HHS Public Access

Author manuscript

Ticks Tick Borne Dis. Author manuscript; available in PMC 2016 December 06.

Published in final edited form as:

Ticks Tick Borne Dis. 2016 July ; 7(5): 670–677. doi:10.1016/j.ttbdis.2016.02.014.

Prediction of G protein-coupled receptor encoding sequences from the synganglion transcriptome of the cattle tick, *Rhipicephalus microplus*

Felix D. Guerrero[a,*], **Anastasia Kellogg**[b], **Alexandria N. Ogrey**[b], **Andrew M. Heekin**[a], **Roberto Barrero**[c], **Matthew I. Bellgard**[c], **Scot E. Dowd**[d], and **Ming-Ying Leung**[b]

[a]USDA-ARS, Knipling-Bushland US Livestock Insect Research Laboratory, 2700 Fredericksburg Rd., Kerrville, TX 78028, USA

[b]The University of Texas at El Paso, 500W. University Avenue, El Paso, TX 79968, USA

[c]Centre for Comparative Genomics, Murdoch University, Perth 6150, WA, Australia

[d]Molecular Research DNA, 503 Clovis Rd., Shallowater, TX 79363, USA

Abstract

The cattle tick, *Rhipicephalus (Boophilus) microplus*, is a pest which causes multiple health complications in cattle. The G protein-coupled receptor (GPCR) super-family presents a candidate target for developing novel tick control methods. However, GPCRs share limited sequence similarity among orthologous family members, and there is no reference genome available for *R. microplus*. This limits the effectiveness of alignment-dependent methods such as BLAST and Pfam for identifying GPCRs from *R. microplus*. However, GPCRs share a common structure consisting of seven transmembrane helices. We present an analysis of the *R. microplus* synganglion transcriptome using a combination of structurally-based and alignment-free methods which supplement the identification of GPCRs by sequence similarity. TMHMM predicts the number of transmembrane helices in a protein sequence. GPCRpred is a support vector machine-based method developed to predict and classify GPCRs using the dipeptide composition of a query aminoacid sequence. These two bioinformatic tools were applied to our transcriptome assembly of the cattle tick synganglion. Together, BLAST and Pfam identified 85 unique contigs as encoding partial or full length candidate cattle tick GPCRs. Collectively, TMHMM and GPCRpred identified 27 additional GPCR candidates that BLAST and Pfam missed. This demonstrates that the addition of structurally-based and alignment-free bioinformatic approaches to transcriptome

[*]Corresponding author. Tel.: +1 830 792 0327; fax: +1 830 792 0314. Felix.Guerrero@ars.usda.gov (F.D. Guerrero).

Competing interests

The authors declare that there are no competing interests.

Author's contributions

FDG conceived the study, participated in the design, data collection, and analysis of the study and helped draft and revise the manuscript. AK and ANO wrote and tested the scripts and led the GPCR prediction and analysis and helped draft the manuscript. AH participated in the bioinformatic assembly and analysis of the transcriptome sequences and drafted the manuscript. SED conducted the sequencing and initial assembly of the transcriptome. ALT participated in the design and data collection and helped revise the manuscript. RB and MB participated in the Pfam analysis and GPCR annotation and helped revise the manuscript. MYL participated in the GPCR prediction and analysis and helped revise the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.ttbdis.2016.02.014.

annotation and analysis produces a greater collection of prospective GPCRs than an analysis based solely upon methodologies dependent upon sequence alignment and similarity.

Keywords

Cattle tick; *Rhipicephalus* (*Boophilus*) *microplus*; EST; Synganglion; Transcriptome; GPCR

1. Introduction

The southern cattle tick, *Rhipicephalus (Boophilus) microplus*, is the vector of pathogens that cause anaplasmosis and babesiosis in cattle (Bock et al., 2004). Cattle infected with these pathogens generally experience reduced milk production, a decrease in weight, and often death in immunologically naive hosts. Tick control strategies are, therefore, an essential part of livestock management practices. The application of chemical treatments remains central to tick control (George et al., 2004). However, resistance to new acaricides has historically appeared in ticks within a relatively few years after acaricide introduction (Kunz and Kemp, 1994).Presently, most acaricidal treatments target the nervous system of the tick (Lees and Bowman, 2007). A thorough understanding of the components of the tick nervous system, specifically signaling molecules and their receptors, would be integral to further the identification of new targets for development of these types of tick control technologies.

