1   **Profiling G protein-coupled receptors of *Fasciola hepatica***

2   **identifies orphan rhodopsins unique to phylum**

3   **Platyhelminthes**

4

5   **Short title: Profiling G protein-coupled receptors (GPCRs) in *Fasciola hepatica***

6

7   Paul McVeigh[1]*, Erin McCammick[1], Paul McCusker[1], Duncan Wells[1], Jane

8   Hodgkinson[2], Steve Paterson[3], Angela Mousley[1], Nikki J. Marks[1], Aaron G. Maule[1]

9

10

11   [1]Parasitology & Pathogen Biology, The Institute for Global Food Security, School of

12   Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn

13   Road, Belfast, BT9 7BL, UK

14

15   [2] Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

16

17   [3] Institute of Integrative Biology, University of Liverpool, Liverpool, UK

18

19   * Corresponding author

20   Email: paul.mcveigh@qub.ac.uk

21

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 146 high confidence GPCRs, representing the largest cohort of GPCRs, and the most complete 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, five frizzled and one smoothened GPCR. Stringent annotation pipelines identified 18 highly diverged rhodopsins in *F. hepatica* that maintained core rhodopsin signatures, but lacked significant similarity with non-flatworm sequences, providing a new sub-group of potential flukicide targets. These facilitated identification of a larger cohort of 76 related sequences from available flatworm genomes, representing new members of existing groups of flatworm-specific rhodopsins. These receptors imply flatworm specific GPCR functions, and/or co-evolution with unique flatworm ligands, and could facilitate development of exquisitely selective anthelminthics. 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 families implicated, in other species, in growth, development and

49   reproduction. This catalogue of liver fluke GPCRs provides a platform for new

50   avenues into our understanding of flatworm biology and anthelmintic discovery.

Author Summary

Fasciola spp. liver fluke are important veterinary pathogens with impacts on human and animal health, and food security, around the world. Liver fluke have developed resistance to most of the drugs used to treat them (flukicides). Since no vaccines exist, we need to develop new flukicides as a matter of urgency. Most anthelmintic drugs used to treat parasitic worm infections operate by impeding the functioning of their nerve and muscle. In flatworms, most nervous signals are received by a type of receptor called a G protein-coupled receptor (GPCR). Since GPCRs control important parasite functions (e.g. movement, egg-laying, feeding), they represent appealing targets for new flukicides, but have not yet been targeted as such. This work exploited the *F. hepatica* genome to determine the quantity and diversity of GPCRs in liver fluke. We found more GPCRs in the *Fasciola* genome than have been reported in any other parasitic worm. These findings provide a foundation that for researchers to determine the functions of these receptors, and which molecules/ligands they are activated by. These data will pave the way to exploring the potential of *F. hepatica* GPCRs as targets for new flukicides.

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 [1-4]. *Fasciola* spp. also infect humans, with fascioliasis considered a neglected tropical disease [5]. Anthelmintic chemotherapy currently carries the burden of fluke control, since there are no liver fluke vaccines [6]. Six flukicidal active compounds are available for general use, with on-farm resistance reported for all except oxyclozanide [7]. Resistance to the frontline flukicide, triclabendazole, also exists in human *F. hepatica* infections [8,9]. 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 [10-12], not least because many existing anthelmintics (dichlorvos, levamisole, morantel, piperazine, pyrantel, macrocyclic lactones, paraherquamide, amino acetonitrile derivatives) act upon receptors or enzymes associated with classical neurotransmission in nematodes [11] The G protein-coupled receptors (GPCRs) that transduce signals from both peptidergic and classical neurotransmitters are of broad importance to helminth neuromuscular function. Despite industry efforts to exploit helminth GPCRs in the context of anthelmintic discovery [13], only a single current anthelmintic (emodepside) has been attributed GPCR-directed activity as part of its mode of action [14-16]. GPCRs are druggable targets, since 33% of human prescription medicines are attributed a GPCR-based mode of action [17].

5

95   Despite two *F. hepatica* genomes [18,19], no GPCR sequences have been reported

96   from *F. hepatica*. In contrast, GPCRs have been profiled in the genomes of

97   trematodes (*Schistosoma mansoni* and *Schistosoma haematobium* [20,21])*,*

98   cestodes (*Echinococcus multilocularis*, *E. granulosus, Taenia solium* and

99   *Hymenolepis microstoma* [22]), and planaria (*Schmidtea mediterranea, Girardia*

100   *tigrina* [21-24]). These datasets illustrated clear differences in the GPCR

101   complements of individual flatworm classes and species, with reduced complements

102   in parasitic flatworms compared to planarians.

103

104   This study profiles the GPCR complement of the temperate liver fluke *F. hepatica* for

105   the first time, permitting comparisons with previously characterised species that

106   inform evolutionarily and functionally conserved elements of flatworm GPCR

107   signalling. We have identified and classified 146 GPCRs by GRAFS family

108   (glutamate, rhodopsin, adhesion, frizzled, secretin) assignment [25], the majority of

109   which are expressed in *Fasciola* RNA-Seq datasets. These include clear

110   orthologues of GPCRs activated by known neurotransmitters, within which we

111   performed the deepest *in silico* ligand-receptor matching analyses to date for any

112   parasitic helminth. The latter predicted ligands for 17 *F. hepatica* GPCRs,

113   designating these as primary targets for deorphanisation. Intriguingly, the dataset

114   included a set of flatworm-expanded GPCRs lacking orthologues outside of phylum

115   Platyhelminthes. Evolution of such GPCRs across the parasitic flatworm classes

116   may have been driven by flatworm-specific functional requirements or co-evolution

117   with flatworm ligands, either of which could help support novel anthelmintic

118   discovery. This dataset provides the first description of GPCRs in liver fluke, laying a

119   foundation for future advances in GPCR-directed functional genomics and flukicide

120   discovery.

121

Results and Discussion

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* (Figure 1, S1 Table). Figure 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 146 sequences containing ≥4TM domains (S1 Table; S2 Text). Twenty sequences containing ≤3 TMs were analysed no further (Figure 1).

Our ≥4TM dataset (146 sequences) was comprised of three glutamate, 135 rhodopsin, two adhesion, five frizzled, and one smoothened 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 (Figure 2, 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

149   parasites. Secretin is the notable exception, at least one of which has been identified

150   in every other species studied, but which was absent from the datasets scrutinized

151   here.

152

153   **Table 1. Comparison of the *Fasciola hepatica* G-protein coupled receptor**
154   **(GPCR) complement with those reported from other flatworms.** Species
155   complements are shown in the context of GRAFS nomenclature [25]. * Saberi et al
156   [24] described 566 GPCRs in *Schmidtea mediterrannea*, of which 516 fall within
157   GRAFS nomenclature.
158

	F. hepatica	*S. mansoni*[21]	*S. mansoni*[20]	*S. haematobium*[20]	*E. multilocularis*[22]	*S. mediterrannea*[21]	*S. mediterrannea*[24]
Glutamate	3	2	2	2	5	9	11
Rhodopsin	135	105	59	53	48	418	461
Adhesion	2	3	-	-	4	9	14
Frizzled/Smoothened	6	5	4	4	5	11	10
Secretin	0	2	5	5	1	1	20
Total	**146**	**117**	**64**	**64**	**83**	**448**	**516***

167

168   **Stringent annotation of flatworm-specific orphan rhodopsin GPCRs in *F.**

169   *hepatica***

170   Encompassing 135 sequences, the rhodopsin family is the largest of the GRAFS

171   classifications in *F. hepatica*. Rhodopsins comprise four subfamilies (α, β, γ and δ)

172   [26]; we identified members of both α and β groups, with nucleotide-activated (P2Y)

173   receptors (γ group), and olfactory (δ group) receptors absent from our dataset

174   (Figures 1, 2; S1 Table). The *F. hepatica* α subfamily contained 38 amine receptors

175   and three opsins, with the β subfamily comprised of at least 47 peptide receptors.

176   Homology-based annotations were supported by an ML phylogeny (Figure 2A),

177   which clearly delineated between amine and opsin α clades, and the peptide-

178   activated β-rhodopsin clades. Amine and peptide receptors were further delineated

179   by additional phylogenetic and structural analyses, permitting high-confidence

180   assignment of putative ligands to 16 GPCRs (see below).

181

182   Six clades contain an additional 44 rhodopsin sequences with low scoring (median E

183   = 5.6e^{-5}) similarity matches to a range of disparate α and β rhodopsins. Due to the

184   subsequent difficulty in designating these clades as amine, peptide or opsin, we

185   labelled them orphan rhodopsins ("R" clades in Figure 2A). Eighteen GPCRs within

186   the orphan clades displayed exceptionally low similarity scores relative to non-

187   flatworm sequences (Figure 2A,B). Seven returned no-significant hits in BLASTp

188   searches against non-flatworm members of the ncbi nr dataset (the most diverse

189   sequence dataset available to the research community), and the remaining eleven

190   scored E>0.01. Domain analysis (InterPro) identified rhodopsin domains

191   (IPR000276 or IPR019430) in thirteen of these (S1 Table, S3 Table), confirming their

192   identity as rhodopsin-like GPCRs. More troublesome to classify were five that, in

193   addition to lacking significant BLASTp identity to non-flatworm sequences, also

194   lacked any identifiable protein domains/motifs (with the exception of TM domains).

195   We annotated these as rhodopsins because: (i) They did not contain motifs/domains

196   representative of any other protein family; (ii) They displayed topological similarity to

197   GPCRs (ten had seven TM domains, seven had six TMs, one had five TM domains);

198   (iii) They contained at least two of the conserved rhodopsin motifs in TM domains 2,

199   3, 6 and 7 similar to those seen in the rest of the *F. hepatica* rhodopsins (Figure 2C;

200   S4 Table). As highly diverged rhodopsins with little or no sequence similarity versus

201   host species, these 18 *F. hepatica* receptors have obvious appeal as potential

202   targets for flukicidal compounds with exquisite selectivity for parasite receptors over

203   those of the host. This potential is contingent on future work demonstrating essential

204   functionality for these receptors; showing their wider expression across flatworm

205   parasites would enable consideration of anthelmintics with multi-species activity. To

206 investigate the latter question, we used BLASTp to search the 18 *F. hepatica*

207 rhodopsins against other available genomes representing phylum Platyhelminthes.

208

An orphan family of lineage-expanded rhodopsins in flatworm genomes

210 Although lacking similarity against non-flatworm datasets, each of the 18 lineage-

211 expanded *F. hepatica* rhodopsins returned high-scoring hits in BLASTp searches

212 against the genomes of other flatworms (WormBase Parasite release WBPS9). All

213 returns were subsequently filtered through a stringent five-step pipeline (Figure 3A)

214 consisting of: (i) Removal of duplicate sequences; (ii) Exclusion of sequences

215 containing fewer than four TM domains; (iii) A requirement for reciprocal BLASTp

216 against the *F. hepatica* genome to return a top hit scoring E<0.001 to one of the

217 original 18 *F. hepatica* queries; (iv) A requirement for BLASTp against ncbi nr non

218 flatworm sequences to return a top hit scoring E>0.01; (v) Removal of sequences

219 lacking conservation of the ubiquitous rhodopsin motifs seen in the divergent *F.*

220 *hepatica* rhodopsins (Figures 2C, 3C). The latter motifs were largely absent from

221 cestode rhodopsins (with the exception of a single sequence from *Diphyllobothrium*

222 *latum*, and three sequences from *Schistocephalus solidus*), and present in only two

223 sequences from a single monogenean (*Protopolystoma xenopodis*). This left our

224 final dataset consisting of 76 "flatworm-specific" rhodopsins (fwRhods; Figure 3B,

225 Table S4) in phylum Platyhelminthes, heavily biased towards trematodes (70

226 sequences). Nineteen sequences from nine species of cestode were omitted from

227 the final dataset despite meeting the inclusion criteria in most respects, because they

228 lacked conservation of ubiquitous rhodopsin motifs (filtering step (v)). Although their

229 further characterisation was beyond the scope of this study, they warrant more

230 detailed examination in future studies as potential cestode-specific rhodopsins. Note

231 that our filtering pipeline also excluded initial hits from *Gyrodactylus salaris*

232 (Monogenea), and the Turbellarians *Macrostomum lignano* and *S. mediterranea*.

233 Individual species complements of fwRhods showed some consistency (Figure 3B);

10

234   the trematodes *F. hepatica* and *Echinostoma caproni* (both phylum Platyhelminthes

235   Order Echinostomida) bore 18 and 19 sequences, respectively, most species of

236   family Schistosomatidae contained 3-4 sequences each. The inclusion of two

237   cestode species and a single monogenean may be an indication of the existence of

238   distantly related rhodopsins in those lineages, rather than a true measure of the

239   extent of cestode and monogenean fwRhod diversity. Again, proper classification of

240   these groups will require further more focused study that was beyond the scope of

241   the current work.

242

243   Our method for identification of fwRhods is supported by a similar BLAST-driven

244   approach used to identify highly diverged "hidden orthologues" in flatworms [27]. It

245   should be noted that the existence of sequences lacking sequence similarity to

246   genes of other species is not a new finding. "Taxonomically-restricted genes"

247   comprise 10-20% of every sequenced eukaryote genome, and may be essential for

248   phylum-specific morphological and molecular diversity [28]. How do our fwRhods

249   compare to previously reported groups of flatworm restricted GPCRs in *S. mansoni,*

250   *S. mediterrannea and E. granulosus* [21,22,24]? Phylogenetic comparisons of these

251   groups (Figure 3D) demonstrated that the previously described *Schmidtea* Srfb

252   cluster [24] and the PROF1 clade (*E. multilocularis, Schmidtea, S. mansoni* [21]) are

253   equivalent, and likely represent a single group. Our phylogeny added 23 fwRhods to

254   this clade, including three from *F. hepatica* (BN1106_s6156B000040, D915_03083,

255   D915_13002). Figure 3D designated the remaining fwRhods within additional pre-

256   existing groups [24], placing 34 within Rho-L (including eight from *F. hepatica*), nine

257   in Srfc (one from *F. hepatica*), four in Rho-R (one from *F hepatica*) and two in Srfa

258   (one from *F. hepatica*). Four fwRhod sequences were omitted from this tree due to

259   poor alignment.

