proteomes (y axis, name of the species specified on top of each graph) against the proteome of the threespine stickleback. Data below the red dotted line correspond to stickleback proteins showing a higher sequence similarity for *S. solidus* as compared to the corresponding control worm species. For each scatter plot, a Wilcoxon signed-rank test was performed on the distributions of *blastp* scores (*p*-value in the upper left corner of each graph). Significant *p*-values were highlighted in bold for higher *blastp* scores between *S. solidus* and its host than between the control species and the host. The percentage of the stickleback proteome showing higher scores with *S. solidus* than with the control species appears on top of each scatterplot.

of the proteome of *S. solidus* using EMBOSS function *shuffleseq* [11,55]. This second pipeline is thus similar to pipeline A, although it does not include other parasite proteomes. By doing so, proteins potentially involved in convergent molecular mechanisms of phenotypic alteration will not be discarded by comparing *S. solidus* proteins to other parasite proteins [68]. Pipeline A thus served as a first exploratory phase that allowed us to classify proteins according to their level of conservation across a wide range of phylogenetic groups, whereas pipeline B was specifically used to identify mimicry candidates between *S. solidus* and the threespine stickleback.

RT-PCR validation

We further confirmed that our refined set of candidates identified through pipeline B were expressed by the parasite by performing a reverse transcription polymerase chain reaction (RT-PCR) on 17 additional worms from a different population than the one used to build the transcriptome and from three different developmental stages, i.e. pre-infective (n = 7), infective (n = 7, > 50 mg) and post-reproduction adults (n = 3, > 350 mg) in a simulated bird gut. Worms were extracted from a population of lab-raised and artificially infected sticklebacks at the University of Leicester, England (UK).

Details of the method and complete primer sequences are available in Additional file 3.

Host contamination control

To eliminate potential fish cDNA contamination that could have been introduced in the transcriptome due to host tissues inadvertently left on the parasite's integument during the dissection, three different bioinformatics controls were used. First, cleaned reads were mapped to the raw genome of *Schistocephalus solidus* using BWA-SW [69] with default parameters. Default parameters for the program BWA are designed to offer the best possible balance between performance and accuracy. The program also automatically adjusts parameters according to read length and error rates [60]. BWA default parameters are thus sufficiently efficient to achieve the goal of discarding low quality reads that do not match the reference genome. Reads that did not map on the genome were discarded from the assembly. The second control was performed at the end of the assembly. Final contigs obtained with both assembly methods (RAY and MIRA) were blasted against the raw genome of *S. solidus* and sequences returning no hits were discarded. A third control was finally used to confirm that the mimicry candidates found with pipeline B (Figure 2) were tapeworm proteins and not cDNA contamination from the host. Candidate proteins were blasted against the proteomes of the host and the parasite (*blastp* searches) and cDNA sequences corresponding to these proteins were also respectively blasted against the genomes of the host and the parasite (*blastn* searches), which allowed us to assess if these sequences (at the levels of nucleic acids and amino acids) were more similar to the tapeworm or the fish proteome. When the e-value, the raw bit score and the length of alignment were systematically greater when blasted against the parasite as compared to the host, candidates were considered as true parasite sequences and not contamination. We also performed a fourth control in the laboratory to determine whether the four candidate mimicry genes do originate from *Schistocephalus solidus* and not from threespine stickleback DNA contamination. To do so, we conducted a PCR validation experiment using DNA from three pools of coracidia, the free swimming stage of *Schistocephalus solidus* that has never been in contact with the fish host, where each pool came from one breeding pair. We also used genomic DNA (gDNA) and cDNA samples from three individual adult worms and host fish (Additional file 3).

Ethical approval

All animal experiments that were performed at the Max Planck Institute for Evolutionary Biology (Plön, Germany) were approved by the 'Ministry of Energy, Agriculture, the Environment and Rural Areas' of the state of Schleswig-

Holstein, Germany (reference number: V 313–72241.123–34). Fish were captured under U.K. Environment Agency permit and with the permission of the landowner. All animal experiments performed at the University of Leicester (England, UK) were undertaken under a U.K. Home Office license (PPL80/2327), in accordance with local and national regulations, and in line with ABS/ASAB guidelines for the ethical treatment of animals in behavioral research (available online at <http://asab.nottingham.ac.uk/ethics/guidelines.php>).

Results and discussion

Transcriptome assembly

The multiple assembly strategy used in this study yielded a total of 9,093 putative protein-coding genes (Table 2, Additional files 4 and 5), which is slightly lower than expected based on recent flatworm genome assemblies (between 10,231 and 12,490 genes, see [54]) but on the same order of magnitude. Two plausible reasons could explain this discrepancy: i) not all life stages and hosts were sampled (earlier larval stages and adult missing) and ii) a relatively low median coverage (average: 7X, median: 6.93X) could have resulted in the elimination of "true gene" transcripts which have insufficient coverage. However, since there is no published consensus transcriptome or complete genome annotation for *S. solidus*, the number of genes and these explanations remain speculative and we consider our predicted proteome dataset (referred to as proteome) to be a conservative estimate. On the other hand, the most relevant developmental stage required to answer the main question asked in this study (nine weeks post-infection, when behavioral changes in the host would be apparent) was sampled. Relevant genes involved in molecular mimicry at this stage and in the fish are therefore likely to be contained in the dataset used to perform our analyses.

Table 2 Transcriptome assembly metrics

Total number of reads	528 595
Number of reads mapping on raw genome ¹	471 018 (89%)
Assembled reads	462 128
Number of gene-contigs	9093
Average length in bp (range; median)	1266 (212 – 7588; 1097)
N50	945
Average coverage (range; median)	7.03X (4 – 1850; 6.93)
Average protein length in amino acids (range; median)	300 (20 – 1992; 258)
Number of annotated proteins ²	5949

¹Genome available from 50 Helminth Genomes Initiative (<http://sanger.ac.uk/>).