The central nervous system (CNS) of the tick is a condensed mass of fused nerve fibers known as the synganglion. The tick esophagus partitions the synganglion into two regions approximately 0.3–0.5 mm in size; the supraesophageal region lies anterior and dorsal to the esophagus and the slightly larger subesophageal region lies posterior and ventral to the esophagus (Szlendak and Oliver, 1992). The synganglion is further divided into an outer cortex, consisting primarily of neuronal cell bodies (perikarya) and an inner neuropile consisting of neuronal axons and dendrites (Prullage et al., 1992). The outer cortex contains the cell bodies of motor-associated neurons and the cell bodies of additional neurosecretory neurons. Axonal pathways from outer cortical neurons form tracts that innervate peripheral organs (Šimo et al., 2014).

Research on tick neurobiology has been slowed by difficulties in maintaining disease-free tick colonies and the lack of a suitable non-parasitic tick species to utilize as a model organism (Lees and Bowman, 2007). Despite the relative scarcity of research on the tick CNS, recent studies have contributed to our understanding of tick neurobiology. Using antibody staining methods, Šimoet al. (2009) identified 15 different immunoreactive compounds expressed in specific peptidergic neurons, endocrine cells, and adjacent secretory cells that were homologous to neuropeptides in insects and crustaceans. The same study also revealed two novel peripheral neuron clusters within the cheliceral and paraspiracular nerves. Christie et al. (2011) mined publicly available transcriptome datasets to identify novel neuropeptides in the tick, *Amblyomma variegatum*, and other chelicerates. Donohue et al. (2010) characterized the synganglion transcriptome of the American dog tick, *Dermacentor variabilis*. They identified cDNA sequences of fourteen putative

neuropeptides and five neuropeptide receptors, in addition to feeding- and mating-related transcripts expressed at various stages of female development. Bissinger et al. (2011) discovered differential expression of several neuropeptide and neuropeptide receptors during tick development. Lees et al. (2010) sequenced the transcriptome of the synganglion of *Rhipicephalus sanguineus* with special attention to identification of neural-specific receptor sequences. The study characterized several novel tar-gets from an acaricide target perspective: two glutamate-gated chloride channels, a leucokinin receptor, a nicotinic acetylcholine receptor, and a chitinase. The synganglion transcriptome from *Ixodes scapularis* female adult ticks was recently sequenced and annotated (Egekwu et al., 2014). Several transcripts were annotated as encoding neuropeptides, neuropeptide receptors, and neurotransmitter receptors. A proteomic study of *I. scapularis* revealed a diverse mix of neuropeptides that shared a close relationship with insect neuropeptides (Neupert et al., 2009).

An important class of signal-transducing receptors in eukaryotes is the G-protein coupled receptors (GPCRs). GPCRs are common drug targets in humans, as over 30% of prescribed medications target this receptor type (Liebmann, 2004). A few GPCRs have been studied in *R. microplus*, including an octopamine receptor (Baxter and Barker, 1999) subsequently shown to most likely be a type-1 tyramine receptor (Gross et al., 2015), a serotonin receptor confirmed in *R. microplus* adults (Chen et al., 2004), and a leukokinin-like GPCR identified in various developmental stages of *R. microplus* (Holmes et al., 2000) and functionally characterized in mammalian cell lines (Holmes et al., 2003). Dopamine, a GPCR substrate, was first identified in *R. microplus* synganglia and associated nerves (Binnington and Stone, 1977). A dopamine D-1 receptor in the salivary glands was later confirmed by Bowman and Sauer (2004). The GPCR database (GPCRDB http://www.gpcr.org/7tm/) has 53 curated GPCRs from the deer tick, *I. scapularis*, and we expect at least a similar number to exist in *R. microplus*. The progress toward obtaining and assembling the genome sequence of *R. microplus* has made several transcriptome datasets available for annotation and analysis (Bellgard et al., 2012). However, many short read next generation sequence datasets contain partial transcript sequences. Thus, analytical approaches that only examine full length transcripts for GPCR-encoding sequences will not be comprehensive. In our study, we sequenced and annotated the synganglion transcriptome of adult *R. microplus* using a Titanium 454 pyrosequencing approach, optimized for long read length. We used both sequence similarity-based and structural similarity-based approaches to predict GPCRs from the synganglion transcriptome and classify them into GPCR families based on the human GPCR classification model (Nordström et al., 2011). Our approach attempted to reliably predict GPCR-like sequences from both whole and partial transcript sequences.