260

11

261   Our approach to classifying flatworm-restricted rhodopsins was to err towards

262   stringency, and this may have resulted in erroneous exclusion of some sequences

263   from the dataset. There is no set definition for lineage specificity in the literature, but

264   ours is the most stringent yet applied to flatworm GPCRs. Applying our BLASTp

265   E≥0.01 cutoff (modified from Pearson [29]) to the previously described groups of

266   flatworm-specific rhodopsins [21,22,24], excludes 57 of the 62 PROF1s described

267   from *S. mansoni* and *S. mediterranea* and 287 of the 318 RhoL/R and Srfa/b/c

268   flatworm-specific clusters in *S. mediterranea*. Further pursuit of the extent of lineage

269   restricted GPCRs in the wider phylum was beyond the scope of this study, but we

270   are currently trawling for taxonomically restricted flatworm GPCRs on a phylum wide

271   scale.

272

273   We have established the existence of a group of rhodopsin GPCRs that appear

274   restricted to, and expanded in, phylum Platyhelminthes. By definition these

275   receptors are orphan (i.e. their native ligands are unknown), so key experiments

276   must focus on identifying their ligands and functions. Such experiments can exploit

277   the expanding molecular toolbox for flatworm parasites, which in *F. hepatica* includes

278   RNA interference (RNAi) [30-33], interfaced with enhanced in vitro maintenance

279   methods, and motility, growth/development and survival assays [31,34,35]. Our

280   phylogeny (Figure 2A) suggests that fwRhods are more similar to peptide than amine

281   receptors. If their heterologous expression can be achieved, one approach to

282   characterisation would be to screen them with the growing canon of peptide ligands

283   from flatworms [36-38], as well as from other genera, in a receptor activation assay.

284   Subsequent localisation of their spatial expression patterns would provide additional

285   data that would inform function.

286

287   **Predicting ligands for *F. hepatica* rhodopsin GPCRs**

288   In addition to the flatworm-specific fwRhod sequences described above, for which

289   the ligands and functions remain cryptic, we also identified many rhodopsins with

290   clear similarity to previously annotated GPCRs. Figure 2A shows the phylogenetic

291   delineation of these sequences into amine-, opsin- and peptide-like receptors,

292   distinctions that are supported by BLASTp comparisons with general (ncbi nr) and

293   lineage-specific (superphylum level) datasets, as well as by gross domain structure

294   (InterProScan) (S1 Table). These data provided a foundation for deeper

295   classification of putative ligand-receptor matches.

296

297   The structure and function of GPCR LBDs can be studied using molecular modelling

298   to predict interactions with receptor-bound ligands. These predictions can then be

299   validated by targeted mutagenesis of residues within the LBD, measuring impacts

300   with downstream signalling assays. Such experiments have been performed in

301   model vertebrates and invertebrates, enabling identification of evolutionarily

302   conserved binding residues/motifs. These data inform assignment of putative

303   ligands to newly discovered receptors. Since mutagenesis experiments have not yet

304   been performed in flatworm GPCRs, we employed a comparative approach to

305   identify 17 *F. hepatica* rhodopsins with LBD motifs diagnostic of receptors for NPF/Y,

306   5-HT, octopamine (Oct) or acetylcholine (ACh) (Figure 4; S4 Table), thus enabling *in*

307   *silico* ligand-receptor matching of these GPCRs.

308

309   Comparison of *F. hepatica* rhodopsins by structural alignment with LBD residues

310   conserved across vertebrate NPY and dipteran NPF receptors [39-44] identified

311   three peptide receptors with more than 75% identity across 9 ligand-interacting

312   positions (Figure 4A). The two highest scoring GPCRs (BN1106_s3169B000088

313   and D915_05685) are also found, in our phylogenetic analysis (S5 Figure) in the

314   same clade as the deorphanized NPF/Y receptors of human (HsNPYR2), *Glossina*

315   *mortisans* (Glomo-NPFR) and *S. mediterranea* (SmedNPYR1). These data

13

316   designate these three *F. hepatica* GPCRs as prime candidates for further work to

317   deorphanize and confirm these receptors as NPF/Y-activated, and to probe the

318   biology of NPF/Y receptors in parasitic flatworms. A single NPF/Y receptor has been

319   functionally characterised in *S. mediterranea*, displaying a role in the maintenance of

320   sexual maturity [24]. If related functions are conserved in liver fluke NPF/Y receptors

321   they could have appeal as therapeutic targets in adult fluke that could interrupt

322   parasite transmission, although their utility for the control of acute fasciolosis, caused

323   by migrating juveniles, would be open to question.

324

325   Broad phylogenetic comparison of our peptide receptor set with a comprehensive

326   collection of deorphanized bilaterian rhodopsin GPCRs (S5 Figure), identified *F.*

327   *hepatica* receptors similar to those for myomodulin, FLP, luqin and Neuropeptide KY

328   (NKY). These ligands have all been predicted or demonstrated in previous

329   biochemical or in silico studies of flatworm neuropeptides [36-38]. We also

330   uncovered *F. hepatica* GPCRs with phylogenetic similarity to allatotropin, allatostatin,

331   thyrotropin-releasing hormone and sex peptide receptors. These ligands have not

332   yet been reported in flatworms, although the existence of allatostatin-like receptors in

333   flatworms is supported by the inter-phyla activity of arthoropod allatostatins in

334   helminth (including flatworm) neuromuscular assays [45].

335

336   No *F. hepatica* neuropeptide sequences have been published yet, but our

337   unpublished data suggest the presence of at least 36 neuropeptide genes in the *F.*

338   *hepatica* genome (Duncan Wells, Queen's University Belfast, personal

339   communication). These ligands would facilitate deorphanisation of heterologously-

340   expressed peptide GPCRs (S1 Table). This is essential work, since although two

341   planarian peptide receptors have been deorphanised [23,24], no flatworm parasite

342   peptide GPCRs have been ligand matched yet. Receptor deorphanisation provides

343   a starting point for drug discovery, by enabling development of agonists or

14

344   antagonists that modulate the interaction of a GPCR with its cognate ligand. Such

345   compounds could form the basis of ligand series for screening pipelines that would

346   lead to new flukicides [46,47].

347

348   Serotonin (5-hydroxytryptamine, 5-HT) is abundant throughout flatworm nervous

349   systems, and is considered the primary flatworm excitatory neurotransmitter [48].

350   Deorphanized GPCRs activated by 5-HT have been described in turbellarians and

351   trematodes, with an *S. mansoni* 5-HT receptor (Sm5HTR) involved in neuromuscular

352   control [49-51]. Five *F. hepatica* rhodopsins (Figure 4B) bore appreciable (≥80%)

353   positional identity in amino acids shown to be key ligand-interacting residues in the

354   human 5HT1A LBD [52,53]. Notably, these residues were also conserved in the

355   deorphanized *S. mansoni* 5-HT receptor (Sm5HTR, Smp_126730) [51]. Three of the

356   sequences (BN1106_s362B000177, BN1106_s81B000700 and

357   BN1106_s10B000515) also resembled Sm5HTR in our phylogenetic analysis,

358   identifying them as likely 5-HT receptors. The remaining two (D915_00277 and

359   BN1106_s1436B000114) appeared phylogenetically more similar to an *S. mansoni*

360   dopamine receptor (Smp_127310) [54]. These annotations provide rational starting

361   points for receptor deorphanization using functional genomic and/or heterologous

362   expression tools. We found that *F. hepatica* dopamine-like receptors, identified by

363   phylogeny (S5 Figure), displayed poor conservation (max 56% overall identity) to the

364   human D2 LBD [55]. Due to this lack of selectivity, we did not annotate any *F.*

365   *hepatica* GPCRs as dopamine receptors.

366

367   Although common in other invertebrates, octopamine has not yet been directly

368   demonstrated as a neurotransmitter in flatworms. Evidence for its presence is

369   indirect, based on tyramine β-hydroxylase (octopamine's biosynthetic enzyme)

370   activity in cestodes and planaria [56,57]. Three rhodopsins (Figure 4C) showed

371   100% conservation of the arthropod octopamine LBD, as determined from

15

372 *Periplaneta americana* and *Bombyx mori* [58,59], with an additional four showing

373 88% conservation. Of these seven rhodopsins, four resolved in close phylogenetic

374 proximity to *Drosophila* mushroom body octopamine receptors (D915_02972),

375 *Drosophila* octopamine beta-receptors (D915_08505 and BN1106_s1016B000108)

376 (S5 Figure) or a *Drosophila* tyramine receptor (D915_05578), denoting these as

377 high-confidence octopamine receptors. These data provide further evidence in

378 support of a functional role for this enigmatic classical neurotransmitter in flatworms.

379

380 Acetylcholine has species-specific impacts on flatworm neuromuscular preparations

381 *in vitro*, with myoinhibitory effects in *Fasciola* [60]. Two putative muscarinic

382 acetylcholine receptors (mAChRs), shared highest LBD identity with a Rat M3 ACh

383 receptor (Figure 4D) [61]. Although these were only 67% identical to the rat

384 sequence, the five ligand-interacting residues within their LBDs were 100% identical

385 to those of a deorphanised *S. mansoni* mAChR, known to be involved in

386 neuromuscular coordination (SmGAR) [62]. These receptors (D915_00814 and

387 BN1106_s1913B000092) were also the most similar to SmGAR in our phylogeny (S5

388 Figure) so we consider them amongst our high confidence candidates for

389 deorphanization.

390

391 **_F. hepatica_ glutamate receptors bear divergent glutamate binding domains**

392 At least three glutamate-like GPCRs exist in *F. hepatica* (Figure 5A, S1 Table). All

393 three are defined by significant BLASTp similarity (median E=$2.3e^{-34}$) to metabotropic

394 glutamate receptors (mGluRs), and/or by the presence of InterPro GPCR family 3

395 (Class C) domains IPR017978, IPR000162 or IPR000337. Phylogenetic analysis of

396 these GPCRs was performed alongside receptors representative of the various Class

397 C subgroups (Figure 5) [63], including Ca^{2+}-sensing receptors, γ-aminobutyric acid

398 type B ($GABA_B$) receptors, metabotropic glutamate (mGluR) receptors, and

399 vertebrate taste receptors; for reference we also included the two previously reported

16

400   mGluRs from *S. mansoni* [21]. One *F. hepatica* GPCR (BN1106_s2924B000081)

401   resolved alongside Smp_052680, which has previously been described as an *S.*

402   *mansoni* mGluR; these receptors form a close outgroup from the mGluR clade,

403   supporting their designation as mGluRs. A second *F. hepatica* glutamate receptor

404   (BN1106_s1717B000113) also has a close *S. mansoni* ortholog (Smp_128940), both

405   of which reside in an orphan outgroup that is of uncertain provenance. The third *F.*

406   *hepatica* glutamate receptor resides within another orphan group, close to human

407   GPR158 and GPR179, two closely related class C GPCRs expressed respectively in

408   the human brain and retina [64]. Although these receptors have been linked with

409   specific disease states [65,66], their ligands remain unknown.

410

411   Divergence within the LBD can inform the ligand selectivity of Class C receptors

412   [21,67]. To further classify the two orphan glutamate GPCRs described above, we

413   generated multiple sequence alignments to analyse the conservation of established

414   agonist-interacting residues between mammalian mGluR and $GABA_B$ receptors and

415   our *F. hepatica* GPCRs. These analyses identified no significant conservation of

416   either mGluR or $GABA_B$ LBD residues (Figure 6B). Figure 6B also includes the

417   previously reported *S. mansoni* glutamate receptors [21], where Smp_052660

418   contained a relatively well-conserved LBD with Smp_062660 appearing more

419   atypical. Since all three *F. hepatica* glutamate GPCRs bear atypical LBDs with

420   respect to both $GABA_B$ and mGluR, it remains difficult to unequivocally define their

421   ligand selectivity on the basis of conserved motifs. Nevertheless, the lack of *in silico*

422   evidence for *F. hepatica* $GABA_B$ GPCRs reflects the dominance of $GABA_A$-like

423   pharmacology, which suggests that flatworm GABA signal transduction is probably

424   entirely mediated by ionotropic receptors [48,68].

425

426   **The Wnt binding domain is conserved in *F. hepatica* frizzled/smoothened**

427   **receptors**

428 Ten frizzled (fzd) GPCRs and a single smoothened (smo) GPCR are recognised in

429 the human genome. In *F. hepatica* we identified five fzd-like sequences and one

430 smo-like sequence (Figure 6; Table S1; Table S7). All of these show high scoring

431 similarity to annotated sequences in the ncbi nr dataset (median E=$3.8e^{-83}$), and all

432 five fzd contain InterPro domain IPR000539, with the single smo containing domain

433 IPR026544 (Table S1). Phylogenetic analysis of these alongside vertebrate and

434 invertebrate receptors placed all in close proximity to existing fzd/smo groups (Figure

435 6A). Four *F. hepatica* fzd had individual direct orthologs with the four known *S.*

436 *mansoni* fzd [21].

437

438 *Frizzled* receptors are activated by cysteine-rich glycoprotein ligands known as Wnts

439 (Wingless and Int-1), and are involved in developmental signalling through at least

440 three different signalling pathways [69]. Crystallography of mouse fz8, docked with

441 *Xenopus* wnt8, identified 14 amino acids within the fz8 CRD that make contact with

442 the Wnt8 ligand [69]. Positional conservation of these residues is apparent when fz8

443 is aligned with the five *F. hepatica* fzd sequences (Figure 6B; S6 Table), suggesting

444 conservation of the wnt-frizzled interaction between liver fluke and vertebrates.

