**Elucidating the kinase repertoire of *Fasciola hepatica* for discovery of putative drug targets**

Edwin K. Murungi, Rosaline Macharia and Henry M. Kariithi

**Abstract**

The trematode *Fasciola hepatica* causes fasciolosis in humans and livestock worldwide resulting in estimated annual global agricultural losses of US $ 3 billion. Treatment of fasciolosis is constrained and largely relies on triclabendazole to which resistance has been reported thus compromising fluke control. Development of new drugs for fasciolosis is therefore imperative. Although kinases play a vital role in diverse biological processes such as metabolism, cell proliferation, apoptosis and differentiation, and represent attractive targets for novel approaches to combat fasciolosis, there is a paucity of information regarding them in *F. hepatica*. We therefore performed a genome-wide analysis to elucidate the kinome of *F. hepatica*. Here, we identified, for the first time, the entire complement of protein kinases (PKs) encoded in *Fasciola hepatica* through sequence similarity searches using Hidden Markov Models (HMMs). Further, phylogenetic reconstruction was undertaken using PhyML and MrBAYES. To identify homologs with approved drugs, the identified kinases were threaded against DrugBank database. Three dimensional models of select kinases homologous to validated drug targets were generated using I-TASSER and docking undertaken using AutoDock to determine binding affinities. We have identified 203 putative PKs that cluster within AGC, CMGC, CAMK, CK1, TK, TKL, STE, OTHER and unclassified groups. Phylogenetic analysis of each of each of these groups reveals a considerable phylogenetic closeness to *Schistosoma haematobium* kinases pointing to plausible similarities in signaling mechanisms between these worms. By defining and curating the kinome of *F. hepatica,* we provide a vital resource that forms the basis of further experimental investigations. Moreover, it provides a foundation for elaborate fundamental explorations into the signaling pathways involved in critical developmental and physiological processes in this trematode.

**Introduction**

The trematode parasite *Fasciola hepatica* globally causes liver fluke disease (fasciolosis) in livestock and humans (Keiser and Utzinger, 2009; Mas-Coma et al., 2005). Productivity losses associated with infection results in tremendous economic losses to the global farming industry. Moreover, WHO has identified fasciolosis as a re-emerging neglected tropical disease associated with endemic and epidemic outbreaks of disease in human populations (WHO, 2012) with rising prevalence due to increased antihelminthic resistance, climate change, increased animal movement and changing farming practices (van Dijk et al., 2012). The World Health Organization estimates that at least 2.4 million people in more than 70 countries are affected by fascioliasis (WHO, 2015). Production losses as a consequence of the disease in domesticated animals play a major role in extending the cycle of poverty in developing countries and imperil food security in developed countries (Piedrafita *et al.,* 2010). Effective control of fasciolosis in both humans and livestock relies on treatment with the benzimidazole derivative triclabendazole (Fairweather, 2005). However, overuse of triclabendazole especially during targeted or mass administration campaigns (Keiser *et al.,* 2005) has led to the development of resistance (Brennan et al., 2007) that may very likely compromise future disease control. Hybridization of *Fasciola* species may further render the agent ineffective. As such, it is imperative to identify and validate new drug targets in *F. hepatica* for the design of alternative chemotherapeutic compounds against fasciolosis.

The life cycle of F. hepatica is complex and involves lymnaeid snails, e.g. Galba truncatula, as intermediate hosts (Cwiklinski et al., 2015b). Undifferentiated fluke eggs are released in the faeces of infected animals by adult flukes in the bile ducts (Robinson and Dalton, 2009). Once washed out of the faeces, the eggs hatch and release free-swimming miracidia that penetrate the snail. Within the infected snail, the fluke multiplies and releases cercariae that encyst on the vegetation to form infective metacercariae. Following ingestion by grazing animals and humans, the metacercariae emerge from their cysts in the intestine as newly excysted juveniles (NEJ), penetrate the intestinal wall and invade the liver causing perforation and haemorrhage (Cwiklinski et al., 2015b). After about 8 weeks in the liver, the parasites then translocate to the bile ducts where they mature and produce fluke eggs that are subsequently released in faeces to initiate another cycle of infection (Molina-Hernandez et al., 2015). The hepatic damage caused by the parasites results in reduced animal performance, fertility and milk production (Charlier et al., 2014). Moreover, F. hepatica modulates the host’s immune system and secretes potent immunosuppressive molecules that depress host immunity and promote co-infection with Salmonella spp and Clostridia spp (Dalton et al., 2013).