2. Materials and methods

2.1. Ticks

R. microplus from Texas and Australia were used for this study. The Texas ticks were from the f32 laboratory generation of the Deutsch strain collected from an outbreak in Webb County, TX, USA in 2001 and reared in the laboratory since the original field collection. The Australian ticks were from the NRFS laboratory strain reared upon Hereford cattle at

the Biosecurity Tick Colony, Animal Research Institute, Yeerongpilly, Queensland, Australia (Stewart et al., 1982). Synganglia were dissected from mixed sex unfed adult *R. microplus* immobilized under phosphate-buffered saline (pH 7.0).Upon dissection, the Australian tick synganglia were immediately placed in a pre-chilled 1.5 ml microcentrifuge tube submerged in dry ice. When 80 synganglia were obtained, RNA*later* ICE (Life Technologies, Grand Island, NY, USA) was added according to the supplier's protocol and the material shipped on dry ice to the United States and stored at −80°C until processed. Two hundred Texas tick synganglia were dissected directly into RNA*later* (Life Technologies) and stored at −80°C according to the manufacturer's protocol.

2.2. RNA extraction procedures

Total RNA was extracted from the synganglia samples using the ToTALLY RNA Isolation Kit (Life Technologies) per manufacturer's recommendation after thawing on ice, centrifugation and removal of excess RNA*later* or RNA*later* ICE. The optional lithium chloride precipitation step suggested by the kit protocol was used to help remove genomic DNA from the RNA. Approximately 10 μg and 5 μg of total RNA was obtained from the Australian and Texas synganglia, respectively. Following agarose gel electrophoretic analysis of the RNA, RNA integrity was good but genomic DNA was detected in the samples, thus, the TURBO DNA-*free* kit (Life Technologies) was used per manufacturer's recommendation to enzymatically remove the genomic DNA. The MicroPoly(A)Purist Kit (Life Technologies) was used to purify polyadenylated RNA from each sample and the Just cDNA Kit (Stratagene, La Jolla, CA USA) was used to prepare cDNA for sequencing.

2.3. Transcriptome sequencing and bioinformatic analysis

The transcriptomes were sequenced by massively parallel pyrosequencing on a 454 GS FLX Titanium platform using DNA preparation and sequencing protocols as described by the manufacturer (Margulies et al., 2005). A total of 507,705 and 1,110,032 unassembled sequences were generated from the Australian and Texas cattle tick samples, respectively, and were submitted to the Short Read Archive of the National Center for Biotechnology Information (Australian: SRX146318 and Texas: SRX145659). Sequence assembly was performed using the MIRA assembler with the EST option (Chevreux et al., 2004). All resulting contigs and unassembled singletons (collectively referred to as unigenes) were used in subsequent analyses (Supplementary files 1 and 2). The Texas tick synganglion transcriptome contigs Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GEEZ00000000. The version described in this paper is the first version, GEEZ01000000. The Australian tick synganglion transcriptome contigs Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accession GEFA00000000. The version described in this paper is the first version, GEFA01000000. In our study, unigenes from the Australian tick samples received the prefix "AT" and unigenes from the Texas tick samples received the prefix "MT". Unigenes were annotated via similarity searches of the UniRef100 database. UniRef100 is generated from the UniProt knowledgebase and merges identical sequence fragments into a single entry, thus increasing the accuracy and speed of sequence homology searches (Bairoch et al., 2005; Suzek et al., 2007). Searches of the UniRef100 database were conducted with BLASTX, translating the query sequence into all six possible reading frames

and using an *E*-value cutoff of 1e − 07. (Altschul et al., 1990). Multiple sequence alignments were performed with MAFFT version 7 online (http://mafft.cbrc.jp/alignment/server/). There are several advanced alignment strategies available online and we selected the G-IN-i alignment option set recommended for sequences with global homology, and all other options set to default (Katoh and Standley, 2013). The unigenes from the Texas synganglion transcriptome were further analyzed (Fig. 1) with a custom open reading frame (ORF)-finding script that examined all 6 ORFs of each unigene and subsequently output any resulting protein coding sequence having a length ≥50 amino acids. The scripts are provided in Supplementary file 3. These ORFs were analyzed by Pfam release 27.0, pfam scan.pl version 1.5, which was downloaded via FTP (http://pfam.xfam.org/) to run locally with the default parameters. Pfam does not accept duplicate names for input sequences. As some unigene reading frames had several translated ORFs ≥50 amino acids and duplicate names might present a problem, a script was written to add the first 8 amino acids to the sequence ID of each ORF to make the IDs unique.