²Based on swissprot database.

Comparisons with full-length proteins

Results from a first glance at the dataset, using full-length *blastp* searches among various parasitic and non-parasitic worms, confirmed the potential for candidate mimicry identification among *S. solidus*' proteins (Figure 2). Full-length *blastp* searches on the proteomes of *S. solidus* and six other worm species (four of which are cestodes and two are non-parasitic free living worms) revealed various levels of protein similarities depending on the species being compared to *G. aculeatus*. When *blastp* scores between *S. solidus* and its host were compared to *blastp* scores between *C. elegans* and *G. aculeatus*, for a given host protein, on average, *S. solidus* shared a significantly higher blast score ($p < 0.0001$, two-tailed Wilcoxon test, Figure 2). This trend was also true when the association *S. solidus*-stickleback was compared to the association *T. solium*-stickleback, i.e. stickleback proteins were, on average, more similar to *S. solidus*' proteins than *T. solium*'s proteins ($p < 0.001$, two-tailed Wilcoxon test, Figure 2). However, when the comparison is performed with any of the three other cestodes, there is no significant difference between distributions of *blastp* scores ($p = 0.1488$ for *S. solidus* vs. *E. multilocularis*, $p = 0.5181$ for *S. solidus* vs. *H. microstoma*, $p = 0.4723$ for *S. solidus* vs. *E. multilocularis*, two-tailed Wilcoxon tests, Figure 2). When the proteome of *C. teleta*, a non-parasitic marine polychaete, was blasted against the stickleback proteome, we observed significantly higher blast scores between the two species, as compared to the scores obtained between *S. solidus* and the stickleback ($p < 0.0001$, two-tailed Wilcoxon test, Figure 2). This could be due to the fact that the proteome of *C. teleta* contains a higher number of proteins than the proteome of *S. solidus* (32 175 and 9 093 respectively). It is thus expected that by chance alone, more similarities can be found when the *blastp* search against the stickleback proteome is performed with a larger set of proteins.

Insights from a general parasitic protein similarity analysis (pipeline A)

After this first round of full-length *blastp* searches, predicted *S. solidus* proteins were analyzed through pipeline A. Predicted *S. solidus* proteins were fragmented and compared to several other proteomes (see Methods, Table 1), which provided a general classification of protein similarity among various phyla (Figure 1, pipeline A). In total, 8,786 proteins passed the similarity thresholds, returning significant hits on various species. Most of these proteins were widely distributed among phyla (total = 86%; controls & parasites = 4877 proteins, 53%; controls & parasites & host = 2981 proteins, 33%), while small proportions were assigned to a given group only (controls = 338, 3.7%; parasites = 562, 6.2%). Based on empirically determined thresholds (Additional file 2, see [11]), only three *S. solidus* proteins showed a high degree of sequence

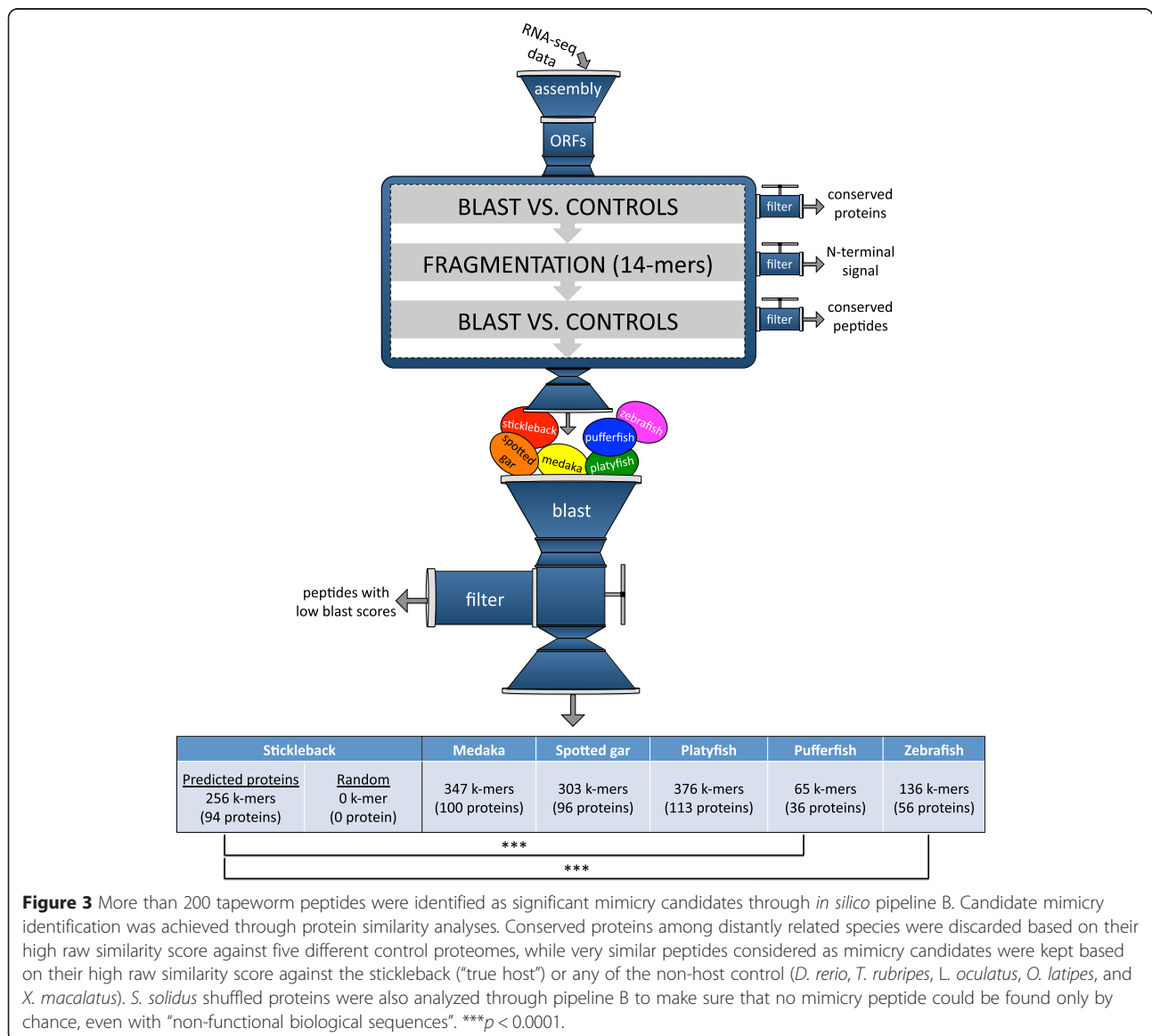
similarity exclusively with host proteins: tektin-4, partial coding sequence from jockey-like mobile element, unknown predicted protein. According to pipeline A, sequences falling into this category were deemed the most interesting candidates for molecular mimicry, i.e. short peptides showing strong homology to a corresponding host protein (between 85% and 100% sequence similarity). However, none of these sequences had a gene identifier that could directly associate them to a molecular mimicry strategy. One of these candidate proteins is similar to tektin-4 [UniProt:GAA27704], which is involved in microtubule cytoskeleton organization, therefore not secreted or expressed at the cell surface. Another candidate corresponds to a polymerase from a mobile element [UniProt: CCD80178], thus not involved in any physiological process or biological function that would relate to the host phenotype. The last candidate, a relatively short unknown protein (73 amino acids) shares only 11 identical residues with an unknown stickleback protein (275 amino acids, Ensembl:ENSGACT00000002277). Since no function can be established based on current information, either for the tapeworm peptide or for the host target, this last candidate will require further studies in order to confirm its role in host behavioral changes.