The significant negative impacts on human health and agriculture, particularly the increasing demand for animal‐derived food products for the sustenance of an increasing global population highlight the necessity of effectively controlling fasciolosis (Cwiklinski et al., 2016). For this objective to be realized, the development of novel chemotherapeutic agents for fasciolosis to augment vaccine development efforts is paramount. Kinases regulate complex cellular processes such as metabolism, cell-cycle progression, apoptosis, cytoskeletal modifications and differentiation (Grevelding et al., 2017; Manning et al., 2002; Kannan et al., 2007) by acting as phosphotransferases that catalyze the transfer of phosphate groups from ATP to specific substrates such as nucleotides, carbohydrates, lipids, amines, vitamins and proteins. Protein kinases (PKs) are classified into protein tyrosine kinases (PTKs), serine/threonine kinases (STKs), dual specificity PKs, histidine PKs and arginine PKs (Grevelding et al., 2017). All these are commonly referred to as eukaryotic protein kinases (ePKs; Hanks and Hunter, 1995) and share a conserved three-dimensional fold that is separated into two subdomains or lobes (Zheng et al., 1993). The N-terminal lobe is composed of a five-stranded β sheet and one prominent α helix, termed helix αC, while the C-terminal lobe is predominantly helical (Huse and Kuriyan, 2002). Substrate and ATP binding sites are located in the deep cleft between the two lobes (Krupa and Srinivasan, 2002). ATP sits beneath a highly conserved phosphate-binding loop that connects strands β1 and β2 and contains a conserved glycine-rich sequence motif, GXGXφG, where φ is usually tyrosine or phenylalanine. The conserved catalytic domain of ePKs is made up of approximately 250 amino acids and is subdivided into 12 subdomains with highly conserved individual amino acids and motifs (Hanks et al., 1988). Based on sequence similarity in their catalytic domains and the presence of accessory domains that serve adaptor or regulatory functions aiding in substrate recruitment, specificity and scaffolding (Deshmukh et al., 2010), ePKs segregate into 8 main typical groups (Stroehlein et al., 2015) including AGC (c-AMP-dependent PKs/protein kinase G/protein kinase C), CaMK (calcium/Calmodulin-regulated kinases), CMGC (cyclin-dependent kinases and close relatives such as MAPKs, mitogen-activated protein kinases), CK1 (cell kinase 1), RGC (receptor guanylate cyclases), STE (MAP kinase cascade kinases), tyrosine kinases (TK; Src, Syk, Fyn, Fes, Abl) and TKLs (tyrosine kinase-like proteins). PKs that do not fit into any of these groups are classified as the OTHER group (Hanks and Hunter, 1995). Several kinase groups are termed at atypical and include Alpha, PIKK, PHDK, and RIO groups (Taylor et al., 2013). Because of their importance in parasite biology, PKs have emerged as viable targets for drug intervention and are therefore of high value also for translational research (Hunter, 2000; Cohen, 2002). Specific kinase groups have a conserved catalytic mechanism and overall structure (Hanks, 2003) and small molecules have been shown to bind to their catalytic cleft (Zhang et al., 2009). Already, several PK inhibitors have been developed and approved particularly for treatment of cancer (Knight et al., 2010; Hu et al., 2014; Rask-Andersen et al., 2014; Wu *et al.,* 2016).

Despite the availability of draft *F hepatica* genome and proteome (Cwiklinski et al., 2015), curation of the kinome for this trematode has not been undertaken. Therefore, the identification and comparative analysis of *F. hepatica* PKs will not only uncover putative targets amenable to disruption by novel chemotherapeutic agents or repurposed drugs but also facilitate deeper understanding into the functioning of various signaling pathways in this trematode. In the present study, we have identified and curated the protein kinases of *F. hepatica*, deciphered their evolution and diversity, mapped the kinases onto signaling pathways and inferred the binding of several chemicals ligands onto a select panel of kinases.