2.4. Identification of GPCR-like sequences through structural prediction

Custom scripts placed the 62,529 predicted ORF (from the custom ORF-finding script described above) sequences over 100 amino acids long into the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM/), utilizing the output format of “one line per protein”, and stored the output into a text file. TMHMM was used to predict the number and locations of transmembrane helices in the ORFs (Krogh et al., 2001). The TMHMM output data for each ORF sequence was then parsed according to length in amino acids and predicted number of helices (Fig. 2). ORFs with lengths between 100 and 234 amino acids, and containing at least 3 predicted trans-membrane helices were parsed into our dataset categorized as “Not Full Length GPCR Candidate ORF”. ORFs of at least 235 aminoacids long and containing at least 6 predicted helices were parsed into our dataset as candidate full-length GPCRs and designated for stop codon analysis. GPCRs contain 7 transmembrane segments. TMHMM is quite accurate but may not predict all transmembrane helical regions. Krogh et al. (2001) reported TMHMM predicted 97.5% of known helical regions in a set of 160 protein sequences with known topologies. Also, transcriptome datasets can contain sequencing or assembly errors. Requiring a minimum of 6 rather than 7 helices to retain an ORF ≥235 amino acids for further analysis is a conservative approach in this part of the prediction analysis in Fig. 2. Meruelo et al. (2012) reported mean values for aminoacid length of transmembrane helices and loops as 26 and 19, respectively. Using these values, an ORF with 6 transmembrane helices and 5 loops would contain at least 251 amino acids. Thus, we chose 235 amino acids as a conservative estimation for parsing a sequence with 6 transmembrane helices as a possible full length ORF. To optimize use of computational resources, ORFs ≥235 amino acids but only possessing 5 or fewer predicted helices were removed from consideration as GPCR candidates. We expect the high accuracy of TMHMM at helix prediction to minimize the number of authentic GPCRs that are eliminated by this amino acid length filter. One caveat is that proteins with an extended N- or C-terminus might be more prone to false elimination. For this candidate full-length set, the unigene nucleotide sequence associated with the ORF was examined for the presence of stop codons before the proposed initiator methionine codon and after the stop codon that breaks the ORF. We included this step as a precaution due to transcriptome datasets often having sequence errors

near the 5′ and 3′ termini. We wanted to increase the certainty of our designation of ORFs as encoding full-length proteins. Unigenes with ORFs containing ≥2 stop codons prior to the putative initiator methionine codon and ≥2 stop codons at the 3′ end of the ORF were parsed to the final category of "Full Length GPCR Candidate ORF". Unigenes with only one stop codon both before and after the protein sequence were parsed to the final category of "Possibly Full Length GPCR Candidate ORF". All the nucleotide sequences associated with the ORFs in the "Not Full Length GPCR Candidate ORF", "Possibly Full Length GPCR Candidate ORF", and "FullLength GPCR Candidate ORF" categories were submitted to GPCR-pred, a support vector machine (SVM) tool which performs three levels of prediction on possible GPCRs using the dipeptide composition of the given sequence (Bhasin and Raghava, 2004). Dipeptide composition is simply a 20 × 20 amino acid matrix that indicates the number of times each possible amino acid neighboring pair occurs in a query sequence (van Heel, 1991). GPCRPred allows GPCR prediction without the usage of topology and was downloaded from the OSDD Linux web site (http://osddlinux.osdd.net/repo/gpcrpred.deb) to analyze the script-predicted GPCR sequences. The predicted tick synganglion GPCRs were classified according to the GRAFS classification system, with comparisons to GPCR data from *Homo sapiens* and *Saccoglossus kowalevskii* as derived from Krishnan et al. (2013).

Contigs predicted to encode GPCRs by the automated BLASTX, Pfam, and GPCRP red were manually curated by BLASTX analysis on the NCBI website (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastx&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) using the nr database and default parameters. The top 20 hits of each contig's BLASTX search result were examined to determine the most likely identity of the ORF encoded by the contig.