Identification of mimicry peptides through pipeline B

In this last sequence similarity analysis performed on *S. solidus* proteome, predicted proteins were fragmented again and compared with control and host proteomes only (see Methods, Figure 1, pipeline B), allowing the identification of very specific peptides potentially involved in the phenotypic changes observed in parasitized sticklebacks. In this analysis, as in pipeline A, peptides (14-mers) were generated based on full-length protein sequences. However, in this second analysis, more stringent criteria were used and this time, tapeworm proteins were not compared to proteomes from vertebrate parasites (see Methods for details). The complete predicted proteome dataset for *S. solidus* (9,093 proteins) was fragmented in 1,240,045 overlapping 14-mer peptides that were successively compared with all control species according to pipeline B (see Methods, Figure 1). After discarding conserved proteins and peptides, empirically determined similarity thresholds with the stickleback proteome allowed the identification of 287 candidate 14-mers (0.023% of all 14-mers) distributed in 94 proteins (Additional file 6). More than half of these candidate peptides (167) shared 100% similarity (i.e. 14 identical residues) with a host sequence. Based on the third control for host cDNA contamination (see Methods), we confirmed that these 14-mer candidates are part of true tapeworm proteins, as these proteins (i.e. proteins containing the candidate peptides) were always more similar to *S. solidus* proteins or any other closely related parasites (e.g.

cestodes; *Echinococcus granulosus*, *Hymenolepis microstoma*, trematodes; *Clonorchis sinensis*, *Opisthorchis viverrini*, *Schistosoma mansoni*) than any fish protein (full-length *blastp* searches). We tested if these proteins were enriched for particular biological functions using GoMiner [66]. No gene ontology (GO) category was over-represented among these candidates (data not shown). To assess the statistical relevance of the candidate 14-mers, the same procedure was carried out with a randomized version of the *S. solidus* proteome. The rationale behind this test is that finding as many mimicry candidates with a randomized version of the proteome as in the predicted proteome would mean that our results have no biological meaning. Shuffled sequences did not return any candidate, thus underlining the potential significance of the candidate 14-mers identified among real sequences (Figure 3).

A final control for the method was performed using fishes that are usually not infected by *S. solidus* in the wild. This additional analysis acts as a validation step aimed at testing the method in a non-specific context, i.e. when parasite-specific peptides are screened against proteins from non-host species (see Methods). By successively screening parasite-specific peptides against five different non-host fish proteomes through pipeline B, we found 136 (in 56 proteins), 303 (in 96 proteins), 347 (in 100 proteins), 65 (in 37 proteins) and 376 (in 113 proteins) candidate mimicry peptides when using *D. rerio*, *L. oculatus*, *O. latipes*, *T. rubripes* and *X. maculatus*, respectively (Figure 3). At the peptide level, we found significantly more mimicry candidates for non-host species when screened against *O. latipes* and *X. maculatus* as compared to the stickleback ($p = 0.0172$



and $p < 0.0001$ respectively, 2-sample tests of equal proportions). On the other hand, we found no significant difference between the numbers of mimicry candidates identified with the stickleback versus *L. oculatus* ($p = 0.51$, 2-sample test of equal proportions). Results also indicate that significantly more mimicry peptides can be found when screening against the stickleback as compared to screening against *D. rerio* or *T. rubripes* ($p < 0.0001$ for both screens, 2-sample tests of equal proportions). When looking at the protein level, we found no significant difference between numbers of candidate mimicry proteins identified with non-hosts *L. oculatus*, *O. latipes*, and *X. maculatus* as compared to the real host ($p = 0.885$, $p = 0.885$, $p = 0.189$ respectively, 2-sample tests of equal proportions). Finally, we found significantly more candidate mimicry proteins when screening against the real host than when screening against either non-hosts *D. rerio* or *T. rubripes* ($p < 0.0001$, 2-sample tests of equal proportions, see Figure 3).