445

446 Two Wnt ligands have been described in *S. mansoni* [70,71] ; our BLAST searches

447 identified at least three Wnt-like sequences in the *F. hepatica* genome

448 (BN1106_s198B000330.mRNA-1, BN1106_s1256B000163.mRNA-1,

449 BN1106_s737B000430.mRNA-1; Figure 6C). These showed conservation of the 23

450 conserved cysteine residues that are diagnostic of Wnt glycoproteins [72]. Norrin, a

451 non-Wnt protein ligand, can also activate Fz4, and the canonical β-catenin pathway.

452 The amino acids involved in norrin binding to the fz4 CRD have also been

453 determined [73], but we did not observe conservation of these in any of the *F.*

454 *hepatica* fzd. Similarly, BLASTp searches of human norrin (Uniprot Q00604) against

455 the *F. hepatica* genome did not return significant hits, suggesting that the norrin-fz

18

456 signaling axis may not function in liver fluke. Smoothened receptors are structurally

457 similar to frizzleds, but operate in a ligand-independent fashion within hedgehog

458 signaling pathways that control several developmental processes [74]. Model

459 organism genomes typically contain only one smoothened gene (SMO); this was the

460 case in *S. mansoni* and *S. mediterranea* [21], and here we have identified a single *F.*

461 *hepatica* smoothened (BN1106_s1509B000194 ; Figure 6; Table S1).

462

463 Fzd/smo GPCRs are involved broadly in the control of cellular development. Our

464 discovery of fzd/smo GPCRs, and their Wnt ligands, in *F. hepatica* opens avenues

465 towards probing molecular aspects of development and differentiation in the putative

466 stem cells/neoblasts of liver fluke [35]. Neoblasts are the cells that impart the

467 regenerative capacity of free-living turbellarian flatworms [75], and neoblast-like cells

468 also represent the only proliferating cells in several parasitic species [76-78].

469 Therefore, these cells are important in understanding fundamental fluke biology and

470 represent potential repositories of unique anthelmintic targets, capable of inhibiting

471 worm growth or development. The presence of both receptor and ligand sequences

472 will permit functional genomic dissection of Wnt-Frizzled ligand-receptor signalling

473 networks, aimed at elucidating their roles in the development and differentiation of

474 liver fluke neoblast- like cells. These FhGPCRs will enable comparisons between

475 the biology of parasitic and free-living flatworms, where Wnt signaling is known to be

476 essential for anterior-posterior polarity in regenerating planaria [79,80].

477

478 **Class B (Adhesion and Secretin) receptors**

479 Class B receptors incorporate both adhesions and secretins. Adhesions are

480 characterised by a long N-terminal extracellular domain (ECD) that includes several

481 functional motifs. These ECDs are auto-proteolytically cleaved into two subunits that

482 subsequently reassemble into a functional dimer [26]. We identified two Class B

483 sequences in the *F. hepatica* genome (S7 Figure, S1 Table), both of which

19

484 (scaffold181_78723-79604, and BN1106_s537B000355) contained GPCR class B

485 InterPro domain IPR000832 and displayed closest BLASTp similarity (E=5.6e^{-7}) to

486 latrophilin-like receptors. These data suggest that both are adhesions, rather than

487 secretins. Phylogenetic analysis of these GPCRs alongside human Class B

488 receptors supports the definition of scaffold181_78723-79604 as an adhesion,

489 alongside two previously reported *S. mansoni* adhesions (Smp_176830,

490 Smp_099670) [21], while the other receptor appears more divergent.

491 BN1106_s536B000355 pairs with another known *S. mansoni* adhesion

492 (Smp_058380) [21], although both sit in closer proximity to human secretins than

493 adhesions.

494

495 Deorphanization of a handful of adhesions matches them with a complex assortment

496 of ligands including collagen, transmembrane glycoproteins, complement proteins

497 and FMRFamide-like neuropeptides [81]. This assortment of potential ligands, and

498 their expression in almost every organ system has led to the proposal of a diverse

499 range of functions for vertebrate adhesions. The *F. hepatica* adhesion complement

500 of two GPCRs is greatly reduced compared to the 33 receptors known in humans; in

501 other flatworms 14, 4 and 1 adhesions have been described in *S. mediterranea, E.*

502 *multilocularis* and *S. mansoni*, respectively [21,22,24]. Functional characterisation

503 will be a challenging task given the wide range of possible functions to be assayed;

504 an appealing starting point would be to investigate roles in neoblast motility prior to

505 differentiation, given that mammalian adhesion GPCRs are involved in the control of

506 cellular migration [81].

507

508 **Developmental expression**

509 Using RNA-Seq methods, we were able to confirm expression of 101 GPCRs across

510 libraries representing several *F. hepatica* life-stages. These datasets included

511 publically available reads from individual developmental stages [18], and a

20

512   transcriptome that we generated in-house for 21-day liver stage ex-vivo juveniles

513   (juv2). Since these datasets were generated independently and clearly display

514   distinct sequence diversities, we avoided any further direct comparisons between

515   Cwiklinski juv1 and our juv2 datasets. Each dataset is analysed separately, below.

516

517   Figure 7A illustrates detection of 83 GPCRs across Cwiklinski's developmentally

518   staged RNA-Seq datasets. These comprised four FZD, thirteen aminergic

519   rhodopsins, two opsins, 41 peptidergic rhodopsins, and 23 orphan rhodopsins. The

520   latter included nine fwRhods. Clustering within Figure 7A's expression heatmap

521   shows clear developmental regulation of GPCR expression, outlining nine GPCRs

522   with relatively higher expression in adults, two with higher expression in 21d

523   juveniles, 64 GPCRs preferentially expressed in either 1h, 3h or 24h NEJs, and six

524   receptors expressed most highly in eggs. GPCR classes appear to be randomly

525   distributed across these expression clusters, giving little opportunity to infer function

526   from expression. Adult-expressed GPCRs include five orphan fwRhods, three

527   peptide receptors including a putative NPF/Y receptor, and a predicted octopamine-

528   gated aminergic rhodopsin. The majority of expressed GPCRs occurred in the NEJ-

529   focused expression cluster. Given data implicating GPCRs in motility,

530   growth/development and sensory perception [11], it is no surprise to find high levels

531   of GPCR expression in the NEJs, which must navigate and burrow their way from the

532   gut lumen into the liver parenchyma, while also sustaining rapid growth from the start

533   of the infection process. The high expression in these stages, of receptors that we

534   predict to be activated by myomodulators such as ACh, FMRFamide, GYIRFamide,

535   myomodulin, myosuppressin and 5-HT, provide tentative support for these

536   predictions. The focused expression of six GPCRs in eggs suggests potential roles

537   in the control of cellular proliferation and fate determination processes that occur

538   during embryonation of liver fluke eggs. This complement did not include frizzled or

539   adhesion GPCRs that are traditionally implicated in the control of development,

21

540 instead consisting of rhodopsins (including an angiotensin-like peptide receptor, two

541 octopamine-like amine receptors, one opsin receptor and one fwRhod receptor).

542

543 Focusing on the pathogenic 21-day juvenile stage, we detected 76 GPCRs in our

544 juv2 datasets, and 29 in the corresponding juv1 samples from Cwiklinski's dataset

545 (Figure 7B). Our juv2 dataset included three glutamate, one adhesion, four frizzled,

546 one smoothened, and 67 rhodopsins. The identity of the receptors expressed here

547 again attest to the key role of neuromuscular co-ordination in this highly motile life

548 stage, which must penetrate and migrate through the liver parenchyma *en route* to

549 the bile ducts. Amongst the receptors expressed in this stage and thought to have a

550 role in neuromuscular function are several activated by classical neurotransmitters

551 including ACh, dopamine and 5-HT. The peptide receptors include some with

552 phylogenetic similarity to receptors for myoactive flatworm peptides (FMRFamide,

553 GYIRFamide, NPF) [11], as well as receptors from other invertebrates activated by

554 peptide ligands known to have excitatory effects on flatworms (allatostatin A,

555 myomodulin, proctolin) [82]. The presence of highly expressed GPCRs with

556 probable neuromuscular functions in liver stage juveniles, points to the importance of

557 studying these receptors with a view to flukicide discovery. The damage caused by

558 migrating juvenile fluke requires that new flukicides are effective against this stage.

559 The neuromuscular GPCRs expressed in migrating juveniles provide compelling

560 targets for new drugs.

561

562 **Conclusions**

563 GPCRs are targets for 33% of human pharmaceuticals [17], illustrating the appeal of

564 GPCRs as putative anthelmintic targets. This study provides the first description of

565 the *F. hepatica* GPCR complement permitting consideration of a GPCR target-based

566 screening approach to flukicide discovery. To facilitate the deorphanization

567 experiments that will precede compound screening efforts, we have described a set

22

568   of high confidence rhodopsin ligand-receptor pairs. We identified these GPCRs,

569   including receptors for ACh, octopamine, 5HT and NPF/Y, through phylogenetic

570   comparison with existing deorphanised receptors and positional conservation of

571   ligand-interacting residues within ligand binding domains. Our additional descriptions

572   of flatworm-specific rhodopsins support the potential for synthetic ligands to be

573   parasite-selective anthelmintics.

574

Materials and Methods

576 **Liver fluke sequence databases**

577 We exploited two *F. hepatica* genome assemblies available from WormBase

578 ParaSite [83], generated by Liverpool University

579 (http://parasite.wormbase.org/Fasciola_hepatica_prjeb6687/Info/Index/; [18], and

580 Washington University, St Louis

581 (http://parasite.wormbase.org/Fasciola_hepatica_prjna179522/Info/Index/; [19].

582

583 **Identification of GPCR-like sequences from *F. hepatica***

584 Figure 1 summarises our GPCR discovery methodology, which employed Hidden

585 Markov Models (HMMs) constructed from protein multiple sequence alignments

586 (MSAs) of previously described *S. mansoni* and *S. mediterranea* GPCR sequences

587 [21]. Individual HMMs were constructed for each GRAFS family [25]. Alignments

588 were generated in Mega v7 (www.megasoftware.net) [84] using the Muscle algorithm

589 with default parameters. HMMER v3 (http://.hmmer.org) was employed to construct

590 family-specific HMMs (*hmmbuild*) from alignments and these were searched

591 (*hmmsearch*) against a predicted protein dataset from *F. hepatica* genome

592 PRJEB6687 consisting of 33,454 sequences [18]. Returned sequences were filtered

593 for duplicates and ordered relative to the *hmmsearch* scoring system, enabling the

594 classification of hits according to the GRAFS family to which they showed most

595 similarity (i.e. highest score, lowest E value). All remaining returns were then used

596 as BLAST queries (BLASTp and tBLASTn) to identify matching, or additional,

597 sequences originating from the PRJEB6687 and PRJNA179522 genomes (Figure 1).

598 Where sequences appeared in both genomes, we kept the longest annotated

599 sequence (S1 Table).

600

601 **GPCR annotation**

24

602 Sequences resulting from HMM searches were filtered by transmembrane (TM)

603 domain composition, using hmmtop (http://www.sacs.ucsf.edu/cgi-bin/hmmtop.py)

604 [85,86]. Sequences containing ≥4 TMs were analysed as described below.

605

Homology analyses

607 All GPCRs were used as BLASTp [87] queries, to identify their closest (highest

608 scoring) match in the ncbi non-redundant (nr) protein sequence dataset

609 (https://blast.ncbi.nlm.nih.gov/Blast.cgi), with default settings and the "Organism" field

610 set to exclude Platyhelminthes (taxid: 6157). All GPCRs were additionally searched

611 against more phylogenetically limited datasets, by using the "Organism" field to limit

612 the BLASTp searches to: (i) Basal phyla, Ctenophora (taxid:10197), Porifera

613 (taxid:6040), Placozoa (taxid:10226), Cnidaria (taxid:6073); (ii) Superphylum

614 Lophotrochozoa (taxid: 1206795), excluding phylum Platyhelminthes (taxid: 6157);

615 (iii) Superphylum Ecdysozoa (taxid: 1206794); (iv) Superphylum Deuterostomia

616 (taxid: 33511). For BLASTp searches against other flatworms, we performed local

617 BLAST+ [88] on the WBPS9 release of WormBase Parasite, which included

618 predicted protein datasets from 30 flatworm species. In all cases, we recorded the

619 single highest scoring hit, or recorded "no significant similarity found" in cases where

620 no hits were returned (Table S1); sequences generating both GPCR hits and "no

621 hits" were retained. Where the top hit was not to a GPCR, that sequence was

622 removed from the dataset.

623

Domain composition

625 GPCR identities were confirmed using InterProScan Sequence Search

626 (www.ebi.ac.uk/interpro/search/sequence-search) [89] and/or HMMER HMMScan

627 (www.ebi.ac.uk/Tools/hmmer/search/hmmscan) [90]. Again, sequences returning

628 non-GPCR domains were omitted from the dataset, with all others retained.

629

25

Motif identification

630

631 As an additional measure of confidence in our identifications, we analysed the

632 presence/absence of key motifs diagnostic of receptor families and subfamilies.

633 These analyses were performed for rhodopsins generally, the ligand binding domains

634 (LBDs) of rhodopsin receptors for acetylcholine (ACh), neuropeptide F/Y (NPF/Y),

635 octopamine and serotonin (5-hydroxytryptamine, 5HT), and for the LBDs of

636 glutamate and frizzled/smoothened families. Motifs were identified via protein

637 multiple sequence alignment (MSA) of GPCRs, performed in MAFFT

638 (www.mafft.cbrc.jp/alignment/server) [91]. Only identical amino acids were accepted

639 at each site, with conservation expressed as % identity across all sites. Motif

640 illustrations (Figures 3 & 7) were generated using WebLogo 3

641 (http://weblogo.threeplusone.com) [92].

