Profiling G protein-coupled receptors of Fasciola hepatica identifies orphan rhodopsins unique to phylum Platyhelminthes

McVeigh, P., McCammick, E., McCusker, P., Wells, D., Hodgkinson, J. E., Paterson, S., ... Maule, A. (2018). Profiling G protein-coupled receptors of Fasciola hepatica identifies orphan rhodopsins unique to phylum Platyhelminthes. International Journal for Parasitology - drugs & drug resistance, 8(1), 87-103. DOI: 10.1101/207316

Published in:
International Journal for Parasitology - drugs & drug resistance

Document Version:
Publisher's PDF, also known as Version of record

Queen's University Belfast - Research Portal: Link to publication record in Queen's University Belfast Research Portal

Publisher rights
Copyright 2018 the authors.
This is an open access article published under a Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen’s institutional repository that provides access to Queen’s research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person’s rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Download date: 19. Jul. 2018
Profiling G protein-coupled receptors of *Fasciola hepatica* identifies orphan rhodopsins unique to phylum Platyhelminthes

Paul McVeigh a,b*, Erin McCammick a, Paul McCusker a, Duncan Wells a, Jane Hodgkinson b, Steve Paterson c, Angela Mousley a, Nikki J. Marks a, Aaron G. Maule a

a Parasitology & Pathogen Biology, The Institute for Global Food Security, School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast, BT9 7BL, UK
b Institute of Infection and Global Health, University of Liverpool, Liverpool, UK
c Institute of Integrative Biology, University of Liverpool, Liverpool, UK

**ARTICLE INFO**

**Keywords:** Anthelmintic Deorphanization Flukicide Genome Invertebrate Nervous system Neuropeptide RNA-Seq

**ABSTRACT**

G protein-coupled receptors (GPCRs) are established drug targets. Despite their considerable appeal as targets for next-generation anthelmintics, poor understanding of their diversity and function in parasitic helminths has thwarted progress towards GPCR-targeted anti-parasite drugs. This study facilitates GPCR research in the liver fluke, *Fasciola hepatica*, by generating the first profile of GPCRs from the *F. hepatica* genome. Our dataset describes 147 high confidence GPCRs, representing the largest cohort of GPCRs, and the largest set of in silico ligand-receptor predictions, yet reported in any parasitic helminth. All GPCRs fall within the established GRAFS nomenclature; comprising three glutamate, 135 rhodopsin, two adhesion, ligand-receptor predictions, yet reported in any parasitic helminth. These receptors imply flatworm specific rhodopsins. These receptors imply flatworm specific GPCR functions, and/or co-evolution with unique flatworm ligands, and could facilitate the development of exquisitely selective anthelmintics. Ligand binding domain sequence conservation relative to deorphanised rhodopsins enabled high confidence ligand-receptor matching of seventeen receptors activated by acetylcholine, neuropeptide F/Y, octopamine or serotonin. RNA-Seq analyses showed expression of 101 GPCRs across various developmental stages, with the majority expressed most highly in the pathogenic intra-mammalian juvenile parasites. These data identify a broad complement of GPCRs in *F. hepatica*, including rhodopsins likely to have key functions in neuromuscular control and sensory perception, as well as frizzled and adhesion/secretin families implicated, in other species, in growth, development and reproduction. This catalogue of liver fluke GPCRs provides a platform for new avenues into our understanding of flatworm biology and anthelmintic discovery.

1. Introduction

*Fasciola* spp. liver fluke are pathogens of veterinary ruminants that threaten the sustainability of global meat and dairy production. Infection with *Fasciola* (fasciolosis/fascioliasis) inhibits animal productivity through liver condemnation, reduced meat and milk yields, and reduced fertility (for recent impact surveys see Abunna et al. (2010), Sariözkan and YalÇin (2011), Howell et al. (2015) and Habarugira et al. (2016). *Fasciola* spp. also infect humans, with fascioliasis considered a neglected tropical disease (Hotez et al., 2008). Anthelmintic chemotherapy currently carries the burden of fluke control, since there are no liver fluke vaccines (Toet et al., 2014). Six flukicidal active compounds are available for general use, with on-farm resistance reported for all except oxyclozanide (Kelley et al., 2016). Resistance to the frontline flukicide, triclabendazole, also exists in human *F. hepatica* infections (Winkelhagen et al., 2012; Cabada et al., 2016). Given the absence of alternative control methods, new flukicides are essential for secure future treatment of veterinary and medical liver fluke infections.

The helminth neuromuscular system is a prime source of molecular targets for new anthelmintics (Martin and Robertson, 2010; McVeigh et al., 2012; Ribeiro and Patocka, 2013), not least because many

* Corresponding author.

E-mail address: paul.mcveigh@qub.ac.uk (P. McVeigh).

https://doi.org/10.1016/j.ijpddr.2018.01.001

Received 2 November 2017; Received in revised form 10 January 2018; Accepted 12 January 2018

Available online 05 February 2018

2211-3207/ © 2018 The Authors. Published by Elsevier Ltd on behalf of Australian Society for Parasitology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/BY/4.0/).
existing anthelmintics (dichlorvos, levamisole, morantel, piperazine, pyrantel, macrocyclic lactones, paraherquamidine, amino acid derivatives) act upon receptors or enzymes associated with classical neurotransmission in nematodes (Wolstenholme, 2011; McVeigh et al., 2012). G protein-coupled receptors (GPCRs) that transduce signals from both peptidergic and classical neurotransmitters are of broad importance to helminth neurotransmission in nematodes (Wolstenholme, 2011; McVeigh et al., 2017). GPCRs are druggable targets, since 33% of human prescription medicines have a GPCR-based mode of action (Santos et al., 2017).

Despite two F. hepatica genomes (Cwiklinski et al., 2015; McNulty et al., 2017), no GPCR sequences have been reported from F. hepatica. In contrast, GPCRs have been profiled in the genomes of trematodes (Schistosoma mansoni and Schistosoma haematobium (Zamanian et al., 2011; Campos et al., 2014)), cestodes (Echinococcus multilocularis, E. granulosus, Taenia solium and Hymenolepis microstoma (Tsai et al., 2013)), and planaria (Schmidtea mediterranea, Girardia tigrina (Omar et al., 2007; Zamanian et al., 2011; Saberi et al., 2016)). These datasets illustrated clear differences in the GPCR complements of individual flatworm classes and species, with reduced complements in parasitic flatworms compared to planarians.

This study profiles the GPCR complement of the temperate liver fluke F. hepatica for the first time, permitting comparisons with previously characterised species that inform evolutionary and functionally conserved elements of flatworm GPCR signalling. We have identified and classified 147 GPCRs by GRAFS family (glutamate, rhodopsin, adhesion, frizzled, secretin) assignment (Fredriksson et al., 2003), the majority of which are expressed in Fasciola RNA-Seq datasets. These include clear orthologues of GPCRs activated by known neurotransmitters, within which we performed the deepest in silico ligand-receptor matching analyses to date for any parasitic helminth. The latter predicted ligands for 17 F. hepatica GPCRs, designating these as primary targets for deorphanisation. Intriguingly, the dataset included a set of flatworm-expanded GPCRs lacking orthologues outside of phylum Platyhelminthes. Evolution of such GPCRs across the parasitic flatworm classes may have been driven by flatworm-specific functional requirements or co-evolution with flatworm ligands, either of which could help support novel anthelmintic discovery. This dataset provides the first description of GPCRs in liver fluke, laying a foundation for future advances in GPCR-directed functional genomics and flukicide discovery.

2. Materials and methods
2.1. Liver fluke sequence databases

We exploited two F. hepatica genome assemblies available from WormBase ParaSite (Howe et al., 2017), generated by Liverpool University (http://parasite.wormbase.org/Fasciola_hepatica_prjeb6687/Info/Index/(Cwiklinski et al., 2015), and Washington University, St Louis (http://parasite.wormbase.org/Fasciola_hepatica_prjna179522/Info/Index/(McNulty et al., 2017).