**Materials and Methods**

**Identification of *Fasciola hepatica* protein kinases**

The draft proteome of *F. hepatica* were retrieved from WormBase Parasite version WBPS9 (WS258) (parasite.wormbase.org/index.html). We used HMMER v3.1b2 (Eddy, 2011) to identify putative *F. hepatica* PKs using a Hidden Markov Model (HMM) library made up of all eukaryotic protein kinase domain definitions obtained from Kinomer (Miranda-Saavedra and Barton, 2007). After identification, the domain architecture of the PK candidates was delineated by comparison with known kinase domains in InterProScan v.5.24-63.0 (Jones et al., 2014) which employs information from domain databases Pfam v.31.0 (Finn et al., 2016), PANTHER v.9.0 (Mi et al., 2013) and SUPERFAMILY v.1.75 (Gough et al., 2001). Sequences with insufficient kinase domain evidence, those lacking critical functional residues particularly the ATP-binding glycine-rich motif and the catalytic aspartate were omitted. Conserved motifs and amino acid residues were identified via sequence alignment. Fragments of putative PKs of less that 200 residues and lacking the functional residues were not considered in this analysis, as they are plausibly pseudogenes. Any sequence containing the VAIK, HRD and DFG motifs was considered to be a kinase (Hanks and Hunter, 1995; Manning et al., 2002). Signal peptides and transmembrane regions were predicted using SignalP4.0 (Petersen et al., 2011) and TMHMM (Krogh et al., 2001), respectively. Subcellular localization for the putative kinases was predicted using TargetP (Emanuelsson et al., 2007).

### Phylogenetic and evolutionary analysis

The amino acid sequences of the putative *F. hepatica* PKs were aligned using MAFFT (Katoh et al., 2002) with default parameters and manually optimized in Jalview 2.9 (Waterhouse et al., 2009). Prior to phylogenetic reconstruction, the best-fit evolutionary model was selected for each alignment using AIC in ProtTest v3 (Darriba et al., 2011). Phylogenetic reconstructions were undertaken using a maximum likelihood approach implemented in PhyML v3.0 program (Guindon et al., 2010) and a Bayesian inference approach implemented in MrBayes v3.2 (Ronquist et al., 2012). For the PhyML analysis, clade support was calculated using SH-like approximate likelihood ratio test, Bayes likelihood test and bootstrap proportions (100 replicates) while for MrBayes, a mixture of models with fixed rate matrices was employed to calculate posterior probabilities. To construct majority rule trees, 1,000,000 trees were generated, of which every 100th tree was sampled after discarding the first 25% of trees as burn-in. Phylogenetic trees were rendered using iTOL (Letunic and Bork, 2016).

**\*\*\*Functional and structural annotation**

We functionally annotated the identified kinases using Blast2GO (Conesa et al., 2005) and assigned then assigned kinase sequences to biochemical pathways based on matches to the KEGG database (Kanehisa and Goto, 2000). Three-dimensional structures of select kinases were predicted using the program I-TASSER v.4.4 (Roy et al., 2010) and structures visualized in PyMOL (Delano, 2002).

**\*\*\*\*Target identification**

**Results and Discussion**

Our survey of the predicted *Fasciola hepatica* proteome (Cwiklinski *et al.,* 2015) using a HMM profile of eukaryotic kinases catalytic domains obtained from Kinomer (Martin *et al.,* 2009) identified 203 putative protein kinase sequences (Table 1) of sufficient length and containing the functionally critical glycine-rich motif or the catalytic aspartate residue (Hanks et al., 1988). This represents XX% of the total proteins encoded in the *F. hepatica* genome. The repertoire of F. hepatica kinases is comparable to that of *S. haematobium* (n = XX) () with average overall amino acid identity value of 35 %. Many of the sequences that constitute the *F. hepatica* kinome cluster within the familiar AGC, CMGC, CK1, TKL and CaMK groups found within the mammalian kinome. Compared to its closely related flukes S. mansoni and S. hematobium, which contain 252 (Andrade et al., 2011) and 269 (Stroehlein et al., 2015) PKs respectively, there is plausibly PKs gene loss in F. hepatica. The most highly populated subfamilies are CMGC, CaMK and AGC containing 44, 39 and 32 proteins respectively.

\*\*\*\*Although several of the PKs contain accessory domains previously identified on kinases, majority lack accessory domains, which points to transcriptional regulation.

***Figure 1***: Phylogenetic tree of the *F. hepatica* AGC kinases. Circular tree of *F. hepatica* and *Schistosoma mansoni* AGCs. To construct the tree, kinase domain regions of each sequence were aligned using MAFFT and the phylogenetic tree then inferred from the resulting alignment using RAxML. The tree image was rendered with the Interactive Tree of Life server (iTOL). A purple circle on a branch indicates bootstrap support greater than 80; larger circles indicate greater bootstrap values.