642

Phylogenetic reconstruction

643

644 Maximum likelihood (ML) phylogenetic trees were constructed using PhyML

645 (http://www.phylogeny.fr) [93], from protein MSA generated in MAFFT

646 (www.mafft.cbrc.jp/alignment/server/) [91]. Alignments were manually edited (in

647 Mega v7) to include only TM domains, by removing extramembrane blocks aligned

648 with human glutamate, rhodopsin, adhesion or frizzled proteins. Trees were

649 constructed from these TM-focused alignments in PhyML using default parameters,

650 with branch support assessment using the approximate likelihood ratio test (aLRT),

651 under "SH-like" parameters. Trees, exported from PhyML in newick format were

652 drawn and annotated in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

653

RNA-Seq analyses

654

655 Expression of *F. hepatica* GPCRs was investigated in publically available and in-

656 house generated RNA-Seq datasets. These included developmentally staged

657 Illumina transcriptome reads associated with one of the *F. hepatica* genome projects

26

658 [18] (reads accessed from the European Nucleotide Archive at

659 http://www.ebi.ac.uk/ena/data/search?query=PRJEB6904). These samples

660 originated from distinct developmental stages of US Pacific Northwest Wild Strain *F.*

661 *hepatica* (Baldwin Aquatics), including egg (*n*=2), metacercariae (met; *n*=4), *in vitro*

662 NEJs 1h post-excystment (NEJ1h; *n*=1), *in vitro* NEJs 3h post-excystment (NEJ3h;

663 *n*=2), *in vitro* NEJs 24h post-excystment (NEJ24h; *n*=2), *ex-vivo* liver-stage juveniles

664 (juv1; *n*=1) and *ex-vivo* adult parasites (Ad; *n*=3). Our in-house datasets were

665 generated from *ex vivo* liver stage *F. hepatica* juveniles (Italian strain, Ridgeway

666 Research Ltd, UK), recovered from rat (Sprague Dawley) hosts at 21 days following

667 oral administration of metacercariae (juv2; *n*=3). Total RNA, extracted with Trizol

668 (ThermoFisher Scientific) from each of the 3 independent biological replicates, was

669 quantified and quality checked on an Agilent Bioanalyzer, converted into paired-end

670 sequencing libraries and sequenced on an Illumina HiSeq2000 by the Centre for

671 Genomic Research at the University of Liverpool, UK. RNA samples were spiked

672 prior to library construction with the ERCC RNA Spike-In Mix (ThermoFisher

673 Scientific) [94]. All read samples were analysed using the TopHat and Cufflinks

674 pipeline [95-100], with mapping against PRJEB6687 genome sequence and

675 annotation files (accessed from WormBase Parasite;

676 http://parasite.wormbase.org/ftp.html). Data were expressed as number of fragments

677 mapped per million mapped reads per kilobase of exon model (FPKM). In juv2

678 datasets we discarded GPCRs represented by fewer than 0.5 FPKM (the minimum

679 linear sensitivity that we detected with our ERCC spike in); for the staged datasets,

680 we included only receptors represented by ≥0.5 FPKM in at least one life stage.

681 Heatmaps were generated with heatmapper (http://www.heatmapper.ca/) [101] set

682 for Average Linkage, and Pearson Distance Measurement.

683

References

1. Abunna F, Asfaw L, Megersa B, Regassa A. Bovine fasciolosis: coprological, abattoir survey and its economic impact due to liver condemnation at Soddo municipal abattoir, Southern Ethiopia. Trop Anim Health Prod. 2010;42:289-92.

2. Habarugira G, Mbasinga G, Mushonga B, Teedzai C, Kandiwa E, Ojok L. Pathological findings of condemned bovine liver specimens and associated economic loss at Nyabugogo abattoir, Kigali, Rwanda. Acta Tropica. 2016;164:27-32.

3. Howell A, Baylis M, Smith R, Pinchbeck G, Williams D. Epidemiology and impact of *Fasciola hepatica* exposure in high-yielding dairy herds. Prev Vet Med. 2015;121:41-48.

4. Sariözkan S & YalÇin C. Estimating the total cost of bovine fasciolosis in Turkey. Ann Trop Med Parasitol. 2011;105:439-444.

5. Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ, Jacobson J. Helminth infections: the great neglected tropical diseases. J Clin Invest. 2008;118:1311-1321.

6. Toet H, Piedrafita DM, Spithill TW. Liver fluke vaccines in ruminants: strategies, progress and future opportunities. Int J Parasitol. 2014; 44:915-927.

7. Kelley JM, Elliott TP, Beddow T, Anderson G, Skuce P, Spithill TW. Current threat of triclabendazole resistance in *Fasiola hepatica*. Trends Parasitol. 2016;32:458-469.

8. Cabada MM, Lopez M, Cruz M, Delgado JR, Hill V, White C Jr. Treatment failure after multiple courses of triclabendazole among patients with fasciolosis in Cusco,

712   Peru: A case series. PLoS Negl Trop Dis. 2016;10: e0004361.

713

714   9. Winkelhagen AJS, Mank T, de Vries PJ, Soetekouw R. Apparent triclabendazole-

715   resistant human *Fasciola hepatica* infection, the Netherlands. Emerg Infect Dis.

716   2012;18:1028-1029.

717

718   10. Martin RJ, Robertson AP. Control of nematode parasites with agents acting on

719   neuro-musculature systems: lessons for neuropeptide ligand discovery. Adv Exp

720   Med Biol. 2010;692:138-154.

721

722   11. McVeigh P, Atkinson L, Marks NJ, Mousley A, Dalzell JJ, Sluder A et al. Parasite

723   neuropeptide biology: seeding rational drug target selection? Int J Parasitol: Drugs

724   Drug Res. 2012;2:76-91.

725

726   12. Ribeiro P & Patocka N. Neurotransmitter transporters in schistosomes: structure,

727   function and prospects for drug discovery. Parasitol Int. 2013;62:629-638.

728

729   13. Lowery DE, Geary TG, Kubiak TM, Larsen MJ. G protein-coupled receptor-like

730   receptors and modulators thereof. Pharmacia and Upjohn Company, United States.

731   2003. Patent US 6632621 B1.

732

733   14. Saeger B, Schmitt-Wrede HP, Dehnhardt M, Benten WPM, Krücken J, Harder A

734   et al. Latrophilin-like receptor from the parasitic nematode *Haemonchus contortus* as

735   a target for the anthelmintic depsipeptide PF1022A. FASEB J. 2001;15:1332-1334.

736

737 15. Harder A, Schmitt-Wrede HP, Krücken J, Marinovski P, Wunderlich F, Willson J

738 et al. Cyclooctadepsipeptides – an anthelmintically active class of compounds

739 exhibiting a novel mode of action. Int J Antimicrob Agents. 2003;22:318-331.

740

741 16. Buxton SK, Neveu C, Robertson AP, Martin RJ. On the mode of action of

742 emodepside: slow effects on membrane potential and voltage-activated currents in

743 *Ascaris suum*. Br J Pharmacol. 2011;164:453-470.

744

745 17. Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG et al. A

746 comprehensive map of molecular drug targets. Nat Rev Drug Disc. 2017;16:19-34

747

748 18. Cwiklinski K, Dalton JP, Dufresne PJ, La Course J, Williams DJL, Hodgkinson J

749 et al. The *Fasciola hepatica* genome: gene duplication and polymorphism reveals

750 adaptation to the host environment and the capacity for rapid evolution. Genome

751 Biol. 2015;16:71. Doi:10.1186/s13059-015-0632-2.

752

753 19. McNulty SN, Tort JF, Rinaldi G, Fischer K, Rosa BA, Smircich P, et al. Genomes

754 of *Fasciola hepatica* from the Americas reveal colonization with *Neorickettsia*

755 endobacteria related to the agents of Potomac horse and human sennetsu fevers.

756 PLoS Genetics. 2017;13: e1006537.

757

758 20. Campos TD, Young ND, Korhonen PK, Hall RS, Mangiola S, Lonie A et al.

759 Identification of G protein-coupled receptors in *Schistosoma haematobium* and *S.*

760 *mansoni* by comparative genomics. Parasit Vectors. 2014;7:242. Doi: 10.1186/1756-

761 3305-7-242.

762

763 21. Zamanian M, Kimber MJ, McVeigh P, Carlson SA, Maule AG, Day TA. The

764 repertoire of G protein-coupled receptors in the human parasite *Schistosoma*

765   *mansoni* and the model organism *Schmidtea mediterranea*. BMC Genomics.

766   2011;12: 596

767

768   22. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL,

769   et al. The genomes of four tapeworm species reveal adaptations to parasitism.

770   Nature. 2013;496:57-63.

771

772   23. Omar HH, Humphries JE, Larsen MJ, Kubiak TM, Geary TG, Maule AG et al.

773   Identification of a platyhelminth neuropeptide receptor. Int J Parasitol. 2007;37:725-

774   733.

775

776   24. Saberi A, Jamal A, Beets I, Schoofs L, Newmark PA. GPCRs direct germline

777   development and somatic gonad function in planarians. PLoS Biol.

778   2016;14(5):e1002457. Doi: 10/1371/journal.pbio.1002457.

779

780   25. Fredriksson R, Lagerström MC, Lundin LG, Schiöth HB. The G-protein-coupled

781   receptors in the human genome form five main families. Phylogenetic analysis,

782   paralogon groups, and fingerprints. Mol Pharmacol. 2003;63:1256-1272.

783

784   26. Lagerström MC, Schiöth HB. Structural diversity of G protein-coupled receptors

785   and significance for drug discovery. Nat Rev Drug Discov. 2008;7:339-357.

786

787   27. Martin-Duran JM, Ryan JF, Vellutini BC, Pang K, Hejnol A. Increased taxon

788   sampling reveals thousands of hidden orthologs in flatworms. Genome Res.

789   2017;27:1263-1272.

790

791   28. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. More than just

792   orphans: are taxonomically-restricted genes important in evolution? Trends Genet.

793   2009;25:404-413.

794

795   29. Pearson WR. UNIT 3.1 An introduction to sequence similarity ("homology")

796   searching. Curr Protoc Bioinformatics DOI: 10.1002/0471250953.bi0301s42.

797

798   30. Dell'Oca N, Basika T, Corvo I, Castillo E, Brindley PJ, Rinaldi G et al. RNA

799   interference in *Fasciola hepatica* newly excysted juveniles: Long dsRNA induces

800   more persistent silencing than siRNA. Mol Biochem Parasitol. 2014;197:28-35.

801

802   31. McGonigle L, Mousley A, Marks NJ, Brennan GP, Dalton JP, Spithill TW et al.

803   The silencing of cysteine proteases in *Fasciola hepatica* newly excysted juveniles

804   using RNA interference reduced gut penetration. Int J Parasitol. 2008;38:149-155.

805

806   32. McVeigh P, McCammick EM, McCusker P, Morphew RM, Mousley A, Abidi A et

807   al. RNAi dynamics in juvenile *Fasciola* spp. liver flukes reveals the persistence of

808   gene silencing *in vitro*. PLoS Negl Trop Dis. 2014;8(9): e3185.

809

810   33. Rinaldi G, Morales ME, Cancela M, Castillo E, Brindley PJ et al. Development of

811   functional genomic tools in trematodes: RNA interference and luciferase reporter

812   gene activity in *Fasciola hepatica*. PLoS Negl Trop Dis 2008;2(7):e260.

813

814   34. McCammick E, McVeigh P, McCusker P, Timson DJ, Morphew RM, Brophy PM

815   et al. Calmodulin disruption impacts growth and motility in juvenile liver fluke. Parasit

816   Vectors. 2016;9:46.

817

818   35. McCusker P, McVeigh P, Rathinasamy V, Toet H, McCammick E, O'Connor A et

819   al. Stimulating neoblast-like cell proliferation in juvenile *Fasciola hepatica* supports

820   growth and progression towards the adult phenotype *in vitro*. PLoS Negl Trop Dis.

821   2016;10(9):e0004994.

822

823   36. McVeigh P, Mair GR, Atkinson L, Ladurner P, Zamanian M, Novozhilova E et al.

824   Discovery of multiple neuropeptide families in the phylum Platyhelminthes. Int J

825   Parasitol. 2009;39:1243-1252.

826

827   37. Collins JJ 3rd, Hou X, Romanova EV, Lambrus BG, Miller CM, Saberi A et al.

828   Genome-wide analyses reveal a role for peptide hormones in planarian germline

829   development. PLoS Biol. 2010;8(10):e1000509.

830

831   38. Koziol U, Koziol M, Preza M, Costábile A, Brehm K, Castillo E. De novo

832   discovery of neuropeptides in the genomes of parasitic flatworms using a novel

833   comparative approach. Int J Parasitol. 2016;46:709-721.

834

835   39. Åkerberg, H, Fällmar H, Sjödin P, Boukharta L, Gutiérrez-de-Terán H, Lundell I,

836   et al. Mutagenesis of human neuropeptide Y/peptide YY receptor Y2 reveals

837   additional differences to Y1 in interactions with highly conserved ligand positions.

838   Regul Pept. 2010;163:120–910.

839

840   40. Berglund MM, Fredriksson R, Salaneck E, Larhammar D. Reciprocal mutations of

841   neuropeptide Y receptor Y2 in human and chicken identify amino acids important for

842   antagonist binding. FEBS Lett. 2002;518:5–910

843

844   41. Fällmar H, Åkerberg H, Gutiérrez-de-Terán H, Lundell I, Mohell N, Larhammar D.

845   Identification of positions in the human neuropeptide Y/peptide YY receptor Y2 that

846   contribute to pharmacological differences between receptor subtypes.

847   Neuropeptides. 2011;45:293–300.

848

849   42. Sautel M, Martinez R, Munoz M, Peitsch MC, Beck-Sickinger AG, Walker P. Role

850   of a hydrophobic pocket of the human Y1 neuropeptide Y receptor in ligand binding.