2.2. Identification of GPCR-like sequences from F. hepatica

Fig. 1 summarises our GPCR discovery methodology, which employed Hidden Markov Models (HMMs) constructed from protein multiple sequence alignments (MSAs) of previously described S. mansoni and S. mediterranea GPCR sequences (Zamanian et al., 2011). Individual HMMs were constructed for each GRAFS family (Fredriksson et al., 2003). Alignments were generated in Mega v7 (www.megasoftware.net) (Kumar et al., 2016) using the Muscle algorithm with default parameters. HMMER v3 (http://hmmer.org) was employed to construct family-specific HMMs (hmmbuild) from alignments and these were searched (hmmsearch) against a predicted protein dataset from F. hepatica genome PRJEB6687 consisting of 33,454 sequences (Cwiklinski et al., 2015); default parameters were used for hmmsearch and hmmbuild. Returned sequences were filtered for duplicates and ordered relative to the hmmsearch scoring system, enabling the classification of hits according to the GRAFS family to which they showed most similarity (i.e. highest score, lowest E value). All remaining returns were then used as BLAST queries (BLASTp and tBLASTn with default parameters) to identify matching, or additional, sequences originating from the PRJEB6687 and PRJNA179522 genomes (Fig. 1). Where sequences appeared in both genomes, we kept the longest annotated sequence (S1 Table).

2.3. GPCR annotation

Sequences resulting from HMM searches were filtered by transmembrane (TM) domain composition, using hmmtop (http://www.sacs.ucsf.edu/cgi-bin/hmmtop.py) (Tuesday and Simon, 1998, 2001). Sequences containing ≥ 4 TMs were analysed as described below.

2.3.1. Homology analyses

All GPCRs were used as BLASTp (Altschul et al., 1990) queries, to identify their closest (highest scoring) match in the ncbi non-redundant (nr) protein sequence dataset (https://blast.ncbi.nlm.nih.gov/Blast.cgi), with default settings and the “Organism” field set to exclude Platyhelminthes (taxid: 6157). All GPCRs were additionally searched against more phylogenetically limited datasets, by using the “Organism” field to limit the BLASTp searches to: (i) Basal phyla, Ctenophora (taxid:10197), Porifera (taxid:6040), Placozoa (taxid:10226), Cnidaria (taxid:6073); (ii) Superphylum Lophotrochozoa (taxid: 1206795), excluding phylum Platyhelminthes (taxid: 6157); (iii) Superphylum Edyszoa (taxid: 1206794); (iv) Superphylum Deuterostomia (taxid: 33511). For BLASTp searches against other flatworms, we performed local BLAST+ (Camacho et al., 2008) on the WBPS9 release of WormBase Parasite, which included predicted protein datasets from 30 flatworm species. In all cases, we recorded the single highest scoring hit, or recorded “no significant similarity found” in cases where no hits were returned (Table S1); sequences generating both GPCR hits and “no significant similarity” were retained. Where the top hit was not to a GPCR, that sequence was removed from the dataset.

2.3.2. Domain composition

GPCR identities were confirmed using InterProScan Sequence Search (www.ebi.ac.uk/interpro/search/sequence-search) (Jones et al., 2014) and/or HMMER HMMScan (www.ebi.ac.uk/Tools/hmmer/search/hmmscan) (Finn et al., 2015), with default parameters. Again, sequences returning non-GPCR domains were omitted from the dataset, with all others retained.

2.3.3. Motif identification

As an additional measure of confidence in our identifications, we analysed the presence/absence of key motifs diagnostic of receptor families and subfamilies. These analyses were performed for rhodopsins generally, the ligand binding domains (LBDs) of rhodopsin receptors for acetylcholine (ACh), neuropeptide F/Y (NPF/Y), octopamine and serotonin (5-hydroxytryptamine, 5HT), and for the LBDs of glutamate and frizzled/smoothed families. Motifs were identified via protein multiple sequence alignment (MSA) of GPCRs, performed in MAFFT (www.mafft.cbrc.jp/alignment/server) (Katoh et al., 2017), using “E-INS-i” parameters, for sequences with multiple conserved domains. Only identical amino acids were accepted at each site, with conservation expressed as % identity across all sites. Motif illustrations were generated using WebLogo 3 (http://weblogo.threeulomse.com) (Crooks et al., 2004).
Fig. 1. Methods for discovery and annotation of *Fasciola hepatica* G protein coupled receptors (FhGPCRs). (A) Hidden Markov Models (HMMs) representing glutamate, rhodopsin, adhesion, frizzled/smoothened and secretin families, and two rhodopsin subfamilies, were built from protein multiple sequence alignments of *Schistosoma mansoni* and *Schmidtea mediterranea* GPCRs. HMMs were built and searched respectively using the *hmmbuild* and *hmmsearch* modules of HMMER v3.0. Searches were performed against two publically available *F. hepatica* genomes using *hmmsearch* and BLAST tools. Each putative FhGPCR sequence was assessed for transmembrane (TM) domain composition with *hmmtop* before classification using tools including BLASTp, Interproscan and CLANS. (B) The largest proportion (49%) of FhGPCRs carried the full complement of 7 TMs, with 88% of sequences bearing at least 4 TMs. (C) GRAFS composition of 147 FhGPCRs carrying ≥4 TMs. (D) Rhodopsins were subject to further classification, including BLASTp vs datasets representing major non-flatworm animal phyla and superphyla. These rhodopsin homology classifications fed back into phylogenetic analyses versus deorphanised bilaterian GPCRs to confirm their putative ligand selectivity, with a final analysis of ligand binding domain composition comparing conservation of ligand interacting residues for characterised GPCRs reported in the literature with our *F. hepatica* assignments.
Table 1 Comparison of the Fasciola hepatica G-protein coupled receptor (GPCR) complement with those reported from other flatworms. Species complements are shown in the context of GRAFS nomenclature (Fredriksson et al., 2003). *Zamanian et al. (2011); **Campos et al. (2014); †Tsai et al., 2013; ‡Saberi et al., 2016. Saberi et al. (2016) described 566 GPCRs in Schmidtea mediterranea, of which 516 fall within GRAFS nomenclature.

| F. hepatica | S. mansoni * | S. mansoni b | S. haematobium b | E. multilocularis c | S. mediterranea a | S. mediterranea d |
|-------------|-------------|-------------|-----------------|-------------------|------------------|------------------|
| Glutamate   | 3           | 2           | 2               | 2                 | 5                | 11               |
| Rhodopsin   | 135         | 105         | 59              | 53                | 48               | 418              |
| Adhesion    | 2           | 3           | –               | 4                 | 4                | 9                |
| Frizzled/Smoothered | 6 | 5 | 4 | 4 | 5 | 3 | 11 | 10 |
| Secretin    | 1           | 2           | 5               | 5                 | 1                | 1                |
| Total       | 147         | 117         | 64              | 64                | 83               | 448              |
|             |             |             |                 |                   | 516*             |                  |

2.3.4. Phylogenetic reconstruction

Maximum likelihood (ML) phylogenetic trees were constructed using PhyML (http://www.phylogeny.fr) (Dereeper et al., 2008), from protein MSA generated in MAFFT (www.maff.brc.jp/alignment/server/). Alignments were manually edited (in Mega v7) to include only TM domains, by removing extra-membrane blocks aligned with human glutamate, rhodopsin, adhesion or frizzled proteins. Trees were constructed from these TM-focused alignments in PhyML using default parameters, with branch support assessment using the approximate likelihood ratio test (aLRT), under “SH-like” parameters. Trees, exported from PhyML in newick format were drawn and annotated in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

2.4. RNA-seq analyses

Expression of F. hepatica GPCRs was investigated in publically available and in-house generated RNA-Seq datasets. These included developmentally staged Illumina transcriptome reads associated with the Cwiklinski et al. (2015) F. hepatica genome (reads accessed from the European Nucleotide Archive at http://www.ebi.ac.uk/ena/data/search?query=PRJEB6687). These samples originated from distinct developmental stages of US Pacific Northwest Wild Strain F. hepatica (Baldwin Aquatics), including egg (n = 2), metacercariae (met; n = 4), in vitro NEJs 1 h post-ecystment (NEJ1h; n = 1), in vitro NEJs 3 h post-ecystment (NEJ3h; n = 2), in vitro NEJs 24 h post-ecystment (NEJ24h; n = 2), ex vivo–live stage juveniles (juv1; n = 1) and ex vivo adult parasites (Ad; n = 3). Our in-house datasets were generated from ex vivo live stage F. hepatica juveniles (Italian strain, Ridgeway Research Ltd, UK), recovered from rat (Sprague Dawley) hosts at 21 days following oral administration of metacercariae (juv2; n = 3). All animal use was approved by Queen’s University Belfast’s Animal Welfare and Ethical Review Body, and performed under Home Office project license PPL2764.