3. Results and discussion

3.1. Characterization of the overall transcriptome

Transcripts from the Australian cattle tick synganglion samples were sequenced and assembled into 42,275 contigs with a mean length of 710 nucleotides (Supplementary file 1). Transcripts from the Texas cattle tick synganglion samples were sequenced and assembled into 43,468 contigs with a mean length of 725 nucleotides (Supplementary file 2). Out of 85,743 total contigs, 33,511 received significant BLASTX scores (*E*-value < 1e − 07) from searches of the Uniref100 database (Supplementary file 4). The most frequently occurring top hit species was *I. scapularis* (14,497 times) followed by *Amblyomma maculatum* (7513 times). Over 74% of the top hit species were from the genera of *Ixodes*, *Rhipicephalus, Amblyomma*, or *Dermacentor*.

As an overall characterization of the Texas cattle tick synganglion transcriptome, Pfam analysis was performed on the 294,260 predicted ORFs ≥50 amino acids resulting from the custom ORF-finding script, described above and presented in Supplementary file 3. Fig. 3 shows the most common Pfam clans while Supplementary file 5 contains the entire analysis result. The Beta Propeller (Beta propeller CL0186, 981 occurrences) and P-loop Containing Nucleoside Triphosphate Hydrolase Superfamily (P-loop NTPase CL0023, 825 occurrences) clans were found to be the most prevalent clans in the synganglion transcriptome. The Major

Facilitator Superfamily (MFS CL0015, 102 occurrences), Ion Channel (VIC) Superfamily (Ion channel CL0030, 72 occurrences), and Family AG Protein-coupled Receptor-like Superfamily (GPCR A CL0192, 72 occurrences) clans were of special interest to our research program aimed at finding targets for developing novel control technologies. Consequently, we sought to identify as many candidate G protein-coupled receptor-encoding contigs in our dataset as possible.

3.2. Identification of GPCR candidates in the Texas cattle tick synganglion transcriptome

The results from our BLASTX analysis of the 85,743 assembled synganglion transcriptome contigs are listed in Supplementary file 4, including *E*-value and Uniref100 top hit. Searching the top hit descriptions of Supplemental file 4 entries using the search term of "G-protein coupled receptor", we found 22 Texas tick and 14 Australia tick contigs meeting this criterion. These 36 contigs are collectively listed in Supplemental file 6. Only 2 of the entries in this set of GPCR-like contigs hit to a non-tick sequence, likely because the BLASTX analysis is based upon sequence similarity. Interestingly, 9 of the 22 Texas tick contigs were not identified in the Pfam analysis discussed above as belonging to the GPCR clan (data not shown). This illustrated a problem in analyzing large datasets with a single in silico method. At this stage, we decided to focus upon the Texas cattle tick synganglion dataset for two main reasons. First, the Texas cattle tick species, *R. microplus*, is the top priority species for our research team. Second, the cattle ticks from Australia have recently been reclassified to *Rhipicephalus australis* (Estrada-Peñaet al., 2012).

In an attempt to predict and annotate as many GPCRs as possible, we wished to supplement our sequence similarity-based BLASTX and Pfam searches for GPCRs with a structural similarity-based search. Zamanian et al. (2011) used a transmembrane prediction-oriented approach to mine the newly available genome sequence of the human blood fluke, *Schistosoma mansoni*, and the planarian, *Schmidtea mediterranea*, for sequences encoding putative GPCRs. This group used the new genome sequences to predict the proteome of their target organism, using a Hidden Markov Model-based protocol to predict transmembrane helices in the proteome. Full length or nearly full length ORFs (based on the number of predicted transmembrane domains) were examined with BLASTP to identify homology to GPCRs. We based our search for cattle tick synganglion putative GPCRs on this approach. Additionally, the TMHMM prediction method has been shown to be capable of predicting transmembrane helices with high accuracy (>97%) and a low incidence of false positive and false negative predictions (Krogh et al., 2001). Thus, we chose the structurally-based approach of the TMHMM script (http://www.cbs.dtu.dk/services/TMHMM/) to search the Texas synganglion transcriptome for proteins containing predicted transmembrane helices and used this information to help assemble a dataset of cattle tick synganglion GPCR candidates (Fig. 2).