Overall, similar numbers of candidate mimicry proteins can be found when using either the real host or three out of five non-host fishes (i.e. *L. oculatus*, *O. latipes*, and *X. maculatus*). If *S. solidus* uses a molecular mimicry strategy to complete its life cycle, we can assume that the mimicry proteins produced by the parasite will be very similar to fish proteins (if this is a case of mimicry created by sequence similarity, whereas three-dimensional structural, as well as functional mimicry [70] cannot be identified using this method). Finding similar results when screening against non-host fishes as compared to the real fish host may not be surprising, considering that the mimicry strategy can be targeting any common pathway found in fishes (or vertebrates). Moreover, since we did not use any fish or vertebrate proteome among the control group for pipeline B, proteins and peptides highly conserved across vertebrate species were kept throughout the analyses. We can thus assume that some of the mimicry peptides identified

when screening against non-host fishes reflect peptide/protein conservation across fish or vertebrate taxa. Results suggest that this mimicry identification method can efficiently isolate high profile candidates. However, it cannot assess their true biological role in the interaction between the parasite and its host as functional studies are required to perform this task.

Secreted proteins: the most plausible candidates

Schistocephalus solidus is an endoparasite living in the abdominal cavity of its host. As a consequence, if it uses mimicry proteins to affect physiological and cellular pathways in its host, the “phenotype-altering signal” is most likely to come from a protein that is secreted by the parasite and released in the bodily fluids of the host or that is expressed at the cell surface. Strong candidates for molecular mimicry should thus be secreted and/or membrane proteins. Among all proteins containing 14-mer candidates, only four could be labeled as secreted or cell-surface localized (i.e. containing a secretory signal peptide) based on SignalIP results (see Methods, Table 3 & Additional file 7). Interestingly, one of these candidates belongs to the WNT proteins family (14-residue homology), WNT4 [Uniprot:A0A068WB45], a large group of signaling molecules involved in cell-to-cell interactions and various developmental pathways [71]. This candidate is also very similar to the protein WNT5B (found in the fish *Danio rerio*, Uniprot:NP_571012), which plays a role in the development of discrete regions of tissues. Previous proteomics investigations found that molecules from the Wnt family were over-expressed in the head of crickets and grasshoppers infected by a behavior-altering hairworm [21,22]. Specifically, protein fragments from the Wnt family involved in the central nervous system (CNS) development and produced by the parasite were found in infected orthopterans exhibiting abnormal suicidal behaviors. Even though the WNT4 protein identified in our study is not known to be

Table 3 Secreted proteins identified as mimicry candidates through pipeline B

Parasite protein	No. of 14-mers with significant BLASTp results	Ave. Identity	GO terms	Protein ID ¹ (Access. no)	Function ²
Ssc1854	1	12/14	Ion transport; Zinc transport	Zinc transporter, ZIP12 (Q504Y0)	Zinc influx transporter.
Ssc2726	2	12/13	Transcription regulation; Transcription	Lysyl oxidase homolog 2B, LOXL2B (A1L1V4)	Mediates the post-translational oxidative deamination of lysine residues
Ssc5533	2	12/14	Ion transport; Transport	Palmitoyltransferase, (ZDHCC17) EUB64135	Transferase activity. Lipid modification activity involved in protein trafficking and function in the central nervous system.
Ssc6185	1	13/14	Wnt signaling pathway	Protein WNT4 A0A068WB45	Signaling molecule possibly involved in the development of discrete regions of tissues.

¹Best blast hit when screened against Uniprot (<http://uniprot.org>).

²Information on function taken from Uniprot.

directly involved in the development of the CNS in fishes, the common point between these different host-parasite systems could be a general disruption of cell-to-cell communication, leading to various changes in behavior [23,72]. Another *S. solidus* peptide matched to a membrane zinc transporter [Uniprot:Q504Y0] (14-residue homology) suspected to be involved in the development of schizophrenia in humans [73]. Its biological function in *S. solidus* or as a mimicry protein in its fish host remains to be investigated. Two other parasite-specific 14-mers showed high similarity (19-residue homology) with a zinc finger protein responsible for a palmitoyltransferase activity in the stickleback [ZDHHC18, Ensembl: ENSGACT0000009617]. Palmitoylation represents one of many different types of post-translational modifications of proteins. Specifically, it involves the addition of the palmitate lipid in a thioester linkage on cysteine residues [74,75]. It has been shown that palmitoylation of neuronal proteins like PSD-95 in humans can lead to changes in synaptic plasticity, thus potentially changing the way the information is transmitted throughout the central nervous system [76]. While empirical evidence supports the role of these lipid

modifications in the dynamic regulation of protein function and neuronal signaling [76], no direct link can yet be established between this protein and phenotypic changes in the stickleback. Overall, pipeline B allowed us to identify significant candidates with potential roles in cell signaling or cognitive pathways, but their direct impact on host phenotypes and their level of implication in the host-parasite molecular cross talk cannot be assessed unless functional studies are carried out.