Total RNA, extracted with Trizol (ThermoFisher Scientific) from each of the 3 independent biological replicates, was quantified and quality checked on an Agilent Bioanalyzer, converted into paired-end sequencing libraries and sequenced on an Illumina HiSeq2000 by the Centre for Genomic Research at the University of Liverpool, UK. RNA samples were spiked prior to library construction with the ERCC RNA Spike-In Mix (ThermoFisher Scientific) (Jiang et al., 2011). All read samples were analysed using the TopHat, Cufflinks, Cuffmerge, Cuffdiff pipeline with default parameters, (Langmead et al., 2009; Trapnell et al., 2009, 2010, 2012a, 2012b; Roberts et al., 2011), with mapping against PRJEB6687 genome sequence and annotation files (accessed from WormBase Parasite; http://parasite.wormbase.org/ftp.html). Data were expressed as number of fragments mapped per million mapped reads per kilobase of exon model FPKM). In juv2 datasets we discarded GPCRs represented by fewer than 0.5 FPKM (the minimum linear sensitivity that we detected with our ERCC spike in); for the staged datasets, we included only receptors represented by ≥ 0.5 FPKM in at least one life stage. Heatmaps were generated with heatmapper (http://www.heatmapper.ca/) (Babicki et al., 2016) set for Average Linkage, and Pearson Distance Measurement.

3. Results and discussion

3.1. A first look at GPCRs in the F. hepatica genome

This study represents the first description of the GPCR complement of the temperate liver fluke, F. hepatica. Using HMM-led methods to examine available F. hepatica genome datasets, we identified 166 GPCR-like sequences in F. hepatica (Figs. 1 and S1 Table). Fig. 1B shows that 49.7% contained 7 TM domains, with 88% of sequences containing at least four TMs. The remainder of this manuscript focuses on 147 sequences containing ≥ 4 TM domains (S1 Table; S2 Text). Twenty-two sequences containing ≤ 3 TMs were not analysed further (Fig. 1).

Our ≥ 4TM dataset (147 sequences) was comprised of three gluta-mate, 135 rhodopsin, two adhesion, five frizzled, one smoothened, and one secretin GPCR. Sequence coverage was generally good in terms of TM and extracellular domain representation, so we did not attempt to extend truncated sequences into full-length receptors. The overall dataset contained excellent representation of seven TM domains, while N-terminal extracellular LBDs and cysteine-rich domains (CRD) were also detected (in glutamate, frizzled/smoothened, adhesion families). However, we could not identify N-terminal secretory signal peptides in any sequence, suggesting incomplete sequence coverage at extreme N-termini. Rhodopsins are designated by ubiquitously conserved motifs on TMs 2, 3, 6 and 7. All rhodopsin sequences contained at least one of these motifs (Fig. 2 and S3 Table), including in the highly diverged flatworm-specific rhodopsins described below.

Table 1 compares the F. hepatica GPCR complement with other flatworms, illustrating that F. hepatica has the largest GPCR complement reported from any parasitic flatworm to date. The bulk of the expansion involves rhodopsins, while the other GRAFS families are comparable between F. hepatica and other flatworm parasites.

3.2. Stringent annotation of flatworm-specific orphan rhodopsin GPCRs in F. hepatica

Encompassing 135 sequences, the rhodopsin family is the largest of the GRAFS classifications in F. hepatica. Rhodopsins comprise four subfamilies (α, β, γ and δ) (Lagerström and Schiöth, 2008); we identified members of both α and β groups, with nucleotide-activated (P2Y) receptors (γ group), and olfactory (δ group) receptors absent from our dataset (Figs. 1 and 2; S1 Table). The F. hepatica α subfamily contained 38 amine receptors and three opsins, with the β subfamily comprised of at least 47 peptide receptors. Homology-based annotations were supported by an ML phylogeny (Fig. 2A), which clearly delineated between amine and opsin α clades, and the peptide-activated β-rhodopsin clades. Amine and peptide receptors were further delineated by additional phylogenetic and structural analyses, permitting high-confidence assignment of putative ligands to 16 GPCRs (see section 3.4).

Six clades contained an additional 44 rhodopsin sequences with low scoring (median E = 5.6e−5) similarity matches to a range of disparate α and β rhodopsins. Due to the subsequent difficulty in designating these clades as amine, peptide or opsin, we labelled them orphan rhodopsins (‘R’ clades in Fig. 2A). Eighteen GPCRs within the orphan
A

B

C

Ubiquitous rhodopsin motifs

F. hepatica rhodopsins
n=117

F. hepatica fwRhods
n=18

(caption on next page)
clades displayed exceptionally low similarity scores relative to non-flatworm sequences (Fig. 2A and B). Seven returned non-significant hits in BLASTp searches against non-flatworm members of the ncbi nr dataset (the most diverse sequence dataset available to the research community), and the remaining eleven scored E > 0.01. Domain analysis (InterPro) identified rhodopsin domains (IPR000276 or IPR019430) in thirteen of these (S1 Table, S3 Table), confirming their identity as rhodopsin-like GPCRs. More troublesome to classify were five that, in addition to lacking significant BLASTp identity to non-flatworm sequences, also lacked any identifiable protein domains/motifs (with the exception of TM domains). We annotated these as rhodopsins because: (i) They did not contain motifs/domains representative of any other protein family; (ii) They displayed topological similarity to GPCRs (ten had seven TM domains, seven had six TMs, one had five TM domains); (iii) They contained at least two of the conserved rhodopsin motifs in TM domains 2, 3, 6 and 7 similar to those seen in the rest of the F. hepatica rhodopsins (Fig. 2C, S4 Table). As highly diverged rhodopsins with little or no sequence similarity versus host species, these 18 F. hepatica receptors have obvious appeal as potential targets for flukicidal compounds with exquisite selectivity for parasite receptors over those of the host. This potential is contingent on future work demonstrating essential functionality for these receptors; showing their wider expression across flatworm parasites would enable consideration of anthelmintics with multi-species activity. To investigate the latter question, we used BLASTp to search the 18 F. hepatica rhodopsins against other available genomes representing phylum Platyhelminthes

3.3. An orphan family of lineage-expanded rhodopsins in flatworm genomes

Although lacking similarity against non-flatworm datasets, each of the 18 lineage-expanded F. hepatica rhodopsins returned high-scoring hits in BLASTp searches against the genomes of other flatworms (WormBase Parasite release WBPS9). All returns were subsequently filtered through a stringent five-step pipeline (Fig. 3A) consisting of: (i) Removal of duplicate sequences; (ii) Exclusion of sequences containing fewer than four TM domains; (iii) A requirement for reciprocal BLASTp against the F. hepatica genome to return a top hit scoring E < 0.001 to one of the original 18 F. hepatica queries; (iv) A requirement for reciprocal BLASTp against ncbi nr non-flatworm sequences to return a top hit scoring E > 0.01; (v) Removal of sequences lacking conservation of the ubiquitous rhodopsin motifs seen in the divergent F. hepatica rhodopsins (Figs. 2C and 3C). The latter motifs were largely absent from cestode rhodopsins (with the exception of a single sequence from Diphyllobothrium latum, and three sequences from Schistoscoeca solidus), and present in only two sequences from a single monogenean (Protostrongylus xenopodis). This left our final dataset consisting of 76 “flatworm-specific” rhodopsins (fwRhods; Fig. 3B, Table S4) in phylum Platyhelminthes, heavily biased towards trematodes (70 sequences). Nineteen sequences from nine species of cestode were omitted from the final dataset despite meeting the inclusion criteria in most respects, because they lacked conservation of ubiquitous rhodopsin motifs (filtering step (v)). Although their further characterisation was beyond the scope of this study, they warrant more detailed examination in future studies as potential cestode-specific rhodopsins. Note that our filtering pipeline also excluded initial hits from Gyrodactylus salaris (Monogenea), and the turbellarians Macrostromum lignano and S. mediterranea. Individual species complements of fwRhods showed some consistency (Fig. 3B); the trematodes F. hepatica and Echinostoma caproni (both phylum Platyhelminthes, order Echinostomida) bore 18 and 19 sequences, respectively, most species of family Schistosomatidae contained 3–4 sequences each. The inclusion of two cestode species and a single monogenean may be an indication of the existence of distantly related rhodopsins in those lineages, rather than a true measure of the extent of cestode and monogenean fwRhod diversity. Proper classification of these groups will require further, Class-focused study.