We directed all 43,468 Texas synganglion contigs into our custom ORF-finding script described in Section 2.3 and shown in Supplementary file 3. The output from our ORF-finding script was used as input to the TMHMM server and the resulting TMHMM server output was parsed based on the ORF length and the number of TMHMM-predicted transmembrane helices. ORFs were analyzed for stop codons as described above and parsed

into our dataset as GPCR candidates in categories noted as Full Length GPCR Candidate ORF, Possibly Full Length GPCR Candidate ORF, and Not Full Length GPCR Candidate ORF. After analysis, these categories contained 32, 72, and 742 candidate GPCR-encoding ORFs, respectively (Supplementary file 7). TMHMM script-predicted GPCRs were submitted to the GPCRpred tool which predicted 26 out of the 32 “FullLength GPCRs” to be GPCRs. Of the “Possible Full Length GPCRs”, 49 out of 72 of the sequences were predicted by GPCRpred to be GPCRs. Only 50 of the 742 “Not Full Length GPCRs” were predicted as GPCRs. In total, GPCRpred predicted 125 of the 846 script/TMHMM-predicted sequences as GPCRs. We must note the predictions for the “Not Full Length GPCRs” are considered less reliable, as the GPCRpred SVM was trained on full length sequences only and is not optimal for GPCR predictions using partial ORFs (Bhasin and Raghava, 2004). Combining the predicted GPCRs from the TMHMM-GPCRpred analyses with those from the BLASTX-Pfam analysis of all the synganglion transcriptome contigs resulted in a dataset of 351 contigs encoding candidate GPCRs expressed in the Texas cattle tick synganglion. These 351 contigs and the associated BLASTX, Pfam, custom scripts, and GPCRpred results are presented in Supplementary file 8. Supplementary file 8 lists each of these candidate GPCRs by contig number and gives annotation information, including method of prediction, BLASTX *E*-value, best hit defline, best hit species, and GPCR family classification.

Each of these 351 contigs was manually curated by manual BLASTX analysis against NCBI’s nr database and careful examination of the BLASTX top 20 hits and *E*-values resulting from each search. Contigs with BLASTX hits to a known non-GPCR sequence at *E*-value < 1.00E − 75 were removed from consideration as encoding a candidate GPCR. This resulted in a final dataset of 112 candidate GPCR-encoding contigs. Supplementary file 8 is arranged to show which of the original 351 candidate GPCRs were removed from the candidate list by the manual curation process. Some of these 112 final candidate GPCRs were predicted by more than 1 of the BLASTX, Pfam, and TMHMM/GPCRPred methodologies. Fig. 4 is a Venn diagram showing the number of GPCR candidates predicted by each method and the overlaps between prediction methods. Fig. 4 demonstrates the value of the three-pronged (BLASTX, Pfam, and TMHMM/GPCRPred) analytical approach that uses both sequence-and structural-based prediction of GPCRs. There is over-lap between the method results, as 39 contig candidates were predicted to be GPCRs by both BLASTX and Pfam, 3 candidates were predicted by both TMHMM/GPCRPred and Pfam, and 11 candidates were predicted by all three approaches. However, 27, 20, and 12 candidates were predicted only by TMHMM/GPCRPred, BLASTX, and Pfam methodologies, respectively.

As Krishnan et al. (2013) had compared the GRAFS family classification of GPCRs from hemichordate, echinoderm, and chordate species, we compared the GPCRs from the *R. microplus* synganglion to *H. sapiens* and the hemichordate *S. kowalevskii* (Fig. 5). GPCRs from the Rhodopsin family predominated in all three species. Rhodopsins make up 70%, 89%, and 82% of the GPCRs from *R. microplus* synganglia, *H. sapiens*, and *S. kowalevskii*, respectively. *R. microplus* appeared to possess proportionally more from the Secretin family (16%) than both *H. sapiens* (2%) and *S. kowalevskii* (0.4%), and this might be due to the nature of the synganglion requiring more of the Secretin family of GPCRs. Secretin plays a central role in water balance and cattle ticks ingest significant amounts of blood, requiring

rapid movement and elimination of water to concentrate blood components of nutritional value. Perhaps this is the reason that the Secretin family of GPCRs is so relatively abundant in our dataset.

A recent study by Richards et al. (2015) used transcriptome data to predict transmembrane proteins from cattle tick nymphs, larvae, and adult female ovaries, salivary glands, and midguts. There was no overlap between their transmembrane protein dataset and our set of 112 candidate GPCR-encoding contigs. This is not surprising, as their study did not use neural tissues and their transcriptome dataset was dominated by sequences derived from a cattle tick Sanger EST-based transcriptome, BmiGI Ver. 2.1 (Wang et al., 2007), that contained approximately 13,500 assembled contigs. The pooled RNA sample that was used to produce the cDNA library and ultimately the ESTs for BmiGI Ver. 2.1 contained predominantly larval material and lacked dissected synganglia. The Sanger technology did not yield the deep sequencing that current technologies produce. Transcripts of lesser abundance, such as GPCRs, would be less likely to appear in Sanger-based datasets. There is certainly overlap between the predicted identities of our overall synganglion transcriptome Contig dataset (Supplementary files 1 and 2) and the dataset from Richards et al. (2015), however not with our dataset of predicted GPCRs. Bissinger et al. (2011) reported the 454 pyrosequencing-derived synganglion transcriptome from the American dog tick, *D. variabilis*, and there was considerable over-lap in the identified transcripts from their dataset and ours. We did not have access to their complete GO analysis terms. However, they discovered transcripts encoding ORFs with sequence similarities to dopamine, octopamine, and muscarinic acetylcholine receptors, all of which are GPCRs.