We also confirmed that these four mimicry candidates originate from the parasite and not from host DNA contamination (see Methods & Additional file 3: Figure S1). Our results showed that each candidate gene produced an amplification product for all parasite DNA templates (coracidia gDNA, adult gDNA, and adult cDNA) but not for fish DNA samples, except *palmitoyltransferase* which only amplified in gDNA with a larger than expected amplicon size (Additional file 3: Figure S1). *Wnt04*, *palmitoyltransferase* and *lysyl oxidase homolog 2B* produced the expected length of PCR products (approximately 190 bp, 180 bp and 170 bp, respectively) for the different types of parasite DNA templates. However, the amplicon for

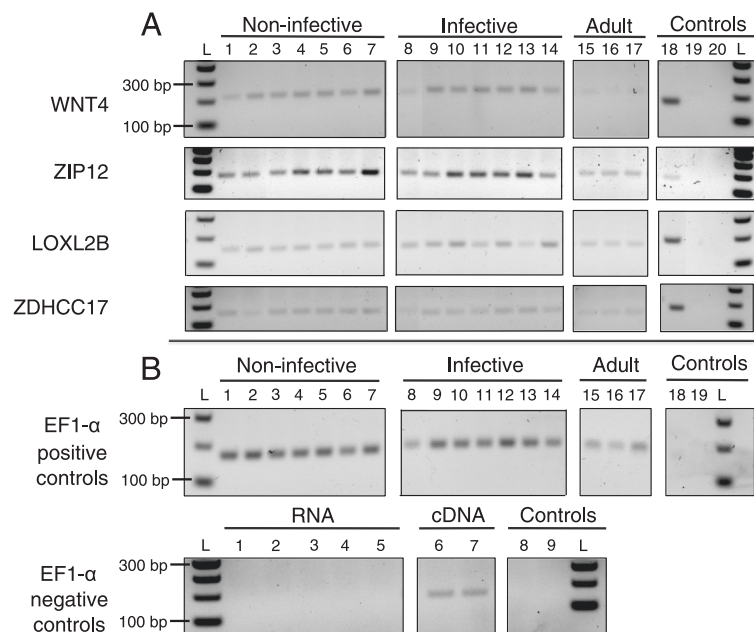


Figure 4 Expression of mimicry candidate genes in three different *S. solidus* developmental stages confirmed by retrotranscription polymerase chain reaction (RT-PCR). RT-PCR was performed on 17 different *S. solidus* worms from three life stages (non-infective, infective and post-reproduction adults) to confirm that the four best mimicry candidate genes, i.e. secreted proteins, are expressed in at least one developmental stage of the parasite. **A**) Gel electrophoresis for each gene, *wnt4* (WNT4), zinc transporter (ZIP12), lysyl oxidase (LOXL2B) and palmitoyltransferase (ZDHCC17). Wells 1–7 contain cDNA from non-infective worms (mass < 50 mg), wells 8–14 contain cDNA from infective worms (mass > 100 mg) and wells 15–17 contain cDNA from adult worms (mass > 350 mg, after egg production outside the fish). Well 18 = positive control with elongation factor 1 alpha (EF1-α), a gene commonly expressed in all tissues and developmental stages. Well 19 = negative control (no template). Well 20 = negative control (no primers). **B**) Additional controls to confirm that our positive control gene (EF1-α) is expressed in all worms and all life stages and to confirm that RNA samples used to perform the RT-PCR reactions are DNA-contamination free. Positive control: non-infective stage (wells 1–7), infective stage (wells 8–14), adult stage (wells 15–17) and controls (well 18 = no template, well 19 = no primers). Negative control: wells 1–7 (RNA samples from all three life stages, i.e. two non-infective, two infective and one adult worms respectively), wells 6–7 (positive controls with cDNA from one infective worm and one non-infective worm respectively), wells 8–9 (negative control, no template and no primers respectively).

membrane zinc transporter was larger in parasite gDNA (approximately 280 bp) than in parasite cDNA samples (160 bp), perhaps because the primer pair used to amplify this gene spans an intron or because of alternative splicing events. By performing this additional control, we were able to confirm the absence of host DNA contamination in our final candidates.

RT-PCR validation

Bearing in mind the importance of validation when it comes to results obtained through complete *in silico* methods, we conducted an additional analysis on parasites bred in the laboratory to confirm that our candidate genes are expressed in different developmental stages (see Methods & Additional file 3). We sampled 17 additional parasites, performed RNA extractions on whole worms, designed and tested primers and performed RT-PCR reactions on all worms. Results indicate that the four candidate genes are expressed in three developmental stages spanning the entire growth period within the fish host and this holds true for all of the additional worms sampled (Figure 4). Such empirical evidence confirms the expression of the candidates in different life stages. However, RT-PCR data is used as a test of presence/absence of expression and does not quantify at which level each gene is expressed in each worm life stage. Small differences in expression levels among developmental stages (i.e. not detectable through the validation method we used) could produce variable phenotypes depending on the background physiological state and the expression levels of other non-studied genes. Secreted proteins identified in this study as potential candidates for molecular mimicry exhibit consistent expression throughout key stages associated with host phenotypic alterations, which prompts the development of follow-up studies.

Conclusion

In this study, we identified mimicry candidate peptides from a behavior-altering cestode that showed high sequence similarity with specific host proteins. Two different *in silico* pipeline analyses were built and used to identify these candidates, which acts as useful analytical tools that can be used in any host-parasite system to perform the same task. The expression of the candidate protein-coding genes in three developmental stages of the parasite was also confirmed by RT-PCR, thus confirming their importance throughout *S. solidus*' life cycle. Candidates identified through these analyses were selected based on sequence similarity only and should not be considered as evidence for any mechanistic link between infection and phenotypic changes in physiology and behavior. Further proteomics and transcriptomics analyses as well as functional assays in different life stages of the parasite and in uninfected fish should help

understand the role of these candidate proteins during the infection of the stickleback and reinforce our knowledge on the molecular bases of complex ecological interactions taking place between a parasite and its host.