Our method for identification of fwRhods is supported by a similar BLAST-driven approach used to identify highly diverged “hidden orthologues” in flatworms (Martin-Duran et al., 2017), as well as by similar, less stringent, methods used to identify PROF1 GPCRs (Zamanian et al., 2011). It should be noted that the existence of sequences lacking sequence similarity to genes of other species is not a new finding. “Taxonomically-restricted genes” comprise 10–20% of every sequenced eukaryote genome, and may be essential for phylum-specific morphological and molecular diversity (Khalturin et al., 2009). We also considered how our fwRhods compare to previously reported groups of flatworm restricted GPCRs in S. mansoni, S. mediterranea and E. granulosus (Zamanian et al., 2011; Tsai et al., 2013; Saberi et al., 2016). Phylogenetic comparisons (Fig. 3D) demonstrated that the previously described Schmidtea Srfb cluster (Saberi et al., 2016) and the PROF1 clade (E. multilocularis, Schmidtea, S. mansoni (Zamanian et al., 2011; Tsai et al., 2013) are equivalent, and likely represent a single group. Our phylogeny added 23 fwRhods to this clade, including three from F. hepatica (BN1106_s6156B000040, D915_03083, D915_13002). Fig. 3D designated the remaining fwRhods within additional pre-existing groups (Saberi et al., 2016), placing 34 within Rho-L (including eight from F. hepatica), nine in Srfc (one from F. hepatica), four in Rho-R (one from F. hepatica) and two in Srfb (one from F. hepatica). Four fwRhod sequences were omitted from this tree due to poor alignment.

There is no set definition for lineage specificity in the flatworm GPCR literature, with the previous studies describing PROF1 (Zamanian et al., 2011), Srfb/b/c and Rhol/R (Saberi et al., 2016) receptors employing distinct methods and criteria (we have employed a similar, but more stringent, E value-driven approach to the former). An additional compounding factor is that many flatworm GPCRs described as taxonomically restricted still return high scoring matches from BLASTp searches of non-flatworm sequence datasets. For example, applying our BLASTp E ≥ 0.01 cutoff (modified from Pearson, 2013) to these published groups, would exclude 57 of the 62 PROF1s described from S. mansoni and S. mediterranea (most of the excluded sequences in this case can be explained by expansion in the ncbi nr dataset since their description in 2011), and 287 of the 318 Rhol/R and Srfb/b/c flatworm-specific clusters in S. mediterranea. This indicates the difficulty in interpreting existing definitions of “lineage specificity” or “taxonomic restriction” amongst flatworm GPCRs, and we therefore feel justified in applying our own simple, but more stringent definition for
A. Diagram showing the workflow for identifying Rhodopsin motifs in Platyhelminthes:

- F. hepatica lineage-expanded GPCRs (18) are identified through BLASTp and WormBase Parasite.
- Reciprocal BLASTp compares with WBPS9 and Platyhelminthes.
- BLASTp ncbi nr excl Platyhelminthes is used to filter results with E<0.01.
- HMMScan domain search is applied to identify no domains.
- HMMTop >3TM filters results.
- Rhodopsin motifs are identified for 76 fwRhods in Phylum Platyhelminthes.

B. Phylogenetic tree showing the distribution of Rhodopsin motifs across various Platyhelminthes species:

- Trematoda includes species like Monogenea (2) and Cestoda (4) with motifs.
- F. hepatica (18) also shows motifs.
- Other species such as E. caproni (19), S. japonicum (1), and S. mansoni (5) exhibit similar motifs.

C. Sequence alignment showing motifs in F. hepatica fwRhods (n=18) and Phylum Platyhelminthes fwRhods (n=58):

- The alignment highlights key amino acid sequences YVQGDRYPMTTAPNEDAL FFVVUGATTAYLLATQF.

D. Circular phylogenetic network highlighting the relationships and motifs among various Platyhelminthes species:

- The network includes nodes representing different species and their motifs.
- The network is color-coded to differentiate species and motifs.

(caption on next page)
Fig. 3. Identification of flatworm-specific rhodopsins (fwRhods) in genomes from phylum Platyhelminthes. (A) The 18 Fasciola hepatica GPCRs in our dataset that had poor BLASTp similarity (E > 0.01) to non-flatworm sequences in the ncbi nr dataset (isGPCRs), were used as queries in BLASTp searches of flatworm genomes in Parasite (release WBPS9). All hits scoring E < 0.01 were back-searched by BLASTp against our F. hepatica GPCR dataset. Sequences scoring E < 0.01 against one of the original F. hepatica GPCRs were retained as matches. These sequences were then filtered to identify those lacking matches in ncbi nr, lacking non-GPCR protein domains, possessing at least 4 transmembrane (TM) domains, and containing rhodopsin motifs consistent with those seen in the majority of F. hepatica rhodopsins (see C). (B) This process identified 76 fwRhods in phylum Platyhelminthes, the majority (70%) of which were from class Trematoda. Small numbers were returned from classes Cestoda and Monogenea. Note that no fwRhods fitting these criteria were identified in class Turbellaria. (C) Sequence diversity within ubiquitous rhodopsin motifs of 18 F. hepatica fwRhods (upper panel), compared to those motifs in the 58 fwRhods identified in the wider phylum (lower panel); motifs are broadly similar between Turbellaria. (D) Maximum likelihood phylogeny of 76 fwRhods, alongside flatworm-specific rhodopsins described previously (70 platyhelminth rhodopsin orphan family 1 (PROF1) (Zamanian et al., 2011; Tsai et al., 2013), and 245 phylum (lower panel); motifs are broadly similar between Turbellaria. (C) Sequence diversity within ubiquitous rhodopsin motifs of 18 existing groups.

3.4. Predicting ligands for F. hepatica rhodopsin GPCRs

In addition to the flatworm-specific fwRhod sequences described above, for which the ligands and functions remain cryptic, we also identified many rhodopsins with clear similarity to previously annotated GPCRs. Fig. 2A shows the phylogenetic delineation of these sequences into amine-, opsin- and peptide-like receptors, distinctions that are supported by BLASTp comparisons with general (ncbi nr) and lineage-specific (superphylum level) datasets, as well as by gross domain structure (InterProScan) (S1 Table). These data provided a foundation for the deeper classification of putative ligand-receptor matches.

The structure and function of GPCR LBDs can be studied using molecular modelling to predict interactions with receptor-bound ligands. These predictions can then be validated by targeted mutagenesis to newly discovered receptors. Since mutagenesis experiments have not yet been performed in flatworm GPCRs, we employed a comparative approach to identify 17 F. hepatica rhodopsins with LBD motifs diagnostic of receptors for NPF/Y, 5-HT, octopamine (Oct) or acetylcholine (ACh) (Fig. 4; S4 Table), thus enabling in silico ligand-receptor matching of these GPCRs.

Comparison of F. hepatica rhodopsins by structural alignment with LBD residues conserved across vertebrate NPY and dipteran NPF receptors (Sautel et al., 1995, 1996; Berglund et al., 2002; Akberg et al., 2010; Füllmair et al., 2011; Vogel et al., 2013) identified three peptide receptors with more than 75% identity across 9 ligand-interacting positions (Fig. 4A). The two highest scoring GPCRs (BN1116_s3169B000088 and D915_05685) are also found, in our phylogenetic analysis (S5 Figure) in the same clade as the deorphanized NPF/Y receptors of human (HsNPYR2), Glossina mortisans (Glonmo-NPR) and S. mediterranea (SmedNPYR1). These data designate these three F. hepatica GPCRs as prime candidates for further work to deorphanize and confirm these receptors as NPF/Y-activated, and to probe the biology of NPF/Y receptors in parasitic flatworms. A single NPF/Y receptor has been functionally characterised in S. mediterranea, displaying a role in the maintenance of sexual maturity (Saberi et al., 2016). If related functions are conserved in liver fluke NPF/Y receptors they could have appeal as therapeutic targets in adult fluke that could interrupt parasite transmission, although their utility for the control of acute fasciolosis, caused by migrating juveniles, would be open to question.