Thus, our study of the transcriptome from the synganglia of the one-host cattle tick, *R. microplus*, is a significant addition to the tick transcriptomic information gleaned from Bissinger et al. (2011) for *D. variabilis* and Lees et al. (2010) from *R. sanguineus*. In addition, we have used methods based upon both sequence similarity and predicted protein structure to develop a database of candidate GPCRs from the synganglion of *R. microplus*. Table 1 presents the non-redundant set of predicted GPCRs and GPCR signaling-related ORFs obtained from our analysis of the Texas *R. microplus* synganglion transcriptome. Table 1 shows predicted protein name and the contig(s) found to encode the GPCR. We found 26, 6, and 7 GPCR candidates whose identity clearly represented the Rhodopsin, Secretin, and Glutamate GPCR families. In addition, there were 32, 7, and 1 contigs that were predicted to be in the Rhodopsin, Secretin, and Glutamate GPCR families but a protein identity could not be determined and they remain orphaned GPCR candidates. Two predicted GPCRs could not be assigned to family. Also in Table 1, we listed contigs predicted to encode proteins that are related to the GPCR signaling pathway, ORFs representing G proteins, GPCR kinase, adenylyl cyclase, inositol triphosphate receptor, and arrestin. This represents the first comprehensive dataset of candidate GPCRs from *R. microplus*, a species with global impacts upon animal health and economics of farmers and ranchers. This dataset can be a valuable asset to further studies in tick neurobiology with specific relevance to research into development of novel tick control technologies.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This article reports the results of research only. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation of endorsement by the U.S. Department of Agriculture. The authors wish to thank Dr. Louise Jackson (Biosecurity Queensland, QDAFF), Dr. Ala Lew-Tabor (The University of Queens-land, St. Lucia, Qld, Australia) and Mr. Jason Tidwell (USDA-ARSCFTRL, Edinburg, TX) for help with obtaining tick materials. We also acknowledge technical support for the Pfam analysis from Mr. Gerardo Cardenas and Mr. Sergio Munoz. FDG and AMH acknowledge funding support from USDA-ARS Knipling-Bushland US Livestock Insects Research Laboratory CRIS project 6205-32000-031-00. AK, ANO and MYL acknowledge funding support from USDA-NIFA-HSI grant 2012-38422-19910 and NIH grants 5G12RR008124 and 5G12MD007592. FDG acknowledges the receipt of a fellowship from the OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems in 2009. USDA is an equal opportunity provider and employer.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990; 215:403–410. [PubMed: 2231712]

Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckman B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Natale DA, O'Donovan C, Redaschi N, Yeh LSL. The universal protein resource (UniProt). Nucl. Acids Res. 2005; 33:D154–D159. [PubMed: 15608167]

Baxter GD, Barker SC. Isolation of a cDNA for an octopamine-like, G-protein coupled receptor from the cattle tick, *Rhipicephalus microplus*. Insect Biochem. Mol. Biol. 1999; 32:815–820.

Bellgard M, Moolhuijzen PM, Guerrero FD, Schibeci D, Rodriguez-Valle M, Peterson DG, Dowd SE, Barrero R, Hunter A, Miller RJ, Lew-Tabor AE. Cattle Tick Base: an integrated Internet-based bioinformatics resource for *Rhipicephalus (Boophilus) microplus*. Int. J. Parasitol. 2012; 42:161–169. [PubMed: 22178513]

Bhasin M, Raghava GPS. GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors. Nucl. Acids Res. 2004; 32(Suppl. 2):W383–W389. [PubMed: 15215416]

Binnington KC, Stone BF. Distribution of catecholamines in the cattle tick *Boophilus microplus*. Comp. Biochem. Physiol. C: Comp. Pharmacol. 1977; 58:21–28.