Several signaling cascades such as mitogen-activated protein kinase/protein kinase C (Mpk1/Pkc1) and cyclic adenosine monophosphate/protein kinase A (cAMP/Pka1) pathways, allow *Cryptococcus neoformans* to sense, respond and adapt to host stresses encountered throughout the course of infection.

**Conclusion and Recommendations**

**References**

1. Keiser J, Utzinger J. Food-borne trematodiasis. *Clin. Microbiol. Rev.,* 2009, 22: 466-483.
2. Mas-Coma S, Bargues MD, Valero MA. Fascioliasis and other plant-borne trematode zoonoses. *Int. J Parastiol.,* 2005, 35: 1255-1278.
3. WHO. Accelerating work to overcome the global impact of neglected tropical diseases; a roadmap for implementation – Executive summary. 2012.
4. Van Dijk J, Sargison ND, Kenyon F, Skuce PJ. Climate change and infectious disease: helminthological challenges to farmed ruminants in temperate regions. Animal, 2010, 4: 377-392.
5. Keiser J, Engels D, Buscher G, Utzinger J. Triclabendazole for the treatment of fascioliasis and paragonimiasis. *Expert Opin. Investig. Drugs*, 2005, 14: 1513-1526.
6. Fairweather I. Triclabendazole: new skills to unravel an old(ish) enigma. J. Helminthol., 2005, 79: 227-234.
7. Robinson MW, Dalton JP. Zoonotic helminth infections with particular emphasis on fasciolosis and other trematodiasis. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 2009, 364: 2763-2776.
8. Piedrafita D, Spithill TW, Smith RE, Raadsma HW. Improving animal and human health through understanding liver fluke immunology. *Parasite Immunol.,* 2010, 32: 572-581.
9. WHO. WHO estimates of the global burden of foodborne diseases. 2015.
10. Cwiklinski K, Allen K, LaCourse J, Williams DJ, Paterson S, Hodgkinson JE. Characterisation of a novel panel of polymorphic microsatellite loci for the liver fluke, Fasciola hepatica, using a next generation sequencing approach. Infection, Genetics and Evolution, 2015, 32: 298-304.
11. Cwiklinski K, Dalton JP, Dufresne PJ, Course JL, Williams DJL, Hodgkinson J, Paterson S. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals adaptation to the host environment and the capacity for rapid evolution. *Genome Biology*, 2015, 16: 71.
12. Cwiklinski K, O’Neill SM, Donnelly S, Dalton JP. A prospective view of animal and human fasciolosis. Parasite Immunol., 2016, 38: 558-568.
13. Brennan GP, Fairweather I, Trudgett A, Hoey E, McCoy, McConville M, Meaney M, Robinson M, McFerran N, Ryan L, Lanusse C, Mottier L, Alvarez L, Solana H, Virkel G, Brophy PM. Understanding triclabendazole resistance. *Exp. Mol. Pathol.,* 2007, 82: 104-109.
14. Wu P, Nielsen TE, Clausen MH. Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. *Drug Discov. Today*, 2017, 21: 5-10.
15. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. InterProScan 5: genome-scale protein function classification. Bioinformatics, 2014, 30: 1236-1240.
16. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Research, 2016, Database Issue 44: D279-D289.
17. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. Nature Protocols, 2013, 8: 1551-1566.
18. Gough J, Karplus K, Hughey R, Chothia C. Assignement of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. J Mol. Biol., 2001, 313: 903-919.
19. Eddy SR. Accelerated profile HMM searches. PLoS Comp. Biol., 2011, 7:e1002195
20. Miranda-Saavedra D, Barton GJ. Classification and functional annotation of eukaryotic protein kinases. Proteins, 2007, 68: 893-914.
21. Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G. Structural and functional diversity of the microbial kinome. PLoS Biol., 2007, 5: e17.
22. Hanks SK, Hunter T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. The FASEB Journal, 1995, 9: 576-596.
23. Rask-Andersen M, Masuram S, Schioth HB. The druggable genome: evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication. Annu. Rev. Pharmacol. Toxicol., 2014, 54: 9-26.
24. Manning G, Plowman GD, Hunter T, Sudarsanam S. Evolution of protein kinase signaling from yeast to man. Trends in Biochemical Sciences, 2002, 27: 514-520.