Broad phylogenetic comparison of our peptide receptor set with a comprehensive collection of deorphanized bilaterian rhodopsin GPCRs (S5 Figure), identified F. hepatica receptors similar to those for myomodulin, FLP, luquin and Neuropeptide KY (NKY). These ligands have all been predicted or demonstrated in previous biochemical or in silico studies of flatworm neuropeptides (McVeigh et al., 2009; Collins et al., 2010; Koziol et al., 2016), as well as from other genera, in a receptor activation assay. Subsequent localisation of their spatial expression patterns would provide additional data that would inform function.

In silico

No F. hepatica neuropeptide sequences have been published yet, but our unpublished data suggest the presence of at least 36 neuropeptide genes in the F. hepatica genome (Duncan Wells, Queen’s University Belfast, personal communication). These ligands would facilitate deorphanisation of heterologously-expressed peptide GPCRs (S1 Table). This is essential work, as although two planarian peptide receptors have been deorphanised (Omar et al., 2007; Saberi et al., 2016), no flatworm parasite peptide GPCRs have been ligand matched. Receptor deorphanisation provides a starting point for drug discovery, by enabling development of agonists or antagonists that modulate the interaction of a GPCR with its cognate ligand. Such compounds could form the basis of ligand series for screening pipelines to support the discovery of new potential flukicides (Yoshida et al., 2012; Stockert and Devi, 2015).

Serotonin (5-hydroxytryptamine, 5-HT) is abundant throughout flatworm nervous systems, and is considered the primary flatworm excitatory neurotransmitter (Ribeiro et al., 2005). Deorphanized GPCRs activated by 5-HT have been described in turbellarians and trematodes, taxonomically restricted GPCRs from Fasciola, and their orthologues in other flatworms. Despite taking a slightly different approach to previous work, the taxonomically-restricted nature of our fwRhods was validated in every case by comparative analysis with other tools. The WormBase Parasite community resource provides comparative genomics analyses for every gene in available parasite genomes. These are driven by Ensembl Compara pipelines (Vilella et al., 2009; Howe et al., 2017) that identify orthologues and paralogues for each parasite gene represented by a gene model. These tools confirm that all of the sequences we have designated as fwRhod in F. hepatica and other flatworms, lack orthologues outside of phylum Platyhelminthes, and our phylogenetic analyses confirm that they represent new members of existing groups.

We have established the existence of a group of rhodopsin GPCRs that appear restricted to, and expanded in, phylum Platyhelminthes. By definition these receptors are orphan (i.e. their native ligands are unknown), so key experiments must focus on identifying their ligands and functions. Such experiments can exploit the expanding molecular toolbox for flatworm parasites, which in F. hepatica includes RNA interference (RNAi) (McGonigle et al., 2008; Rinaldi et al., 2008; Dell’Oca et al., 2014; McVeigh et al., 2014) interface with enhanced in vitro maintenance methods, and motility, growth/development and survival assays (McGonigle et al., 2008; McCamnick et al., 2016; McCusker et al., 2016). Our phylogeny (Fig. 2A) suggests that fwRhods are more similar to peptide than amine receptors. If their heterologous expression can be achieved, one approach to characterisation would be to screen them with the growing canon of peptide ligands from flatworms (McVeigh et al., 2009; Collins et al., 2010; Koziol et al., 2016), as well as from other genera, in a receptor activation assay. Subsequent localisation of their spatial expression patterns would provide additional data that would inform function.

Amphipathic

This is essential work, as although two planarian peptide receptors have been deorphanised (Omar et al., 2007; Saberi et al., 2016), no flatworm parasite peptide GPCRs have been ligand matched. Receptor deorphanisation provides a starting point for drug discovery, by enabling development of agonists or antagonists that modulate the interaction of a GPCR with its cognate ligand. Such compounds could form the basis of ligand series for screening pipelines to support the discovery of new potential flukicides (Yoshida et al., 2012; Stockert and Devi, 2015).

Serotonin (5-hydroxytryptamine, 5-HT) is abundant throughout flatworm nervous systems, and is considered the primary flatworm excitatory neurotransmitter (Ribeiro et al., 2005). Deorphanized GPCRs activated by 5-HT have been described in turbellarians and trematodes,
Fig. 4. Conservation of ligand-interacting residues between 17 *Fasciola hepatica* G protein-coupled receptors (GPCRs) and structurally characterised homologues from other species. (A) Neuropeptide F/Y receptor ligand binding residues as characterised by mutagenesis in human neuropeptide Y receptor NPY1R (Sautel et al., 1995, 1996; Berglund et al., 2002; Åkerberg et al., 2010; Fällmar et al., 2011), and conserved in *Anopheles gambiae* (Ag) and *Drosophila melanogaster* (Dm) neuropeptide F receptors (NPFR) (Vogel et al., 2013). Numbering relative to HsNPY1R. (B) Serotonin (5-hydroxytryptamine; 5HT) receptor ligand binding residues as characterised by mutagenesis in human 5HT receptor (Hs5HT1A) (Nakamura et al., 2015), and conserved in *Schistosoma mansoni* 5HTreceptor (Patocka et al., 2014). Numbering relative to Hs5HT1A. (C) Octopamine receptor (OaR) ligand binding residues as characterised by homology modelling of the *Periplaneta americana* (Pa) (Hirashima and Huang, 2008), and mutational analysis of the *Bombyx mori* (Bm) (Huang et al., 2007) octopamine receptor ligand binding domain. Numbering relative to PaOAR, except for Y412 which is shown relative to BmOAR. (D) Acetylcholine receptor ligand binding residues as characterised by homology modelling of the *S. mansoni* G protein-coupled acetylcholine receptor (SmGAR) (MacDonald et al., 2016); numbering relative to SmGAR. In each case, only *F. hepatica* sequences displaying at least 75% identity across the stated ligand binding residues are shown. Relative positions of residues across seven transmembrane domains (TM1-7) are shown.

|   | TM1 | TM2 | TM3 | TM4 | TM5 | TM6 | TM7 |
|---|-----|-----|-----|-----|-----|-----|-----|
| A |     |     |     |     |     |     |     |
| HsNPY1R (AAA59920) |     |     |     |     |     |     |     |
| AgNPFR (AA781602) | E W C Q V | W | Q | N D |     |     |     |
| DmNPFR (AAK50050) | E W C Q V | W | Q | N D |     |     |     |
| BN1106_s3169B00088 | L W C Q V | W | Q | N D |     |     |     |
| D915_05685 | W W C Q V | W | Q | N E |     |     |     |
| BN1106_s19B000334 | E W C Q V | W | Q | N E |     |     |     |

|   | TM1 | TM2 | TM3 | TM4 | TM5 | TM6 | TM7 |
|---|-----|-----|-----|-----|-----|-----|-----|
| B |     |     |     |     |     |     |     |
| HsSHT1A (NP_000515) | D82 | C109 | D116 | D134 | H135 | C187 | S199 | F200 | N386 |
| SmSHTR (Smp_126730) |     |     |     |     |     |     |     |     |     |
| BN1106_s1436B000114 | D | C | D | D | R | Y | C | S | T | – |
| BN1106_s10B000515 | D | C | D | D | R | Y | C | S | T | – |
| BN1106_s362B000177 | D | C | D | D | R | Y | C | S | T | – |
| BN1106_81B000700 | D | C | D | D | R | Y | C | S | T | – |
| D915_00277 | D | C | D | D | R | Y | C | S | T | – |

|   | TM1 | TM2 | TM3 | TM4 | TM5 | TM6 | TM7 |
|---|-----|-----|-----|-----|-----|-----|-----|
| C |     |     |     |     |     |     |     |
| PaOAR (AY333178.1) |     |     |     |     |     |     |     |
| BmOAR (AB255163.1)** |     |     |     |     |     |     |     |
| BN1106_s1016B000108 | D | V | T | I | S | W | Y | F |     |
| D915_02972 | D | V | T | I | S | W | Y | F |     |
| D915_08505 | D | V | T | I | S | W | Y | F |     |
| BN1106_s2834B000232 | D | V | T | I | S | A | W | F |     |
| D915_01850 | D | V | T | I | S | A | W | F |     |
| D915_05578 | D | I | T | I | S | Y | W | F |     |
| D915_06018 | D | V | T | I | S | N | W | F |     |