851   Mol Cell Endocrinol. 1995;112:215–221.

852

853   43. Sautel M, Rudolf K, Wittneben H, Herzog H, Martinez R, Munoz M, et al.

854   Neuropeptide Y and the non-peptide antagonist BIBP 3226 share an overlapping

855   binding site at the human Y1 receptor. Mol Pharmacol. 1996;50:285–292.

856

857   44. Vogel KJ, Brown MR, Strand MR. Phylogenetic investigation of peptide hormone

858   and growth factor receptors in five dipteran genomes. Front Endocrinol (Lausanne).

859   2013;4:193. Doi: 10/3389/fendo.2013.00193.

860

861   45. Mousley A, Moffett CL, Duve H, Thorpe A, Halton DW, Geary TG et al.

862   Expression and bioactivity of allatostatin-like neuropeptides in helminths. Int J

863   Parasitol. 2005;35:1557-1567.

864

865   46. Yoshida M, Miyazato M, Kangawa K. Orphan GPCRs and methods for identifying

866   their ligands. Methods Enzymol. 2012;514:33-44.

867

868   47. Stockert JA, Devi LA. Advancements in therapeutically targeting orphan GPCRs.

869   Front. Pharmacol. 2015;6:100. Doi:10.3389/fphar.2015.00100.

870

871   48. Ribeiro P, El-Shehabi F, Patocka N. Classical transmitters and their receptors in

872   flatworms. Parasitology. 2005;131:S19-S40.

873

874   49. Nishimura K, Unemura K, Tsushima J, Yamauchi Y, Taniguchi T, Kaneko S et al.

875   Identification of a novel planarian G-protein-coupled receptor that responds to

876   serotonin in *Xenopus laevis* oocytes. Biol Pharm Bull. 2009;32:1672-1677.

877

878   50. Zamanian M, Agbedanu PN, Wheeler NJ, McVeigh P, Kimber MJ, Day TA. Novel

879   RNAi-mediated approach to G protein-coupled receptor deorphanization: proof of

880   principle and characterization of a planarian 5-HT receptor. PLoS One.

881   2012;7(7):e40787.

882

883   51. Patocka N, Sharma N, Rashid M, Ribeiro P. Serotonin signaling in *Schistosoma*

884   *mansoni*: A serotonin-activated G protein-coupled receptor controls parasite

885   movement. PLoS Pathog. 2014;10(1):e1003878.

886

887   52. Wacker D, Wang C, Katritch V, Han GW, Huang XP, Vardy E et al. Structural

888   features for functional selectivity at serotonin receptors. Science. 2013;340:615-619.

889

890   53. Wang C, Jiang Y, Ma J, Wu H, Wacker D, Katritch V et al. Structural basis for

891   molecular recognition at serotonin receptors. Science. 2013;340:610-614.

892

893   54. Taman A, Ribeiro P. Investigation of a dopamine receptor in *Schistosoma*

894   *mansoni*: functional studies and immunolocalization. Mol Biochem Parasitol.

895   2009;168:24-33.

896

897   55. Kalani MY, Vaidehi N, Hall SE, Trabanino RJ, Freddolino PL, Kalani MA et al.

898   The predicted 3D structure of the human D2 dopamine receptor and the binding site

899   and binding affinities for agonists and antagonists. Proc Natl Acad Sci USA.

900   2004;101:3815-3820.

901

902   56. Ribeiro P, Webb RA. The occurrence and synthesis of octopamine and

903   catecholamines in the cestode *Hymenolepis diminuta*. Mol Biochem Parasitol.

904   1983;7:53-62.

905

906   57. Nishimura K, Kitamura Y, Inoue T, Umesono Y, Yoshimoto K, Taniguchi T et al.

907   Characterization of tyramine beta-hydroxylase in planarian *Dugesia japonica*: cloning

908   and expression. Neurochem Int. 2008;53:184-192.

909

910   58. Hirashima A, Huang H. Homology modeling, agonist binding site identification,

911   and docking in octopamine receptor of *Periplaneta americana*. Comput Biol Chem

912   2008;32:185-90.

913

914   59. Huang J, Hamasaki T, Ozoe F, Ohta H, Enomoto K, Kataoka H, et al.

915   Identification of critical structural determinants responsible for octopamine binding to

916   the alpha-adrenergic-like *Bombyx mori* octopamine receptor. Biochemistry.

917   2007;46:5896-903.

918

919   60. McVeigh P, Maule AG. Flatworm neurobiology in the postgenomic era. In: Byrne

920   JH, Editor. Oxford Handbook of Invertebrate Neurobiology. Oxford University Press,

921   2015.

922

923   61. Kruse AC, Hu J, Pan AC, Arlow DH, Rosenbaum DM, Rosemond E et al.

924   Structure and dynamics of the M3 muscarinic acetylcholine receptor. Nature.

925   2012;482:552-556.

926

927   62. MacDonald K, Kimber MJ, Day TA, Ribeiro P. A constitutively active G protein-

928   coupled acetylcholine receptor regulates motility of larval *Schistosoma mansoni*. Mol

929   Biochem Parasitol. 2016;202:29-37.

930

931 63. Wellendorph P, Brauner-Osborne H. Molecular basis for amino acid sensing by

932 family C G-protein-coupled receptors. Br J Pharmacol. 2009;156:869-84.

933

934 64. Orlandi C, Masuho I, Posokhova E, Cao Y, Ray T, Hasan N. GPR158 and

935 GPR179: a subfamily of orphan GPCRs as a new class of signaling modulators.

936 FASEB J. 2013; 27(1). Supplement 1095.2.

937

938 65. Orhan E, Prezeau L, El Shamieh S, Bujakowska KM, Michiels C, Zagar Y et al.

939 Further insights into GPR179: expression, localization and associated pathogenic

940 mechanisms leading to complete congenital stationary night blindness. Invest

941 Ophthalmol Vis Sci. 2013;54: 8041-8050.

942

943 66. Patel N, Itakura T, Jeong S, Liao CP, Roy-Burman P, Zandi E. Expression and

944 functional role of orphan receptor GPR158 in prostate cancer growth and

945 progression. PLoS One. 2015;10(2): e0117758.

946

947 67. Mitri C, Parmentier ML, Pin JP, Bockaert J, Grau Y. Divergent evolution in

948 metabotropic glutamate receptors. A new receptor activated by an endogenous

949 ligand different from glutamate in insects. J Biol Chem. 2004;279:9313-9320.

950

951 68. Mendonça-Silva DL, Gardino PF, Kubrusly RC, De Mello FG, Noël F.

952 Characterization of a GABAergic neurotransmission in adult *Schistosoma mansoni*.

953 Parasitology. 2004;129:137-146.

954

955 69. Janda CY, Waghray D, Levin AM, Thomas C, Garcia KC. Structural basis of Wnt

956 recognition by Frizzled. Science. 2012;337:59-64.

957

958   70. Li HF, Wang XB, Jin YP, Xia YX, Feng XG, Yang JM et al. Wnt4, the first

959   member of the Wnt family identified in *Schistosoma japonicum*, refulates worm

960   development by the canonical pathway. Parasitol Res. 2010;107:795-805.

961

962   71. Ta N, Feng X, Deng L, Fu Z, Hong Y, Liu J et al. Characterization and expression

963   analysis of Wnt5 in *Schistosoma japonicum* at different developmental stages.

964   Parasitol Res. 2015;114:3261-3269.

965

966   72. Willert K, Nusse R. Wnt proteins. Cold Spring Harb Perspect Biol.

967   2012;4(9):a007864.

968

969   73. Smallwood PM, Williams J, Xu Q, Leahy DJ, Nathans J. Mutational analysis of

970   Norrin-Frizzled4 recognition. J Biol Chem. 2007;282:4057-4068.

971

972   74. Ayers KL, Thérond PP. Evaluating Smoothened as a G-protein-coupled receptor

973   for Hedgehog signalling. Trends Cell Biol. 2010;20:287-298.

974

975   75. Gehrke AR, Srivastava M. Neoblasts and the evolution of whole-body

976   regeneration. Curr Opin Genet Dev. 2016;40:131-137.

977

978   76. Collins JJ 3rd, Wang B, Lambrus BG, Tharp ME, Iyer H, Newmark PA. Adult

979   somatic stem cells in the human parasite *Schistosoma mansoni*. Nature.

980   2013;494:476-479.

981

982   77. Koziol U, Rauschendorfer T, Zanon Rodriguez L, Krohne G, Brehm K. The

983   unique stem cell system of the immortal larvae of the human parasite *Echinococcus*

984   *multilocularis*. Evodevo. 2014;5(1):10. Doi: 10.1186/2041-9139-5-10.

985

986   78. Wang B, Collins JJ 3rd, Newmark PA. Functional genomic characterization of

987   neoblast-like stem cells in larval *Schistosoma mansoni*. Elife. 2013;2:e00768.

988

989   79. Gurley KA, Rink JC, Sanchez-Alvarado A. beta-catenin defines head versus tail

990   identity during planarian regeneration and homeostasis. Science. 2008;319:323-327.

991

992   80. Petersen CP, Reddien PW. Smed-betacatenin-1 is required for anterioposterior

993   blastema polarity in planarian regeneration. Science. 2008;319:327-330.

994

995   81. Langenhan T, Aust G, Hamann J. Sticky signaling – adhesion class G protein-

996   coupled receptors take the stage. Sci Signal. 2013;6:re3. Doi:

997   10.1126/scisignal.2003825.

998

999   82. Mousley A, Marks NJ, Halton DW, Geary TG, Thompson DP, Maule AG.

1000   Arthropod FMRFamide-related peptides modulate muscle activity in helminths. Int J

1001   Parasitol. 2004;34:755-768.

1002

1003   83. Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite – a

1004   comprehensive resource for helminth genomics. Mol Biochem Parasitol. 2017;215:2-

1005   10.

1006

1007   84. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics

1008   Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870-1874.

1009

1010   85. Tusnday GE, Simon I. Principals governing amino acid composition of integral

1011   membrane proteins: application to topology prediction. J Mol Biol. 1998;283:489-506.

1012

39

1013   86. Tusnday GE, Simon I. The HMMTOP transmembrane topology prediction server.

1014   Bioinformatics. 2001;17:849-850.

1015

1016   87. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment

1017   search tool. J Mol Biol. 1990;215:403-410.

1018

1019   88. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K et al.

1020   BLAST+: architecture and applications. BMC Bioinformatics. 2008;10:421.

1021

1022   89. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C et al. InterProScan 5:

1023   genome-scale protein function classification. Bioinformatics. 2014;

1024   doi:10.1093/bioinformatics/btu031

1025

1026   90. Finn RD, Clements J, Arndt W, Miller BL, Wheeler TJ, Schreiber F et al. HMMER

1027   web server: 2015 update. Nucleic Acids Res. 2015;43:W30-W38.

1028

1029   91. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence

1030   alignment, interactive sequence choice and visualization. Brief Bioinform. 2017; doi:

1031   10.1093/bib/bbx108.

1032

1033   92. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo

1034   generator. Genome Res. 2004;14:1188-1190.

1035

1036   93. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F et al.

1037   Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res.

1038   2008;36:W465-469.

1039

1040   94. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M et al. Synthetic spike-in

1041   standards for RNA-Seq experiments. Genome Res. 2011;21:1543-1551.

1042

1043   95. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient

1044   alignment of short DNA sequences to the human genome. Genome Biol.

1045   2009;10:R25. Doi: 10/1186/gb-2009-10-3-r25.

1046

1047   96. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with

1048   RNA-seq. Bioinformatics. 2009;**25**:1105–1111.

1049

1050   97. Trapnell C, Williams B, Pertea G, Mortazavi A, Kwan G, van Baren J et al.

1051   Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts

1052   and isoform switching during cell differentiation. Nat Biotechnol. 2010;

1053   doi:10.1038/nbt.1621.

1054

1055   98. Trapnell C, Hendrickson D, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential

1056   analysis of gene regulation at transcript resolution with RNA-Seq. Nat Biotechnol.

1057   2012; doi:10.1038/nbt.2450.

1058

1059   99. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR et al. Differential gene

1060   and transcript expression analysis of RNA-seq experiments with TopHat and

1061   Cufflinks. Nat Protoc. 2012;7:562-578.

1062

1063   100. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq

1064   expression estimates by correcting for fragment bias. Genome Biol. 2011;

1065   doi:10.1186/gb-2011-12-3-r22.

1066

1067

1068 101. Babicki S, Arndt D, Marcu A, Liang Y, Grant JR, Maciejewski A et al.

1069 Heatmapper: web-enabled heat mapping for all. Nucleic Acids Res. 2016;44:W147-

1070 W153.

1071

1072 102. Nakamura Y, Ishii J, Kondo A. Applications of yeast-based signaling sensor for

1073 characterization of antagonist and analysis of site-directed mutants of the human

1074 serotonin 1A receptor. Biotechnol Bioeng (2015) 112(9):1906-15.

1075

1076 103. Galvez T, Prezeau L, Milioti G, Franek M, Joly C, Froestl W, Bettler B, Bertrand

1077 HG, Blahos J, Pin JP. Mapping the agonist binding site of $GABA_B$ type 1 subunit

1078 sheds light on the activation process of $GABA_B$ receptors. J Biol Chem (2000) 275,

1079 41166-41174.

1080

1081 104. Geng Y, Bush M, Mosyak L, Wang F, Fan QR. Structural mechanism of ligand

1082 activation in human GABA(B) receptor. Nature (2013) 504(7479):254-9.