Additional files

Additional file 1: Figure S1. CD-HIT-EST parameters used in the *de novo* assembly process. CD-HIT-EST uses an incremental greedy algorithm that sorts sequences in order of decreasing length. The longest sequence becomes the "representative sequence" (R) of the first cluster. All other sequences are then compared to the representative sequence of each cluster. If similarity with the representative sequence is above a certain threshold, the sequence is added to the cluster, otherwise it becomes the representative sequence of a new cluster. In our analysis, the shorter sequences (S) had to cover at least 85% of the length of the representative sequence ($s = 85\%$). Alignment coverage threshold on the representative sequence (aL) was set to 85% and alignment coverage threshold on other sequences in the cluster (aS) was also set to 85%. Sequence similarity threshold was 90%. R_a = portion of the representative sequence (R) that aligns with all other sequences in the cluster. S_a = portion of the shorter sequence (S) that aligns with all other sequences in the cluster. See http://weizhong-lab.ucsd.edu/cd-hit/wiki/doku.php?id=cd-hit_user_guide for further details.

Additional file 2: Figure S2. Ungapped BLAST-p identity thresholds used in this study. Thresholds were empirically determined by Ludin et al. [11] and confirmed in this study after performing several tests. Red line: conserved proteins thresholds, queries showing identities below the line (values in red) were considered as "conserved" and were discarded. Green line: high similarity thresholds, queries showing identities above the line (values in green) were considered as potential mimicry candidates.

Additional file 3: Supplementary methods. Detailed description of the methods used to confirm, by RT-PCR and simple PCR, that the mimicry candidate genes identified in this study are expressed in different life stages of *S. solidus* and that no host contamination can be found in our dataset [77,78].

Additional file 4: Final DNA sequence assembly. Complete set of DNA sequences obtained after performing combined *de novo* assemblies. This cleaned and annotated sequence dataset represents the reference transcriptome used in this study to predict tapeworm proteins and perform the *in silico* mimicry pipelines.

Additional file 5: *Schistocephalus solidus* predicted proteome. Complete set of ORFs predicted from our reference transcriptome. Mimicry protein and peptide identification was performed on this dataset.

Additional file 6: Table S1. 14-mer mimicry candidates. Complete set of 14-mer mimicry candidates as identified through pipeline B.

Additional file 7: Secreted Protein/peptide mimicry candidates. Protein and DNA sequences for the four final candidates labeled as secreted proteins and identified through pipeline B.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FOH designed the mimicry project with input from NAH and CRL. LP, IS, MP and MK performed the transcriptome experiment and contributed samples. IS performed the gDNA PCR used for contamination control. FOH performed the RT-PCR experiment. FOH performed the bioinformatics analyses. FOH drafted the manuscript, FOH, LP, IS, MP, SG, IB, MK, CRL and NAH revised and contributed to the final manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Eric Normandeau and Scott Pavey for their insightful comments on earlier versions of the manuscript. This work was funded by a FRQ-NT Project de Recherche en Equipe grant to NAH and CRL and a Natural Science

and Engineering Research Council of Canada (NSERC) Discovery grant to NAH. FOH would like to thank NSERC for its financial support through the Vanier Canada Graduate Scholarship and the University of Leicester for its support and assistance throughout the lab sampling process. IS, LP, MP and MK would like to thank M Milinski for his continued support and encouragement to their projects and to the German Science Foundation (DFG, grant #KA 2970/1-2) for funding. SG would like to thank UK BBSRC for its financial support via MITBP. CRL was a CIHR New Investigator during this project.

Author details

¹Institut de Biologie Intégrative et des Systèmes (IBIS), Département de Biologie, Université Laval, Pavillon Charles-Eugènes-Marchand, Québec G1V 0A6, Canada. ²Department of Evolutionary Ecology, Max Planck Institute for Evolutionary Biology, August-Thienemann-Str 2, 24306 Ploen, Germany. ³Department of Biology, Adrian Building, Leicester University, University Road, Leicester LE1 7RH, UK.