|   | TM1 | TM2 | TM3 | TM4 | TM5 | TM6 | TM7 |
|---|-----|-----|-----|-----|-----|-----|-----|
| D |     |     |     |     |     |     |     |
| SmGAR (Smp_145540) |     |     |     |     |     |     |     |
| D915_00814 | D | Y |     | Y | H |     | Y |   |
| BN1106_s1913B000092 | D | Y |     | Y | H |     | Y |   |
with an *S. mansoni* 5-HT receptor (Sm5HTR) involved in neuromuscular control (Nishimura et al., 2009; Zamanian et al., 2012; Patocka et al., 2014). Five *F. hepatica* rhodopsins (Fig. 4B) bore appreciable (≥80%) positional identity in amino acids shown to be key ligand-interacting residues in the human 5HT1A LBD (Wacker et al., 2013; Wang et al., 2013a). Notably, these residues were also conserved in the deorphanized *S. mansoni* 5-HT receptor (Sm5HTR, Smp.126730; Patocka et al., 2014). Three of the sequences (BN1106_s362B0000177, BN1106_s362B0000700 and BN1106_s10B000515) also resembled Sm5HTR in our phylogenetic analysis, identifying them as likely 5-HT...
Fig. 5. Fasciola hepatica glutamate G-protein coupled receptors (GPCRs) display divergent phylogeny and ligand binding domain (LBD) composition. (A) Maximum likelihood phylogeny containing three F. hepatica glutamate receptors, alongside representative receptors from the various recognised GPCR Class C subgroups (subclasses indicated by blue boxes: Ca$^{2+}$, Ca$^{2+}$-sensing receptor; GABA$\gamma$, G-aminobutyric acid type B receptors; mGluR, metabotropic glutamate receptors; Orphan, receptors with no known ligand; Taste, vertebrate taste receptors). Two previously reported Schistosoma mansoni glutamate receptors are also included; F. hepatica sequences are coloured red, S. mansoni are coloured blue, all others are black. Node numbers indicate statistical support as determined by approximate likelihood ratio test (aLRT). Tree was midpoint rooted. (B) Conservation of ligand-interacting residues between vertebrate GABA$\beta$, and metabotropic glutamate receptors (mGluR), and F. hepatica class C GPCRs. Agonist-interacting residues were identified by multiple protein sequence alignment of F. hepatica glutamate receptors against mutually-identified ligand interacting residues (those causing a significant reduction in receptor signalling activity), from mouse GABA$\beta$ receptor (top panel), or selected human mGluR subtypes (lower panel). Identical amino acids in F. hepatica/S. mansoni GPCRs are represented by white text on black background, functionally conserved amino acids by black text on grey background. In lower panel, mutations causing a significant reduction in mGluR receptor activity are bold and numbered, with the region of the glutamate molecule bound by each residue indicated (COOH, C-terminus; NH$_2$, N-terminus). For references see (Galvez et al., 2000; Wang et al., 2009; Wellendorph & Bräuner-Osborne, 2009; Geng et al., 2013). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

receptors. The remaining two (D915_00277 and BN1106_s1436B000114) appeared phylogenetically more similar to an S. mansoni dopamine receptor (Smp.127310; Taman and Ribeiro, 2009). These annotations provide rational starting points for receptor deorphanization using functional genomic and/or heterologous expression tools. We found that F. hepatica dopamine-like receptors, identified by phylogeny (SS Figure), displayed poor conservation (max 56% overall identity) to the human D2 LBD (Kalani et al., 2004). Due to this lack of selectivity, we did not annotate any F. hepatica GPCRs as dopamine receptors.

Although common in other invertebrates, octopamine has not yet been directly demonstrated as a neurotransmitter in flatworms. Evidence for its presence is indirect, based on tyramine $\beta$-hydroxylase (octopamine’s biosynthetic enzyme) activity in ceptodes and planaria (Ribeiro and Webb, 1983; Nishimura et al., 2008). Three rhodopsins (Fig. 4C) showed 100% conservation of the arthropod octopamine LBD, as determined from Periplaneta americana and Bombyx mori (Huang et al., 2007; Hirashima and Huang, 2008), with an additional four showing 88% conservation. Of these seven rhodopsins, four resolved in close phylogenetic proximity to Drosophila mushroom body octopamine receptors (D915_02972, Drosophila octopamine beta-receptors (D915_05805 and BN1106_s1016B000108) (SS Figure) or a Drosophila tyramine receptor (D915_05578), denoting these as high-confidence octopamine receptors. These data provide further evidence in support of a functional role for this enigmatic classical neurotransmitter in flatworms.

Acetylcholine has species-specific impacts on flatworm neuromuscular preparations in vitro, with myoinhibitory effects in Fasciola (McVeigh and Maule, 2017). Two putative muscarinic acetylcholine receptors (mAChRs) shared highest LBD identity with a Rat M3 ACh receptor (Fig. 4D) (Kruse et al., 2012). Although these were only 67% identical to the rat sequence, the five ligand-interacting residues within their LBDs were 100% identical to those of a deorphanised S. mansoni mAChR, known to be involved in neuromuscular coordination (SmGAR) (MacDonald et al., 2016). These receptors (D915_00814 and BN1106_s1913B00092) were also the most similar to SmGAR in our phylogeny (SS Figure) so we consider them amongst our high confidence candidates for deorphanization.

3.5. F. hepatica glutamate receptors bear divergent glutamate binding domains

At least three glutamate-like GPCRs exist in F. hepatica (Fig. 5A and S1 Table). All three are defined by significant BLASTp similarity (median E = 2.3e-34) to metabotropic glutamate receptors (mGluRs), and/or by the presence of InterPro GPCR family 3 (Class C) domains (Table S1). These three GPCRs were performed alongside receptors representative of the various Class C subgroups (Fig. 5) (Wellendorph & Bräuner-Osborne, 2009), including Ca$^{2+}$-sensing receptors, $\gamma$-aminobutyric acid type B (GABA$\beta$) receptors, metabotropic glutamate (mGluR) receptors, and vertebrate taste receptors; for reference we also included previously reported mGluRs from S. mansoni (Zamanian et al., 2011; Campos et al., 2014). One F. hepatica GPCR (BN1106_s2924B000081) resolved alongside the mGluR clade, supporting designation as an mGluR. A second F. hepatica glutamate receptor (BN1106_s1717B000113) has a close S. mansoni orthologue (SmP 128940), both of which reside in an orphan outgroup that is of uncertain provenance. The third F. hepatica glutamate receptor resides within another orphan group with human GPR158 and GPR179, two closely-related class C GPCRs expressed respectively in the human brain and retina (Orlandi et al., 2013). Although these receptors have been linked with specific disease states (Orhan et al., 2013; Patel et al., 2015), their ligands remain unknown.

Divergence within the LBD can inform the ligand selectivity of Class C receptors (Wang et al., 2009; Mitri et al., 2004; Zamanian et al., 2011). To further classify the two orphan glutamate GPCRs described above, we generated multiple sequence alignments to analyse the conservation of established agonist-interacting residues between mammalian mGluR and GABA$\beta$ receptors and our F. hepatica GPCRs. These analyses identified no significant conservation of either mGluR or GABA$\beta$ LBD residues (Fig. 6B). Fig. 6B also includes the previously-reported S. mansoni glutamate receptors, where Smp.052660 contains a relatively well-conserved LBD with SmP.062660 appearing more atypical. Since all three F. hepatica glutamate GPCRs bear atypical LBDs with respect to both GABA$\beta$ and mGluR, it remains difficult to define unequivocally their ligand selectivity on the basis of conserved motifs. Nevertheless, the lack of in silico evidence for F. hepatica GABA$\beta$ GPCRs reflects the dominance of GABA$\beta$-like pharmacology, which suggests that flatworm GABA signal transduction is probably entirely mediated by ionotropic receptors (Mendoça-Silva et al., 2004; Ribeiro et al., 2005).