25. Andrade LF, Nahum LA, Avelar LGA, Silva LL, Zerlotini A, Ruiz JC, Oliveira G. Eukaryotic protein kinases (ePKs) of the helminth parasite Schistosoma mansoni. BMC Genomics, 2011, 12:215.
26. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature Methods, 2011, 8: 785-786.
27. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol. Biol., 2001, 305: 567-580.
28. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols, 2007, 2: 953-971.
29. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res., 2002, 30: 3059-3066.
30. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. Bioinformatics, 2009, 25: 1189-1191.
31. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics, 2011, 27: 1164-1165.
32. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol., 2010, 59: 307-321.
33. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol., 2012, 61: 539-542.
34. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res., 2016, 44(W1): W242-W245.
35. Martin DM, Miranda-Saavedra D, Barton GJ. Kinomer v.1.0: a database of systematically classified eukaryotic protein kinases. Nucleic Acids Res., 2009, 37(Database issue): D244-250.
36. Deshmukh K, Anamika K, Srinivasan N. Evolution of domain combinations in protein kinases and its implications for functional diversity. Progress in Biophysics and Molecular Biology, 2010, 102: 1-15.
37. Hunter T. Signaling – 2000 and beyond. Cell, 2000, 100: 113-127.
38. Cohen P. The origins of protein phosphorylation. Nat. Cell Biol., 2002, 4: E127-E130.
39. Stroehlein AJ, Young ND, Jex AR, Sternberg PW, Tan P, Boag PR, Hoffmann A, Gasser RB. Defining the Schistosoma haematobium kinome enables the prediction of essential kinases as anti-schistosome drug targets. Scientific Reports, 2015, 5:17759.
40. Molina-Hernandez V, Mulcahy G, Perez J, Martinez-Moreno A, Donnelly S, O’Neill SM, Dalton JP, Cwiklinski K. Fasciola hepatica vaccine: we may not be there yet but we’re on the right road. Veterinary Parasitology, 2015, 208: 101-111.
41. Charlier J, Vercruysse J, Morgan E, Van Dijk J. Recent advances in the diagnosis, impact on production and prediction of Fasciola hepatica in cattle. Parasitology, 2014, 141: 326-335.
42. Dalton JP, Robinson MW, Mulcahy G, O’Neill SM, Donnelly S. Immunomodulatory molecules of Fasciola hepatica: Candidates for both vaccine and immunotherapeutic development. Veterinary Parasitology, 2013, 195: 272-285.
43. Knight ZA, Lin H, Shokat KM. Targeting the cancer kinome through polypharmacology. Nat. Rev. Cancer, 2010, 10: 130-137.
44. Hu Y, Furtmann N, Bajorath J. Current compound coverage of the kinome: miniperspective. J. Med. Chem., 2014, 58: 30-40.
45. Hanks SK. Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. Genome Biol., 2003, 4: 111.
46. Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. Nat. Rev. Cancer, 2009, 9: 28-39.
47. Taylor CM, Martin J, Rao RU, Powell K, Abubucker S, Mitreva M. Using existing drugs as leads for broad spectrum anthelmintics targeting protein kinases. PLoS Pathogens, 2013, 9: e1003149-10.
48. Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis of functional genomics research. Bioinformatics, 2005, 21: 3674-3676.
49. Kanehisa M, Goto S. Kyoto encyclopedia of genes and genomes. Nucleic Acids Res., 2000, 28: 27-30.
50. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols, 2010, 5: 539-542.
51. Delano WL. The PyMOL Molecular Graphics System. 2002.
52. Grevelding CG, Langner S, Dissous C. Kinases: Molecular stage directors for Schistosome development and differentiation. Trends in Parasitology, 2017, S1471-4922: 30287-30288.
53. Zheng J, Knighton DR, ten Eyck LF, Karlsson R, Xuong N, Taylor SS, Sowadski JM. Crystal structure of the catalytic subunit of cAMP dependent protein kinase complexed with MgATP and peptide inhibitor. Biochemistry, 1993, 32: 2154-2161.
54. Huse M, Kuriyan J. The conformational plasticity of protein kinases. Cell, 2002, 109: 275-282.
55. Krupa A, Srinivasan N. The repertoire of protein kinases encoded in the draft version of the human genome: atypical variations and uncommon domain combinations. Genome Biology, 2002, 3: research0066.1.
56. Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science, 1988, 241: 42-52.