1083

1084 **Figure Captions**

1085 **Fig 1: Methods for discovery and annotation of *Fasciola hepatica* G protein**

1086 **coupled receptors (FhGPCRs).** (A) Hidden Markov Models (HMMs) representing

1087 glutamate, rhodopsin, adhesion, frizzled/smoothened and secretin families, and two

1088 rhodopsin subfamilies, were built from protein multiple sequence alignments of

1089 *Schistosoma mansoni* and *Schmidtea mediterranea* GPCRs [21]. HMMs were built

1090 and searched respectively using the *hmmbuild* and *hmmsearch* modules of HMMER

1091 v3.0. Searches were performed against two publically available *F. hepatica*

1092 genomes using *hmmsearch* and basic local alignment search tool (BLAST) tools.

1093 Each putative FhGPCR sequence was assessed for transmembrane (TM) domain

1094 composition with *hmmtop* before classification using tools including BLASTp,

1095 Interproscan and CLANS. (B) The largest proportion (49%) of FhGPCRs carried the

1096 full complement of 7 TMs, with 88% of sequences bearing at least 4 TMs. (C)

1097 GRAFS composition of 146 FhGPCRs carrying ≥4 TMs. (D) Rhodopsins were

1098 subject to further classification, including BLASTp vs datasets representing major

1099 non-flatworm animal phyla and superphyla. These rhodopsin homology

1100 classifications fed back into phylogenetic analyses versus deorphanised bilaterian

1101 GPCRs to confirm their putative ligand selectivity, with a final analysis of ligand

1102 binding domain composition comparing conservation of ligand interacting residues

1103 for characterised GPCRs reported in the literature with our *F. hepatica* assignments.

1104

1105 **Fig 2: Phylogenetic classification of *Fasciola hepatica* rhodopsin G protein-**

1106 **coupled receptors. (**A) Maximum-likelihood cladogram of *F. hepatica* rhodopsins.

1107 Phylogeny delineated clades containing rhodopsins with distinct homologies (RA,

1108 amine; RP, peptide; RO, opsin: R, orphan rhodopsin). The orphan clades contained

1109 sequences with generally low BLASTp similarity to their closest non-flatworm

1110 BLASTp hit, but concentrated within them were 18 sequences with exceptionally low

1111 (E>0.01) BLASTp similarity to non-flatworm sequences (fwRhods). The tree was

43

1112   midpoint rooted and was generated from a multiple protein sequence alignment

1113   trimmed to TM domains I-VII. Numbers at nodes indicate statistical support from

1114   approximate likelihood ratio test (aLRT). Tip colours are coded according to the E-

1115   value scale (as indicated) of that GPCR's closest BLASTp match in the ncbi nr

1116   database, excluding phylum Platyhelminthes. (B) Summary of sequence similarity

1117   comparisons between GPCRs within each rhodopsin clade, and their closest

1118   BLASTp hits in major phylogenetic groups (Basal: Cnidaria, Ctenophora, Porifera,

1119   Placozoa; superphylum Lophotrochozoa, omitting Platyhelminthes; Superphylum

1120   Ecdysozoa; superphylum Deuterostomia; phylum Platyhelminthes). BLASTp E-value

1121   (median) is summarised in each case, colour coded as a heat map on the same

1122   colour scale as (A). The number of GPCRs comprising each *F. hepatica* clade (*n*) is

1123   also indicated. (C) Sequence diversity within ubiquitous rhodopsin motifs of the

1124   majority (117) of the *F. hepatica* rhodopsins (upper panel), compared to those motifs

1125   in 18 *F. hepatica* fwRhods (lower panel). The mammalian consensus motifs are

1126   illustrated above the top panel, along with an illustration of the location of each motif

1127   within the rhodopsin 7TM domain structure. Some variability is visible within the TM2

1128   and TM6 motifs, but TM3 and TM7 motifs are well conserved.

1129

1130   **Fig 3: Identification of flatworm-specific rhodopsins (fwRhods) in genomes**

1131   **from phylum Platyhelminthes.** (A) The 18 *Fasciola hepatica* GPCRs in our dataset

1132   that had poor BLASTp similarity (E>0.01) to non-flatworm sequences in the ncbi nr

1133   dataset (lsGPCRs), were used as queries in BLASTp searches of flatworm genomes

1134   in WormBase Parasite (release WBPS9). All hits scoring E<0.01 were back-

1135   searched by BLASTp against our *F. hepatica* GPCR dataset. Sequences scoring

1136   E<0.01 against one of the original *F. hepatica* GPCRs were retained as matches.

1137   These sequences were then filtered to identify those lacking matches in ncbi nr,

1138   lacking non-GPCR protein domains, possessing at least 4 transmembrane (TM)

1139   domains, and containing rhodopsin motifs consistent with those seen in the majority

1140   of *F. hepatica* rhodopsins (see C). (B) This process identified 76 fwRhods in phylum

1141   Platyhelminthes, the majority (70) of which were from class Trematoda. Small

1142   numbers were returned from classes Cestoda and Monogenea. Note that no

1143   fwRhods fitting these criteria were identified in class Turbellaria. (C) Sequence

1144   diversity within ubiquitous rhodopsin motifs of 18 *F. hepatica* fwRhods (upper panel),

1145   compared to those motifs in the 58 fwRhods identified in the wider phylum (lower

1146   panel); motifs are broadly similar between *F. hepatica* and the rest of the phylum.

1147   (D) Maximum likelihood phylogeny of 76 fwRhods, alongside flatworm-specific

1148   rhodopsins described previously (70 platyhelminth rhodopsin orphan family 1

1149   (PROF1) [21,22], and 245 *S. mediterranea* G protein coupled receptor [GCRs,

1150   comprising RhoL, RhoR, Srfa, Srfb and Srfc families, reported as lacking non-

1151   flatworm homologues [24]) with branches coloured to indicate Family (dark blue,

1152   PROF1; mid blue, Srfa; cyan, Rho-L; green, Rho-R; orange, Srfb; purple, Srfc; red,

1153   fwRhod). Tree was rooted to a human rhodopsin (P08100) and was generated from

1154   an alignment trimmed to transmembrane domains I-VII. Numbers at nodes indicate

1155   statistical support from approximate likelihood ratio test (aLRT).

1156

1157   **Fig 4: Conservation of ligand-interacting residues between 17 *Fasciola***

1158   ***hepatica* G protein-coupled receptors (GPCRs) and structurally characterized**

1159   **homologues from other species.** (A) Neuropeptide F/Y receptor ligand binding

1160   residues as characterised by mutagenesis in human neuropeptide Y receptor NPY1R

1161   [39-43], and conserved in *Anopheles gambiae* (Ag) and *Drosophila melanogaster*

1162   (Dm) neuropeptide F receptors (NPFR) [44]. Numbering relative to HsNPY1R. (B)

1163   Serotonin (5-hydroxytryptamine; 5HT) receptor ligand binding residues as

1164   characterised by mutagenesis in human 5HT receptor (Hs5HT1A) [102], and

1165   conserved in *Schistosoma mansoni* 5HTR [51]. Numbering relative to Hs5HT1A.

1166   (C) Octopamine receptor (OaR) ligand binding residues as characterised by

1167   homology modelling of the *Periplaneta americana* (Pa) [58], and mutational analysis

1168   of the *Bombyx mori* (Bm) [59] octopamine receptor ligand binding domain.

1169   Numbering relative to PaOAR, except for Y412 which is shown relative to BmOAR.

1170   (D) Acetylcholine receptor ligand binding residues as characterised by homology

1171   modelling of the *S. mansoni* G protein-coupled acetylcholine receptor (SmGAR) [62];

1172   numbering relative to SmGAR. In each case, only *F. hepatica* sequences displaying

1173   at least 75% identity across the stated ligand binding residues are shown. Relative

1174   positions of residues across seven transmembrane domains (TM1-7) are shown. TM

1175   diagrams are not to scale.

1176

1177   **Fig 5: *Fasciola hepatica* glutamate G-protein coupled receptors (GPCRs)**

1178   **display divergent phylogeny and ligand binding domain (LBD) composition.**

1179   (A) Maximum likelihood phylogeny containing three *F. hepatica* glutamate receptors,

1180   alongside representative receptors from the various recognised GPCR Class C

1181   subgroups (subclasses indicated by blue boxes: Ca^{2+}, Ca^{2+}-sensing receptor;

1182 $GABA_B$, γ-aminobutyric acid type B receptors; mGluR, metabotropic glutamate

1183   receptors; Orphan, receptors with no known ligand; Taste, vertebrate taste

1184   receptors). Two previously reported *Schistosoma mansoni* glutamate receptors are

1185   also included; *F. hepatica* sequences are coloured red, *S. mansoni* are coloured

1186   blue, all others are black. Node numbers indicate statistical support as determined

1187   by approximate likelihood ratio test (aLRT). Tree was midpoint rooted. (B)

1188   Conservation of ligand-interacting residues between vertebrate $GABA_B$ and

1189   metabotropic glutamate receptors (mGluR), and *F. hepatica* class C GPCRs.

1190   Agonist-interacting residues were identified by multiple protein sequence alignment

1191   of *F. hepatica* glutamate receptors against mutationally-identified ligand interacting

1192   residues (those causing a significant reduction in receptor signalling activity), from

1193   mouse $GABA_B$ receptor (top panel), or selected human mGluR subtypes (lower

1194   panel). Identical amino acids in *F. hepatica*/*S. mansoni* GPCRs are represented by

1195   white text on black background, functionally conserved amino acids by black text on

1196   grey background. In lower panel, mutations causing a significant reduction in mGluR

1197   receptor activity are bold and numbered, with the region of the glutamate molecule

1198   bound by each residue indicated (COOH, C-terminus; NH_2, N-terminus). For

1199   references see [63,103,104].

1200

1201   **Fig 6: Frizzled/smoothened seven transmembrane receptors and wnt ligands in**

1202   ***Fasciola hepatica.*** (A) Maximum likelihood phylogeny containing six *F. hepatica*

1203   frizzled/smoothened receptors, alongside those from *Schistosoma mansoni,*

1204   *Drosophila melanogaster, Caenorhabditis elegans* and *Homo sapiens* (identified by

1205   FSMP, d, c and h, respectively). *F. hepatica* sequences are coloured red, *S.*

1206   *mansoni* are coloured blue, all other species are coloured black. Radial labels

1207   indicate human frizzled clusters (hClust) I-IV, and the smoothened clade. Node

1208   numbers indicate statistical support as determined by approximate likelihood ratio

1209   test (aLRT). Tree was rooted against a *Dictyostelium* frizzled sequence (dicty-fslJ-1).

1210   Tree composition adapted from [21]. (B) WebLogo comparison of ligand interacting

1211   residues between mouse fz1-10 (top panel) and *F. hepatica* frizzled receptors.

1212   Numbering in top panel x-axis is relative to mouse fz8 [69]. (C) Three wnt-like

1213   sequences exist in *F. hepatica.* Shading indicates positions of 22 characteristic Cys

1214   residues, positions numbered relative to *D. melanogaster* wnt-1 (Dro-wnt-1).

1215

1216   **Fig 7: Expression profiling of 101 G protein-coupled receptors (GPCRs) in**

1217   ***Fasciola hepatica* life stages.** (A) Expression heatmap generated from $\log_2$ FPKM

1218   values of 83 GPCRs identified from developmentally staged RNA-seq libraries. Life

1219   stages are represented in columns (Egg; Met, metacercariae; NEJ_1h, newly-

1220   excysted juvenile collected 1h post excystment; NEJ_3h, NEJ collected 3h post-

1221   excystment; NEJ_24h, NEJ collected 24h post-excystment; Juv_21d, liver stage

1222   juvenile parasites collected from murine livers 21 days following oral administration of

1223   metacercariae; Adult, adult parasites collected from the bile ducts of bovine livers).

47

1224   Rows indicate individual GPCRs, as denoted by the ID and phylogeny columns. The

1225   latter indicates receptor classification and predicted ligand where available (see S1

1226   Table). Expression cluster column indicates clusters of GPCRs with highest

1227   expression focused in particular life stages. (B) Detection of 76 GPCRs in Illumina

1228   RNA-Seq libraries generated from *F. hepatica* 21 day liver-stage juveniles, recovered

1229   *ex vivo* from rat infections. Data show expression of three glutamate (G), one

1230   adhesion (A), four frizzled (F), one smoothened (S) and 67 rhodopsin (R) GPCRs.

1231   The rhodopsins include representatives of amine (RA1, RA3), opsin (RO), peptide

1232   (RP1-7), and orphan (R2,3,4,6). Data points (each at *n*=3) represent mean $\log_2$

1233   FPKM ± 95% confidence intervals, as calculated by cuffdiff. In both panels, flatworm

1234   rhodopsins (fwRhods) are marked in red text. ACh, acetylcholine; AstA. Allatostatin

1235   A; Dop, dopamine; FMRFa, FMRFamide; GHS, growth hormone secretagogue;

1236   GYIRFa, GYIRFamide; Myom, myomodulin; Myos, myosuppressin; NPF/Y,

1237   neuropeptide F/Y; Oct, octopamine; Pkt, prokineticin; Tyr, tyramine; 5HT, 5-

1238   hydroxytryptamine.

1239

1240   **Supporting Information Captions**

1241   **S1 Table.** *Fasciola hepatica* **G protein coupled receptor (GPCR) dataset**

1242   **summary.** Table describes sequences containing 4-9 transmembrane (TM)

1243   domains. Each sequence is defined in terms of GRAFS family, and annotated for

1244   TM composition, sequence length, phylogeny, domain composition and homology

1245   relative to various datasets.

1246

1247   **S2 Text.** *Fasciola hepatica* **G protein-coupled receptor (GPCR) protein**

1248   **sequence dataset.**

1249

1250   **S3 Table. Flatworm-specific rhodopsins (fwRhods) in** *Fasciola hepatica* **and**

1251   **other flatworms.** Species and genome IDs of sequences that have ≥4

1252   transmembrane (TM) domains, and lack high-scoring orthologues in non-flatworms

1253   (BLASTp score E≥0.01 vs ncbi nr excluding Platyhelminthes), and show

1254   conservation of at least two of the four ubiquitous rhodopsin motifs. Each sequence

1255   is annotated for protein domains where present (Pfam HMMScan). Accessions refer

1256   to WormBase ParaSite.