3.6. The Wnt binding domain is conserved in F. hepatica frizzled/smoothened receptors

Ten frizzled (fzd) GPCRs and a single smoothened (smo) GPCR are recognised in the human genome. In F. hepatica we identified five fzd-like sequences and one smo-like sequence (Fig. 6; Table S1; Table S6). All of these show high scoring similarity to annotated sequences in the ncbi nr dataset (median E = 3.8e-83), and all five fzd contain InterPro domain IPR000539, with the single smo containing domain IPR026544 (Table S1). Phylogenetic analysis of these alongside vertebrate and invertebrate receptors placed all in close proximity to existing fzd/smo groups (Fig. 6A). Four F. hepatica fzd had individual direct orthologues with the four known S. mansoni fzd GPCRs (Zamanian et al., 2011; Campos et al., 2014).

Frizzled receptors are activated by cysteine-rich glycoprotein ligands known as Wnts (Wingless and Int-1), and are involved in developmental signalling through at least three different signalling pathways. Crystalllography of mouse fzs, docked with Xenopus wnt8, identified 14 amino acids within the fzs CRD that make contact with the Wnt8 ligand (Janda et al., 2012). Positional conservation of these residues is apparent when fzs is aligned with the five F. hepatica fzd sequences (Fig. 6B; S6 Table), suggesting conservation of the wnt-frizzled interaction between liver fluke and vertebrates.

Two Wnt ligands have been described in S. mansoni (Li et al., 2010; Ta et al., 2015); our BLAST searches identified at least three Wnt-like sequences in the F. hepatica genome (BN1106_s198B000330.mRNA-1,
A

B

Ligand-contacting amino acids
*M. musculus* fz1-10, numbering relative to fz8

Aligned positions
BN1106_s2220B000114
BN1106_s1665B000273
BN1106_s2894B000110
BN1106s33B000335
BN1106_s905B000183

C

Amino acid position relative to Dro-wnt-1

Dro-Wnt-1  ECQ  NCS  ACS  CTC  CGS  CKC  SCT  TCW  FCE  QCN  CGLMC  CAC  CCEVK  CIC  TCL
BN1106_s737B000430  ---  ---  ACS  CDC  CGS  CKC  ACS  CTW  YCR  LCN  CNRKC  CDC  CCEVR  CVR  CCK
BN1106_s198B000330  ---  ---  ACS  GCC  CGS  CKC  SCE  TCW  FCH  RCV  CCEYLC  CMC  CCKV1  QTC  TCN
BN1106_s1256B000163  ECQ  NCT  SCK  GCG  GGG  CRP  L---  ---  YCO  ECL  CVCILCC  CRC  CCRVEG  QTC  VCN

(caption on next page)
BN1106_s1256B000163.mRNA-1, BN1106_s737B000430.mRNA-1; Fig. 6C). These showed conservation of the 23 cysteine residues that are diagnostic of Wnt glycoproteins (Willert and Nusse, 2012). Norrin, a non-Wnt protein ligand, can also activate Fzd4, and the canonical β-catenin pathway. The amino acids involved in norrin binding to the Fzd CRD have also been determined (Smallwood et al., 2007), but we did not observe conservation of these in any of the F. hepatica Fzds. Similarly, BLASTp searches of human norrin (Uniprot Q00604) against the F. hepatica genome did not return significant hits, suggesting that the norrin-Fzd signalling axis may not function in liver fluke. Smoothened receptors are structurally similar to frizzleds, but operate in a ligand-independent fashion within hedgehog signalling pathways that control several developmental processes (Ayers and Thérond, 2010). Model organism genomes typically contain only one smoothened gene (SMO); this was the case in S. mansoni and S. mediterranea (Zamanian et al., 2011), and here we have identified a single F. hepatica smoothened (BN1106_s1509B000194; Fig. 6; Table S1).

Fzd/smo GPCRs are involved broadly in the control of cellular development. Our discovery of fzd/smo GPCRs, and their Wnt ligands, in F. hepatica opens avenues towards probing molecular aspects of development and differentiation in the putative stem cells/neoblasts of liver fluke (McCusker et al., 2016). Neoblasts are the cells that impede the regenerative capacity of free-living turbellarian flatworms (Gehrke and Srivastava, 2016), and neoblast-like cells also represent the only proliferating cells in several parasitic species (Collins et al., 2013; Wang et al., 2013b; Koziol et al., 2014). Therefore, these cells are important in understanding fundamental fluke biology and represent potential repositories of unique anthelmintic targets, capable of inhibiting worm growth or development. The presence of both receptor and ligand sequences will permit functional genomic dissection of Wnt-Frizzled ligand-receptor signalling networks, aimed at elucidating their roles in the development and differentiation of liver fluke neoblast-like cells. These FhGPCRs will enable comparisons between the biology of parasitic and free-living flatworms, where Wnt signalling is known to be essential for anterior-posterior polarity in regenerating planaria (Gurley et al., 2008; Petersen and Reddien, 2008).

3.7. Class B (adhesion and secretin) receptors

Class B receptors incorporate both adhesions and secretins. Adhesions are characterised by a long N-terminal extracellular domain (ECD) that includes several functional motifs. These ECDs are autoproteolytically cleaved into two subunits that subsequently reassemble into a functional dimer (Lagerström and Schüth, 2008). We identified two adhesion sequences in the F. hepatica genome (S7 Figure, S1 Table), both of which (scaffold181_78723–79604, and BN1106_s537B000355) contained GPCR class B InterPro domain IPR000832 and displayed closest BLASTp similarity (E = 5.6e-7) to latrophilin-like receptors. These data suggest that both are adhesions, rather than secretins. We also identified a single secretin-like sequence in our 4-9TM dataset (BN1106_s1217B000278.mRNA-1), which also contained GPCR class B InterPro domain IPR000832, but showed closest BLASTp similarity to a pigment dispersal factor (PDF) receptor (E = 9e-46). Phylogenetic analysis of these receptors alongside human Class B receptors supports the definition of scaffold181_78723–79604 and BN1106_s537B000355 as adhesions, with BN1106_s1217B000278 appearing within the clade of secretin receptors.

Deoration of a handful of adhesions matches them with a complex assortment of ligands including collagen, transmembrane glycoproteins, complement proteins and FMR1amide-like neuropeptides (Langenhan et al., 2013). This assortment of potential ligands, and their expression in almost every organ system has led to the proposal of a diverse range of functions for vertebrate adhesions. The F. hepatica adhesion complement of two GPCRs is greatly reduced compared to the 33 receptors known in humans; in other flatworms 14, 4 and 1 adhesions have been described in S. mediterranea, E. multilocularis and S. mansoni, respectively (Zamanian et al., 2011; Tsai et al., 2013; Saberi et al., 2016). Functional characterisation will be a challenging task given the wide range of possible functions to be assayed; an appealing starting point would be to investigate roles in neoblast motility prior to differentiation, given that mammalian adhesion GPCRs are involved in the control of cellular migration (Langenhan et al., 2013).

3.8. Developmental expression

Using RNA-Seq methods, we were able to confirm the expression of 101 GPCRs across libraries representing several F. hepatica life-stages. These datasets included publically available reads from individual developmental stages (Cwiklinski et al., 2015), and a transcriptome that we generated in-house for 21-day liver stage ex-vivo juveniles (juv2). Since these datasets were generated independently and clearly display distinct sequence diversities, we avoided any further direct comparisons between Cwiklinski juveniles and our juv2 datasets. Each dataset is analysed separately, below.