Received: 10 March 2015 Accepted: 29 March 2015

Published online: 15 April 2015

References

- Dawkins R, Krebs JR. Arms races between and within species. *Proc R Soc Lond B Biol Sci.* 1979;205:489–511.
- Morris SC. Parasites and the fossil record. *Parasitology.* 1981;82:489.
- Klompmaaker AA, Artal P, van Bakel BWM, Fraaije RHB, Jagt JWM. Parasites in the fossil record: a Cretaceous fauna with isopod-infested decapod crustaceans, infestation patterns through time, and a new ichnotaxon. *Plos One.* 2014;9:e92551.
- Poulin R. *Evolutionary ecology of parasites.* 2nd ed. Princeton: Princeton University Press; 2011.
- Poulin R. Parasite manipulation of host behavior: an update and frequently asked questions. In: Brockmann JH, editor. *Advances in the study of behavior*, vol. 41. Burlington: Academic; 2010. p. 151–86 [Advances in the Study of Behavior].
- Vyas A, Kim S-K, Giacomini N, Boothroyd JC, Sapolsky RM. Behavioral changes induced by *Toxoplasma* infection of rodents are highly specific to aversion of cat odors. *Proc Natl Acad Sci U S A.* 2007;104:6442–7.
- House PK, Vyas A, Sapolsky R. Predator cat odors activate sexual arousal pathways in brains of *Toxoplasma gondii* infected rats. *Plos One.* 2011;6:e23277.
- Hébert FO, Aubin-Horth N. Ecological genomics of host behavior manipulation by parasites. *Adv Exp Med Biol.* 2014;781:169–90.
- Prandovszky E, Gaskell E, Martin H, Dubey JP, Webster JP, McConkey GA. The neurotropic parasite *Toxoplasma gondii* increases dopamine metabolism. *Plos One.* 2011;6:e23866.
- Damian RT. Molecular mimicry: antigen sharing by parasite and host and its consequences. *Am Nat.* 1964;98:129–49.
- Ludin P, Nilsson D, Mäser P. Genome-wide identification of molecular mimicry candidates in parasites. *Plos One.* 2011;6:e17546.
- Damian RT. A theory of immunoselection for eclipsed antigens of parasites and its implications for the problem of antigenic polymorphism in man. *J Parasitol.* 1962;48:16.
- Ouaissi MA, Afchain D, Capron A, Grimaud JA. Fibronectin receptors on *Trypanosoma cruzi* trypomastigotes and their biological function. *Nature.* 1984;308:380–2.
- Ouaissi MA, Cornette J, Afchain D, Capron A, Gras-Masse H, Tartar A. *Trypanosoma cruzi* infection inhibited by peptides modeled from a fibronectin cell attachment domain. *Science.* 1986;234:603–7.
- Baruch DI, Gormely JA, Ma C, Howard RJ, Pasloske BL. *Plasmodium falciparum* erythrocyte membrane protein 1 is a parasitized erythrocyte receptor for adherence to CD36, thrombospondin, and intercellular adhesion molecule 1. *Proc Natl Acad Sci U S A.* 1996;93:3497–502.
- Howell DP-G, Levin EA, Springer AL, Kraemer SM, Phippard DJ, Schief WR, et al. Mapping a common interaction site used by *Plasmodium falciparum* Duffy binding-like domains to bind diverse host receptors. *Mol Microbiol.* 2008;67:78–87.
- Hide G, Gray A, Harrison CM, Tait A. Identification of an epidermal growth factor receptor homologue in trypanosomes. *Mol Biochem Parasitol.* 1989;36:51–9.
- Ghansah TJ, Ager EC, Freeman-Junior P, Villalta F, Lima MF. Epidermal growth factor binds to a receptor on *Trypanosoma cruzi* amastigotes inducing signal transduction events and cell proliferation. *J Eukaryot Microbiol.* 2002;49:383–90.
- Spiliotis M, Kroner A, Brehm K. Identification, molecular characterization and expression of the gene encoding the epidermal growth factor receptor orthologue from the fox-tapeworm *Echinococcus multilocularis*. *Gene.* 2003;323:57–65.
- Vicogne J, Cailliau K, Tulasne D, Browaeys E, Yan YT, Fafeur V, et al. Conservation of epidermal growth factor receptor function in the human parasitic helminth *Schistosoma mansoni*. *J Biol Chem.* 2004;279:37407–14.
- Biron DG, Marche L, Ponton F, Loxdale HD, Galeotti N, Renault L, et al. Behavioural manipulation in a grasshopper harbouring hairworm: a proteomics approach. *Proc Biol Sci.* 2005;272:2117–26.
- Biron DG, Ponton F, Marche L, Galeotti N, Renault L, Demey-Thomas E, et al. "Suicide" of crickets harbouring hairworms: a proteomics investigation. *Insect Mol Biol.* 2006;15:731–42.
- Biron DG, Loxdale HD. Host-parasite molecular cross-talk during the manipulative process of a host by its parasite. *J Exp Biol.* 2013;216:148–60.
- Barber I, Schar sack JP. The three-spined stickleback- *Schistocephalus solidus* system: an experimental model for investigating host-parasite interactions in fish. *Parasitology.* 2010;137:411.
- Wedekind C, Milinski M. Do three-spined sticklebacks avoid consuming copepods, the first intermediate host of *Schistocephalus solidus*? - an experimental analysis of behavioural resistance. *Parasitology.* 1996;112:371–83.
- Jakobsen PJ, Wedekind C. Copepod reaction to odor stimuli influenced by cestode infection. *Behav Ecol.* 1998;9:414–8.
- Hammerschmidt K, Koch K, Milinski M, Chubb JC, Parker GA. When to go: optimization of host switching in parasites with complex life cycles. *Evolution.* 2009;63:1976–86.
- Bråten T. Host specificity in *Schistocephalus solidus*. *Parasitology.* 1966;56:657–64.
- Hammerschmidt K, Kurtz J. *Schistocephalus solidus*: establishment of tapeworms in sticklebacks – fast food or fast lane? *Exp Parasitol.* 2007;116:142–9.
- Clarke AS. Studies on the life cycle of the pseudophyllidean cestode *Schistocephalus solidus*. *Proc Zool Soc London.* 1954;124:257–302.
- Barber I. Sticklebacks as model hosts in ecological and evolutionary parasitology. *Trends Parasitol.* 2013;29:556–66.
- Schar sack JP, Kalbe M, Derner R, Kurtz J, Milinski M. Modulation of granulocyte responses in three-spined sticklebacks *Gasterosteus aculeatus* infected with the tapeworm *Schistocephalus solidus*. *Dis Aquat Org.* 2004;59:141–50.
- Schar sack JP, Gossens A, Franke F, Kurtz J. Excretory products of the cestode, *Schistocephalus solidus*, modulate in vitro responses of leukocytes from its specific host, the three-spined stickleback (*Gasterosteus aculeatus*). *Fish Shellfish Immun.* 2013;35:1779–87.
- Milinski M. Parasites determine a predator's optimal feeding strategy. *Behav Ecol Sociobiol.* 1984;15:35–7.
- Barber I, Ruxton G. Temporal prey distribution affects the competitive ability of parasitized sticklebacks. *Anim Behav.* 1998;56:1477–83.
- Tierney JF, Crompton DW. Infectivity of plerocercoids of *Schistocephalus solidus* (Cestoda: Ligulidae) and fecundity of the adults in an experimental definitive host, *Gallus gallus*. *J Parasitol.* 1992;78:1049–54.
- Barber I, Walker P, Svensson PA. Behavioural responses to simulated avian predation in female three spined sticklebacks: the effect of experimental *Schistocephalus solidus* infections. *Behaviour.* 2004;141:1425–40.
- Barber I, Downey LC, Braithwaite VA. Parasitism, oddity and the mechanism of shoal choice. *J Fish Biol.* 1998;53:1365–8.
- Giles N. Behavioural effects of the parasite *Schistocephalus solidus* (Cestoda) on an intermediate host, the three-spined stickleback, *Gasterosteus aculeatus* L. *Anim Behav.* 1983;31:1192–4.
- Milinski M. Risk of predation of parasitized sticklebacks (*Gasterosteus aculeatus* L.) under competition for food. *Behaviour.* 1985;93:203–16.
- Godin J-GJ, Sproul CD. Risk taking in parasitized sticklebacks under threat of predation: effects of energetic need and food availability. *Can J Zool.* 1988;66:2360–7.
- Ness JH, Foster SA. Parasite-associated phenotype modifications in threespine stickleback. *Oikos.* 1999;85:127–34.
- Overli O, Pall M, Borg B, Jobling M, Winberg S. Effects of *Schistocephalus solidus* infection on brain monoaminergic activity in female three-spined sticklebacks *Gasterosteus aculeatus*. *P R Soc B.* 2001;268:1411–5.