1257

1258   **S4 Table. Rhodopsin ubiquitous motifs and ligand binding domains for ACh,**

1259   **NPF/Y 5-HT, octopamine.** Note data are in individual tabs. Rhodopsin: Sequence

1260   motifs extracted from alignment of *F. hepatica* rhodopsins, corresponding to

1261   ubiquitous rhodopsin motifs of TMs 2, 3, 6 and 7; Acetylcholine, NPF, 5-HT,

1262   Octopamine: Amino acids extracted from alignment of *F. hepatica* rhodopsins with

1263   mutationally or structurally-characterised GPCRs (comparators). Summary: Percent

1264   identity of ACh, NPF, 5-HT and Oct receptor LBDs, indicating most conserved LBD

1265   sequences showing at least 75% identity.

1266

1267 **S5 Fig. Phylogenetic comparison of *Fasciola hepatica* GPCRs with**

1268 **deorphanised bilaterian GPCRs.** (A) Peptide receptors (*F. hepatica* black or

1269 magenta as described in Fig x); (B) Amine receptors (*F. hepatica* dark blue); In all

1270 cases, non-flatworm receptors are coloured light blue. In (A) outer labels indicate

1271 positions of receptors for neuropeptide families previously reported in flatworms

1272 (McVeigh et al., 2009; Collins et al., 2010; Koziol et al., 2016); in (B), outer labels

1273 represent major groups containing phylogenetically similar *F. hepatica* sequences.

1274 Trees were midpoint rooted, maximum likelihood phylogenies of transmembrane

1275 domains I-VII. Numbers at nodes indicate statistical support from approximate

1276 likelihood ratio test (aLRT). Scale bars at the centre of each tree indicate number of

1277 substitutions per site. Abbreviations: ACh, acetylcholine; CCAP, crustacean

1278 cardioactive peptide; Dop, dopamine; FLP, FMRFamide-like peptide; GrH,

1279 gonadotropin-releasing hormone; Luq, luqin; Mmd, myomodulin; NKY, neuropeptide

1280 KY; NPF/Y, neuropeptide F/Y; Oct, octopamine; PK, pyrokinin; SIFa, SIFamide; Tyr,

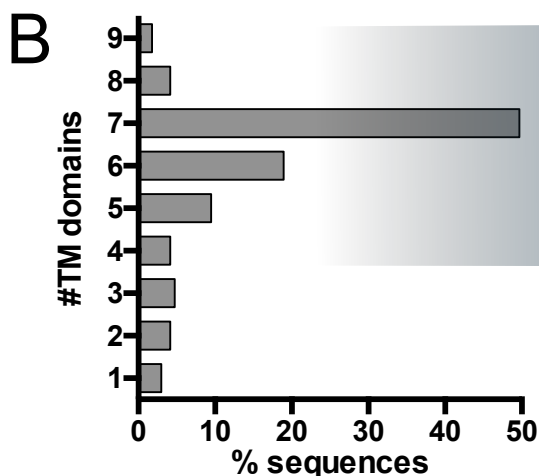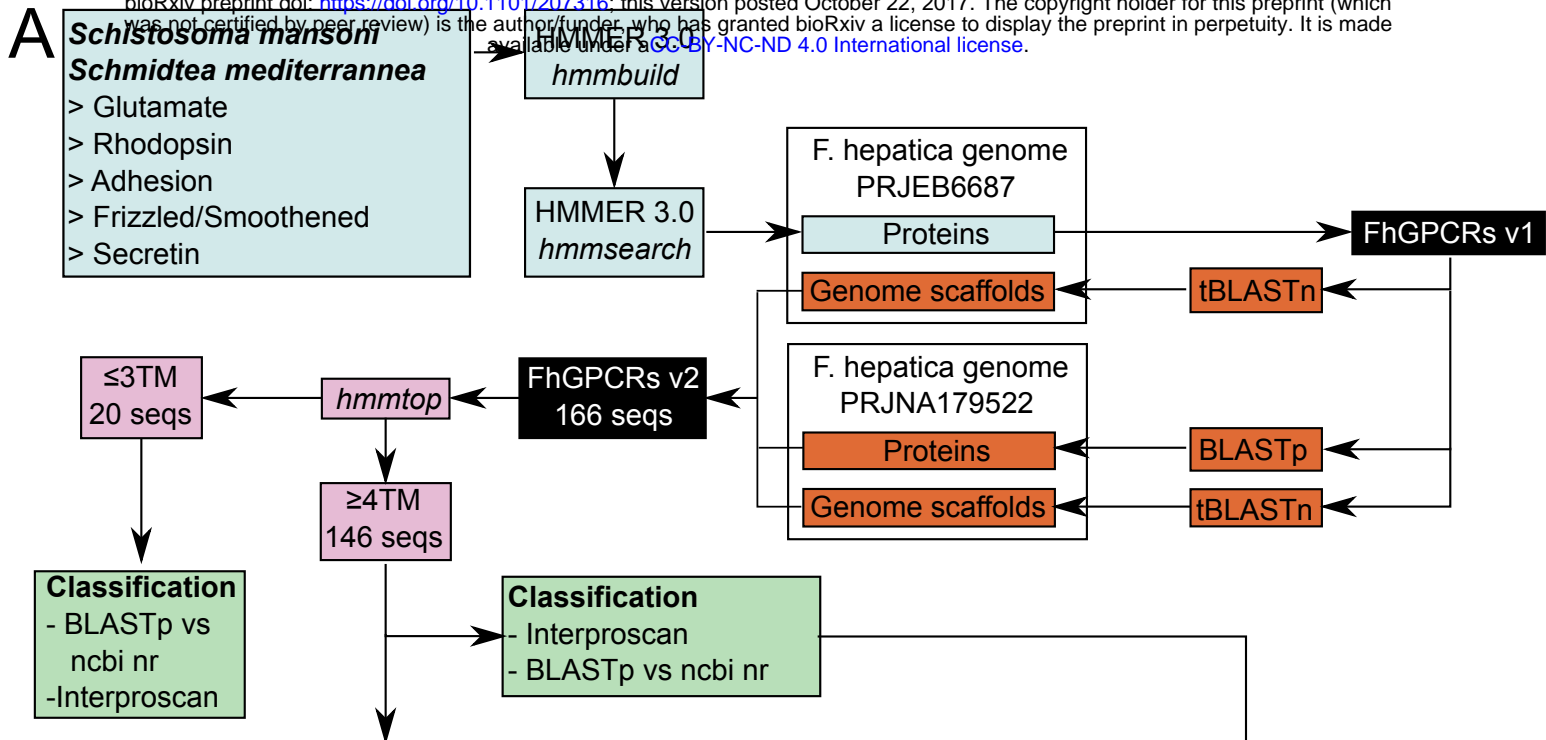1281 tyramine; SmGPR, schistosome GPCRs; 5HT, 5-hydroxytryptamine.

1282

1283 **S6 Table. Frizzled receptor ligand binding domain motifs.** Amino acids

1284 extracted from alignment of *F. hepatica* frizzled and smoothened GPCRs with

1285 mutationally characterised mouse fz8, also showing positionally-conserved residues

1286 in mouse fz1-10 (top panel) and *F. hepatica* frizzled receptors. Numbering in top row

1287 relative to mouse fz8 (Janda et al., 2012). Green boxes indicate identical residues in

1288 *F. hepatica* vs mammalian Fzd.

1289

1290 **S7 Figure: Adhesion receptor phylogeny.** Maximum likelihood phylogeny of

1291 *Fasciola hepatica* adhesion/secretin-like GPCRs alongside class B GPCRs from

1292 human and *Schistosoma mansoni*. Tree was a midpoint rooted, maximum likelihood

1293 phylogeny of transmembrane domains I-VII. Numbers at nodes indicate statistical

50

1294   support from approximate likelihood ratio test (aLRT). Scale bars at the centre of

1295   each tree indicate number of substitutions per site.

1296

A

Schistosoma mansoni
*Schmidtea mediterrannea*
> Glutamate
> Rhodopsin
> Adhesion
> Frizzled/Smoothened
> Secretin

HMMER 3.0
hmmbuild

HMMER 3.0
hmmsearch

F. hepatica genome
PRJEB6687
Proteins
Genome scaffolds

FhGPCRs v1

tBLASTn

F. hepatica genome
PRJNA179522
Proteins
Genome scaffolds

BLASTp

tBLASTn

≤3TM
20 seqs

hmmtop

FhGPCRs v2
166 seqs

≥4TM
146 seqs

Classification
- BLASTp vs ncbi nr
-Interproscan

Classification
- Interproscan
- BLASTp vs ncbi nr

B

#TM domains (y-axis: 1–9)
% sequences (x-axis: 0, 10, 20, 30, 40, 50)

C

4-9TM dataset summary
146 GPCRs
Glutamate   =   3
Adhesion   =   2
Frizzled   =   5
Smoothened   =   1
Rhodopsin   =   135

D

Rhodopsin subclassification (Figure 2)

RP: E<0.01 non-FW peptide GPCRs
RA: E<0.01 non-FW amine GPCRs
RO: E<0.01 non-FWopsin GPCRs
R: E≥0.01 nonFW sequences

> **FhRhodopsins BLASTp vs ncbi nr:**
- Non-flatworms
  Basal Phyla (Cnidaria, Ctenophora, Placozoa, Porifera)
- Lophotrochozoa
- Ecdysozoa
- Deuterostomia
- Phylum Platyhelminthes