Fig. 7A illustrates detection of 83 GPCRs across Cwiklinski’s developmentally staged RNA-Seq datasets. These comprised four Fzd, thirteen amnergic rhodopsins, two opsins, 41 peptidergic rhodopsins, and 23 orphan rhodopsins. The latter included nine fwrhods. Clustering within Fig. 7A’s expression heatmap shows clear developmental regulation of GPCR expression, outlining nine GPCRs with relatively higher expression in adults, two with higher expression in 21d juveniles, 64 GPCRs preferentially expressed in either 1 h, 3 h or 24 h NEJs, and six receptors expressed most highly in eggs. GPCR classes appear to be randomly distributed across these expression clusters, giving little opportunity to infer function from expression. Adult-expressed GPCRs include five orphan fwrhods, three peptiderge receptors including a putative NP/V/Y receptor, and a predicted octopamine-gated amnergic rhodopsin. The majority of expressed GPCRs occurred in the NEJ-focused expression cluster. Given data implicating GPCRs in motility, growth/development and sensory perception (McVeigh et al., 2012), it is no surprise to find high levels of GPCR expression in the NEJs, which must navigate and burrow their way from the gut lumen into the liver parenchyma, while also sustaining rapid growth from the start of the infection process. The high expression in these stages, of receptors that we predict to be activated by myomodulators such as ACh, FMR1amide, GYIRFamide, myosupressin and 5-HT, provide tentative support for these predictions. The focused expression of six GPCRs in eggs suggests potential roles in the control of cellular proliferation and fate determination processes that occur during embryonation. This complement did not include frizzled or adhesion GPCRs that are traditionally implicated in the control of development, instead consisting of rhodopsins (including an angiotensin-like peptide receptor, two octopamine-like amine receptors, one opsin receptor and one fwrhod receptor).

Focusing on the pathogenic 21-day juvenile stage, we detected 76 GPCRs in our juv2 datasets, and 29 in the corresponding juv1 samples
from Cwiklinski’s dataset (Fig. 7B). Our jv2 dataset included three glutamate, one adhesion, four frizzled, one smoothened, and 67 rhodopsins. The identity of the receptors expressed here again attest to the key role of neuromuscular co-ordination in this highly motile life stage, which must penetrate and migrate through the liver parenchyma en route to the bile ducts. Amongst the receptors expressed in this stage

P. McVeigh et al.  IJP: Drugs and Drug Resistance 8 (2018) 87–103
and thought to have a role in neuromuscular function are several ac-
tivated by classical neurotransmitters including ACh, dopamine and 5-
HT. The latter indicates receptor classi-
 venues, points to the importance of studying these receptors with a
view to discovery. The damage caused by migrating juvenile
fluke requires that new fluclides are effective against this stage. The
neuromuscular GPCRs expressed in migrating juveniles provide comp-
telling targets for new drugs.

4. Conclusions

GPCRs are targets for 33% of human pharmaceuticals (Santos et al.,
2017), illustrating the appeal of GPCRs as putative anthelmintic targets.
This study provides the first description of the F. hepatica GPCR com-
plement permitting consideration of a GPCR target-based screening
approach to fluclide discovery. To facilitate the deorphanization ex-
periments that will precede compound screening efforts, we have de-
scribed a set of high confidence rhodopsin ligand-receptor pairs. We
identified these GPCRs, including receptors for ACh, octopamine, SHF
and NPF/Y, through phylogenetic comparison with existing deorpha-
nised receptors and positional conservation of ligand-interacting re-
sidues within ligand binding domains. Our additional descriptions of
new members of existing flatworm-specific rhodopsin groups in Fascioulle
and other species support the potential for synthetic ligands to be para
tise-selective antihelmintics.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.
doi.org/10.1016/j.jipddr.2018.01.001.

References

Ahunna, F., Arafah, L., Megeria, B., Regassa, A., 2010. Bovine fasciolosis:coprol
tary survey and its economic impact due to liver condemnation at Soddo muni-
cipal abattoir, Southern Ethiopia. Trop. Anim. Health Prod. 42, 289–292.
Åkerberg, H., Fullman, P., Bouahora, L., Gutiérrez-de-Terán, H., Lundell, L.,
Mohell, N., Larhammar, D., 2010. Mutagenesis of human neuropeptide Y/peptide Y
receptor 2Y reveals additional differences to Y1 in interactions with high
conserved ligand positions. Regul. Pept. 163, 120–910.
Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local align-
ment search tool. J. Mol. Biol. 215, 403–410.
Ayers, K.L., Thérond, P.P., 2010. Evaluating smoothened as a G-protein-coupled receptor
for hedgehog signalling. Trends Cell Biol. 20, 287–94.
Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J.R., Maciejewski, A., Wishart, D.S.,
P. McVeigh et al.
IJP: Drugs and Drug Resistance 8 (2018) 87–103
Mohell, N., Larhammar, D., 2010. Mutagenesis of human neuropeptide Y/peptide YY
receptor NPF/Y, through phylogenetic comparison with existing deorpha-
nised receptors and other species support the potential for synthetic ligands to be
parasite-selective antihelmintics.
P. McVeigh et al.

LiP: Drug and Drug Resistance (2018) 87–103

Jiang, L., Schlesinger, F., Davis, C.A., Zhang, Y., Li, R., Salit, M., Gingeras, T.R., Oliver, B., 2011. Synthetic spike-in standards for RNA-Seq experiments. Genome Res. 21, 1543–1551.

Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McCwill, H., Maslen, J., Mitchell, A., Nally, G., Pearsall, S., Quinn, A.J., Sangrador-Vegas, A., Schreemerjew, M., Yang, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (9), 1236–1240.

Kalan, M.Y., Vaidhehi, N., Hall, S.E., Trabanino, R.J., Feddolino, P.L., Kalani, M.A., Flornato, W.B., Kam, W.V., Goddard Ted, W.A., 2004. The predicted 3D structure of the human D2 dopamine receptor and the binding site binding for agonists and antagonists. Proc. Natl. Acad. Sci. U.S.A. 101, 3815–3820.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment of short DNA sequences to the human genome. Genome Biol. 10 (3), R25.

Kelley, J.M., Elliott, T.P., Beddow, T., Anderson, G., Skuce, P., Spithill, T.W., 2016. Current threat of tricklaboralzence resistance in Fusococcus hepaticus. Trends Parasitol. 32, 755–768.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., Bosch, T.C.G., 2009. More than just scales protein function classi

Kalani, M.Y., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., Kalani, M.A., Maslen, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., 2011. Synthetic spike-in standards for RNA-Seq experiments. Genome Res. 21, 1372–1375.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. http://dx.doi.org/10.1093/bib/bbx108.

Kello, J.M., Elliott, T.P., Beddow, T., Anderson, G., Skuce, P., Spithill, T.W., 2016. Current threat of tricklaboralzence resistance in Fusococcus hepaticus. Trends Parasitol. 32, 755–768.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., Bosch, T.C.G., 2009. More than just scales protein function classi

Khalani, M.Y., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., Kalani, M.A., Maslen, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., 2011. Synthetic spike-in standards for RNA-Seq experiments. Genome Res. 21, 1372–1375.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. http://dx.doi.org/10.1093/bib/bbx108.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., Bosch, T.C.G., 2009. More than just scales protein function classi

Khalani, M.Y., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., Kalani, M.A., Maslen, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., 2011. Synthetic spike-in standards for RNA-Seq experiments. Genome Res. 21, 1372–1375.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. http://dx.doi.org/10.1093/bib/bbx108.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., Bosch, T.C.G., 2009. More than just scales protein function classi

Khalani, M.Y., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., Kalani, M.A., Maslen, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., 2011. Synthetic spike-in standards for RNA-Seq experiments. Genome Res. 21, 1372–1375.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. http://dx.doi.org/10.1093/bib/bbx108.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., Bosch, T.C.G., 2009. More than just scales protein function classi

Khalani, M.Y., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., Kalani, M.A., Maslen, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., 2011. Synthetic spike-in standards for RNA-Seq experiments. Genome Res. 21, 1372–1375.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. http://dx.doi.org/10.1093/bib/bbx108.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., Bosch, T.C.G., 2009. More than just scales protein function classi

Khalani, M.Y., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., Kalani, M.A., Maslen, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., 2011. Synthetic spike-in standards for RNA-Seq experiments. Genome Res. 21, 1372–1375.

Katoh, K., Rozewicki, J., Yamada, K.D., 2017. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings Bioinf. http://dx.doi.org/10.1093/bib/bbx108.

Khalturin, K., Hemmrich, G., Fraune, S., Augustin, R., Bosch, T.C.G., 2009. More than just scales protein function classi

Khalani, M.Y., Vaidehi, N., Hall, S.E., Trabanino, R.J., Freddolino, P.L., Kalani, M.A., Maslen, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., 2011. Synthetic spike-in standards for RNA-Seq experiments. Genome Res. 21, 1372–1375.