44. Hammerschmidt K, Kurtz J. Ecological immunology of a tapeworms' interaction with its two consecutive hosts. In: Webster JP, editor. *Advances in parasitology*, vol. 68. London: Elsevier; 2009. p. 111–37 [Advances in Parasitology].
45. Blanca JM, Pascual L, Ziarsolo P, Nuez F, Canizares J. ngs_backbone: a pipeline for read cleaning, mapping and SNP calling using Next Generation Sequence. *BMC Genomics*. 2011;12:285.
46. Chevreur B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB)*. 1999;99:45–56.
47. Boisvert S, Laviolette F, Corbeil J. Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. *J Comput Biol*. 2010;17:1519–33.
48. Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PGD. Evaluating characteristics of De Novo assembly software on 454 transcriptome data: a simulation approach. *Plos One*. 2012;7:e31410.
49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
50. Surget-Groba Y, Montoya-Burgos JI. Optimization of de novo transcriptome assembly from next-generation sequencing data. *Genome Res*. 2010;20:1432–40.
51. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.
52. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2.
53. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010;26:680–2.
54. Tsai IJ, Zarowiecki M, Holroyd N, Garcarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*. 2013;496:57–63.
55. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet*. 2000;16:276–7.
56. Fontanillas E, Welch JJ, Thomas JA, Bromham L. The influence of body size and net diversification rate on molecular evolution during the radiation of animal phyla. *BMC Evol Biol*. 2007;7:95.
57. Stebbins CE, Galán JE. Structural mimicry in bacterial virulence. *Nature*. 2001;412:701–5.
58. Guan KL, Dixon JE. Protein tyrosine phosphatase activity of an essential virulence determinant in *Yersinia*. *Science*. 1990;249:553–6.
59. Stuckey JA, Schubert HL, Fauman EB, Zhang ZY, Dixon JE, Saper MA. Crystal structure of *Yersinia* protein tyrosine phosphatase at 2.5 Å and the complex with tungstate. *Nature*. 1994;370:571–5.
60. Kaniga K, Uralil J, Bliska JB, Galán JE. A secreted protein tyrosine phosphatase with modular effector domains in the bacterial pathogen *Salmonella typhimurium*. *Mol Microbiol*. 1996;21:633–41.
61. Stebbins CE, Galán JE. Modulation of host signaling by a bacterial mimic: structure of the *Salmonella* effector SptP bound to Rac1. *Mol Cell*. 2000;6:1449–60.
62. Käll L, Krogh A, Sonnhammer ELL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol*. 2004;338:1027–36.
63. Meyer A, Biermann CH, Orti G. The phylogenetic position of the Zebrafish (*Danio rerio*), a model system in developmental biology: an invitation to the comparative method. *Proc Biol Sci*. 1993;252:231–6.
64. Broughton RE, Betancur RR, Li C, Arratia G, Orti G. Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. *PLoS Curr Tree Life*. 2013. doi: 10.1371/currents.tol.2ca8041495ffaf0c92756e75247483e.
65. Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nat Genet*. 2014;46:1212–9.
66. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, et al. GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol*. 2003;4:R28.
67. Petersen TN, Brunak S, Heijne von G, Nielsen H. Correspondence. *Nat Methods*. 2011;8:785–6.
68. Ponton F, Lefevre T, Lebarbenchon C, Thomas F, Loxdale HD, Marche L, et al. Do distantly related parasites rely on the same proximate factors to alter the behaviour of their hosts? *Proc Biol Sci*. 2006;273:2869–77.
69. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–60.
70. Tsonis PA, Dwivedi B. Molecular mimicry: structural camouflage of proteins and nucleic acids. *Biochim Biophys Acta*. 2008;1783:177–87.
71. Cadigan KM, Nusse R. Wnt signaling: a common theme in animal development. *Genes Dev*. 1997;11:3286–305.
72. Adamo SA. Parasites: evolution's neurobiologists. *J Exp Biol*. 2013;216:3–10.
73. Bly M. Examination of the zinc transporter gene, SLC39A12. *Schizophr Res*. 2006;81:321–2.
74. El-Husseini AE-D, Bredt DS. Protein palmitoylation: a regulator of neuronal development and function. *Nat Rev Neurosci*. 2002;3:791–802.
75. Fukata M, Fukata Y, Adesnik H, Nicoll RA, Bredt DS. Identification of PSD-95 palmitoylating enzymes. *Neuron*. 2004;44:987–96.
76. El-Husseini AE-D, Schnell E, Dakoji S, Sweeney N, Zhou Q, Prange O, et al. Synaptic strength regulated by palmitate cycling on PSD-95. *Cell*. 2002;108:849–63.
77. Smyth DJ. Studies on tapeworm physiology, the cultivation of *Schistocephalus solidus* in vitro. *J Exp Biol*. 1946;23:47–70.
78. Espínola SM, Ferreira HB, Zaha A. Validation of suitable reference genes for expression normalization in *Echinococcus* spp. larval stages. *PLoS One*. 2014;9:e102228.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