> **Maximum Likelihood Phylogeny**
- vs Deorphanised Bilaterian Rhodopsins

> **Ubiquitous rhodopsin motif conservation**
- TMs 2,3,6,7

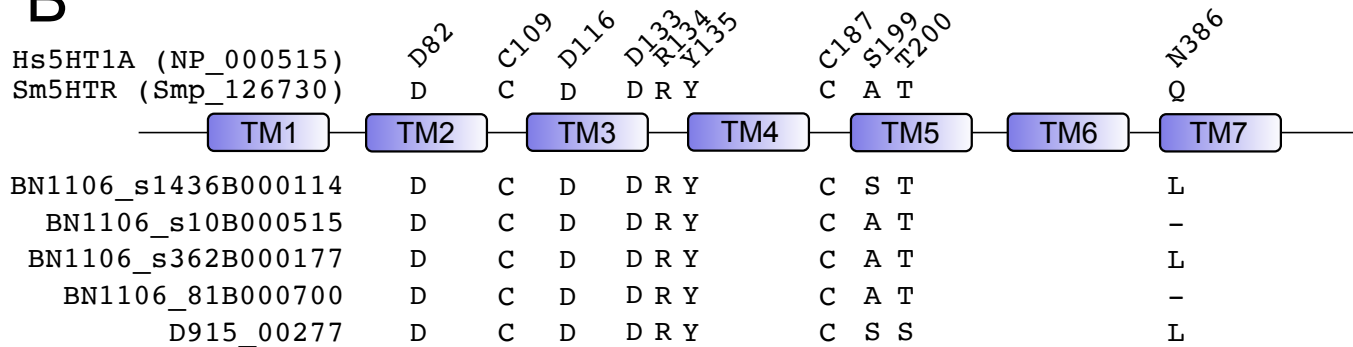> **Ligand binding domain conservation**
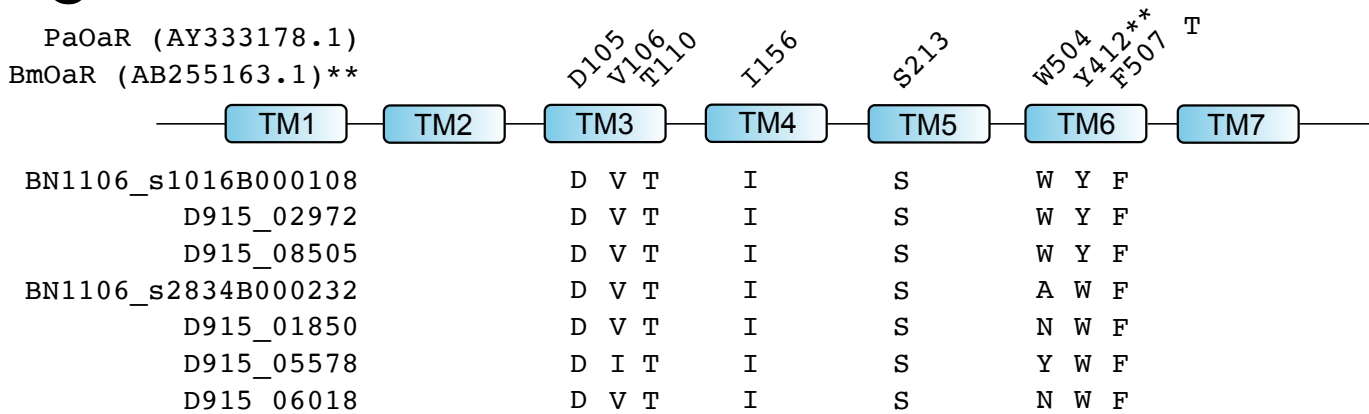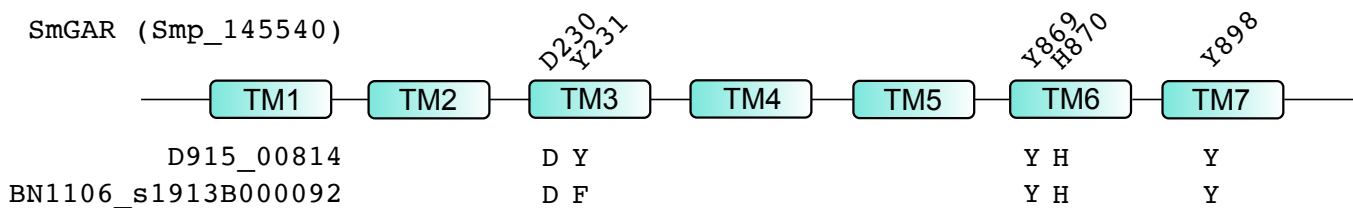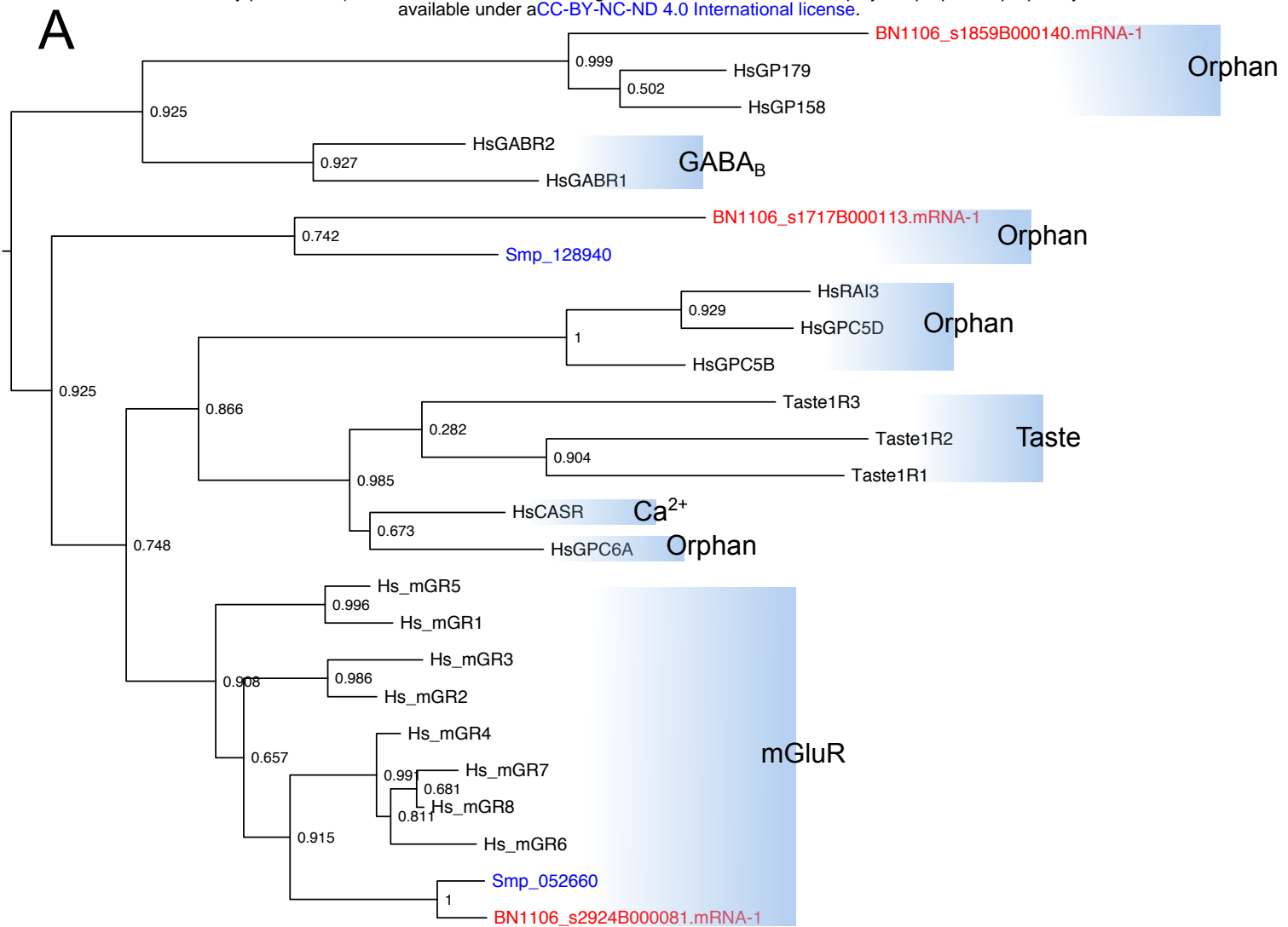- vs structurally characterised GPCRs

A

B
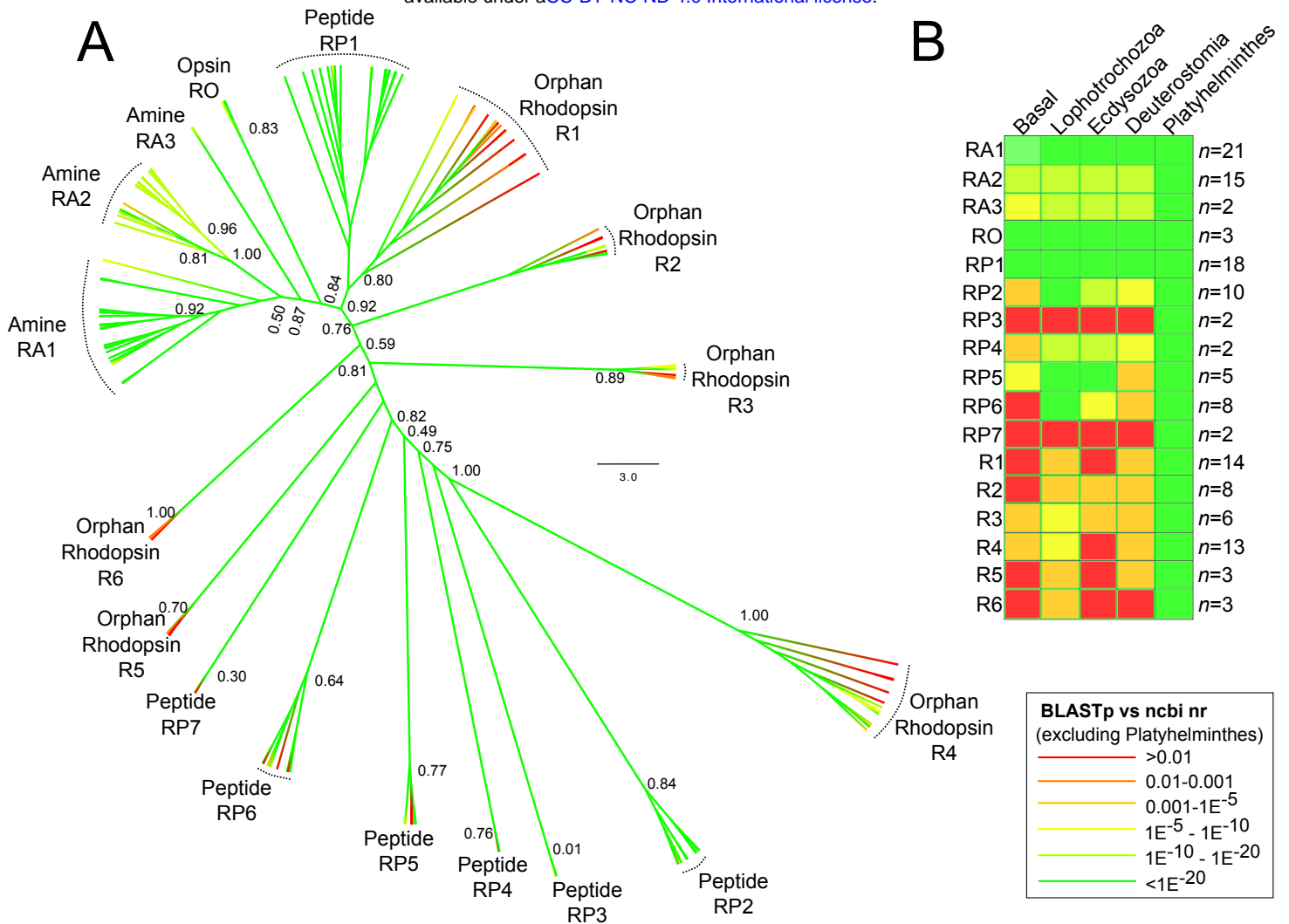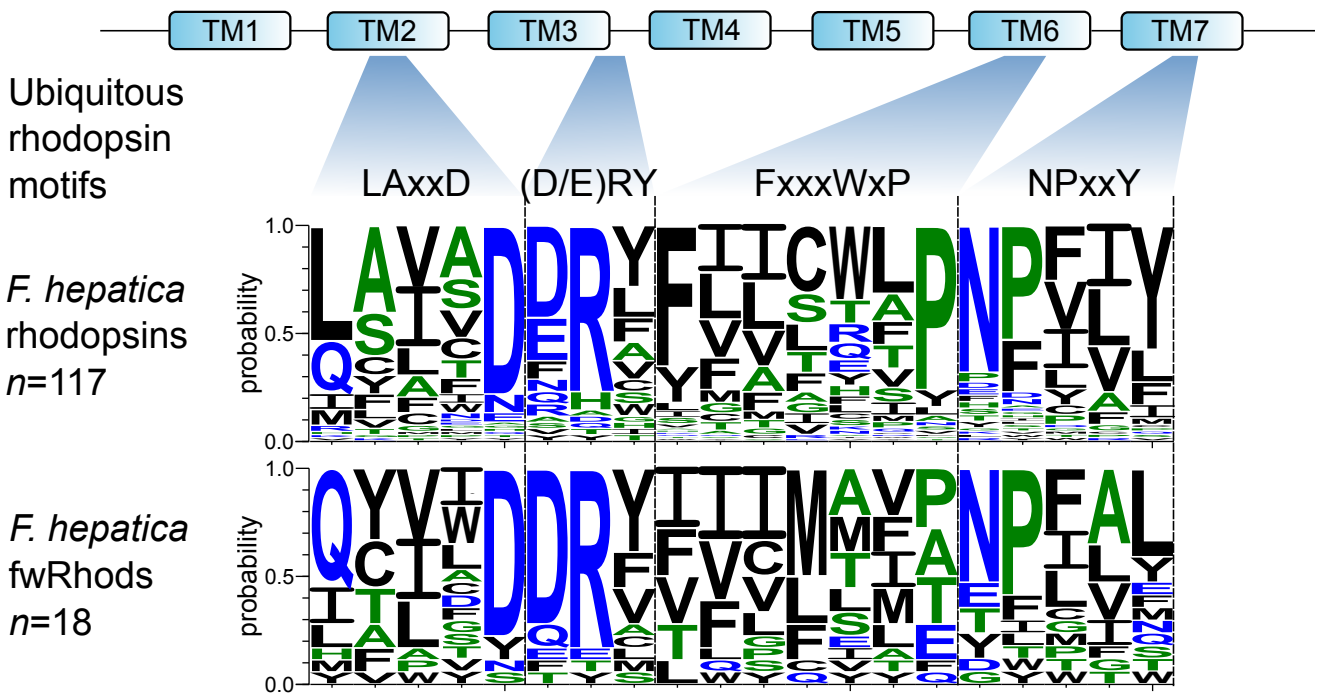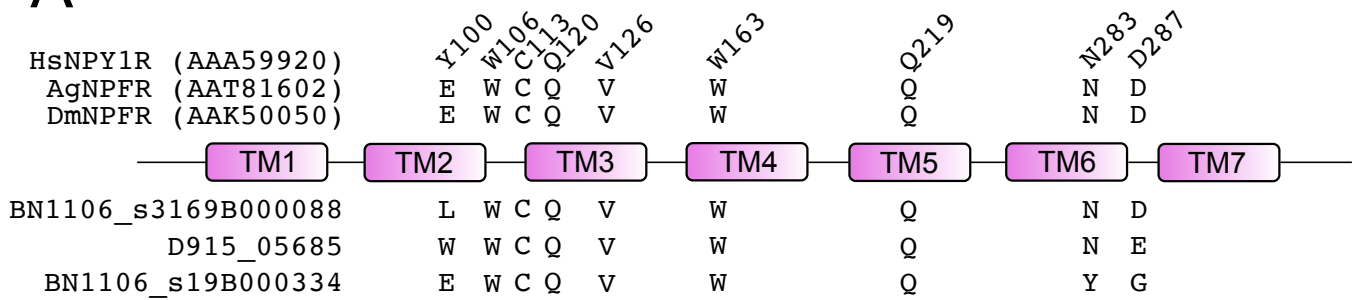
MmGABA$_{B1}$	S246	S269	S270	E465	D471
BN1106_s2924B000081	–	–	–	–	–
BN1106_s1717B000113	S	D	A	D	N
BN1106_s1859B000140	F	L	N	C	Y
Smp_128940	S	A	S	D	N
Smp_052660	S	T	T	E	D

Glu binding region:	COOH		αNH$_2$	αNH$_2$	αNH$_2$	αNH$_2$			αNH$_2$		COOH
mGluR1	Y74	R78	S	S165	T188	D208	Y236	E292	G293	D318	D	K409
mGluR2	R57	R	Y	S145	T168	D188	Y216	R	S	D295	R	K377
mGluR3	R64	R68	Y150	S151	T174	D194	Y222	R277	S	D301	Q306	K
mGluR4	K74	R78	G	S159	T182	D	Y	N	E	D	K	K
mGluR7	N	R	G	S	T	D	Y	N	D	D	K	K
mGluR8	K71	R75	A	S	T	D	Y	N	E	D309	K	K
BN1106_s2924B000081	–	–	–	–	–	–	–	–	–	–	–	–
BN1106_s1717B000113	I	F	I	S	D	A	Y	S	S	N	D	A
BN1106_s1859B000140	–	L	–	–	–	–	A	D	V	N	–	P
Smp_128940	N	L	Q	S	A	I	H	T	S	N	T	S
Smp_052660	R	R	Y	S	T	D	Y	G	M	D	E	K