Bissinger BW, Donohue KV, Khalil SMS, Grozinger CM, Sonenshine DE, Zhu J, Roe RM. Synganglion transcriptome and developmental global gene expression in adult females of the American dog tick, *Dermacentor variabilis* (Acari: Ixodidae). Insect Mol. Biol. 2011; 20:465–491. [PubMed: 21689185]

Bock R, Jackson L, de Vos A, Jorgensen W. Babesiosis of cattle. Parasitology. 2004; 129(Suppl):247–269.

Bowman AS, Sauer JR. Tick salivary glands: function, physiology and future. Parasitology. 2004; 129:S67–S81. [PubMed: 15938505]

Chen A, Holmes SP, Pietrantonio PV. Molecular cloning and functional expression of a serotonin receptor from the Southern cattle tick, *Boophilus microplus* (Acari: Ixodidae). Insect Mol. Biol. 2004; 13:45–54. [PubMed: 14728666]

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. 2004; 14:1147–2115. [PubMed: 15140833]

Christie AE, Nolan DH, Ohno P, Hartline N, Lenz PH. Identification of chelicerate neuropeptides using bioinformatics approaches of publicly accessible expressed sequence tags. Gen. Comp. Endocrinol. 2011; 170:144–155. [PubMed: 20888826]

Donohue KV, Khalil SMS, Ross E, Grozinger CM, Sonenshine DE, Roe RM. Neuropeptide signaling sequences identified by pyrosequencing of the American dog tick synganglion transcriptome

during blood feeding and reproduction. Insect Biochem. Mol. Biol. 2010; 40:79–90. [PubMed: 20060044]

Egekwu N, Sonenshine DE, Bissinger BW, Roe RM. Transcriptome of the female synganglion of the black-legged tick *Ixodes* scapularis (Acari: Ixodidae) with comparison between Illumina and 454 systems. PLoS ONE. 2014; 9:e102667. [PubMed: 25075967]

Estrada-Peña A, Venzal JM, Nava S, Mangold A, Guglielmone AA, Labruna MB, de la Fuente J. Reinstatement of *Rhipicephalus (Boophilus) australis* (Acari: Ixodidae) with redescription of the adult and larval stages. J. Med. Entomol. 2012; 49:794–802. [PubMed: 22897039]

George JE, Pound JM, Davey RB. Chemical control of ticks on cattle and the resistance of these parasites to acaricides. Parasitology. 2004; 129:S353–S366. [PubMed: 15938518]

Gross AD, Temeyer KB, Day TA, Perez de Leon AA, Kimber MJ, Coats JR. Pharmacological characterization of a tyramine receptor from the southern cattle tick, *Rhipicephalus (Boophilus)* microplus. Insect Biochem. Mol. Biol. 2015; 63:47–53. [PubMed: 25958152]

Holmes SP, He H, Chen AC, Ivie GW, Pietrantonio PV. Cloning and transcriptional expression of a leucokinin-like peptide receptor from the Southern cattle tick, *Boophilus microplus* (Acari: Ixodidae). Insect Mol. Biol. 2000; 9:457–465. [PubMed: 11029664]

Holmes SP, Barhoumi R, Nachman RJ, Pietrantonio PV. Functional analysis of a G protein-coupled receptor from the Southern cattle tick *Boophilus microplus* identifies it as the first arthropod myokinin receptor. Insect Mol. Biol. 2003; 12:27–38. [PubMed: 12542633]

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 2013; 30:772–780. [PubMed: 23329690]

Krishnan A, Sällman Almén M, Fredriksson R, Schiöth HB. Remarkable similarities between the hemichordate (*Saccoglossus kowalevskii*) and vertebrate GPCR repertoire. Gene. 2013; 526:122–133. [PubMed: 23685280]

Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 2001; 305:567–580. [PubMed: 11152613]

Kunz SE, Kemp DH. Insecticides and acaricides: resistance and environmental impact. Rev. Sci. Tech. 1994; 13:1249–1286. [PubMed: 7711312]

Lees K, Bowman AS. Tick neurobiology: recent advances and the post-genomic era. Invert. Neurosci. 2007; 7:183–198. [PubMed: 17962985]

Lees K, Woods DJ, Bowman AS. Transcriptome analysis of the synganglion from the brown dog tick, *Rhipicephalus sanguineus*. Insect Mol. Biol. 2010; 19:273–282. [PubMed: 20002796]

Liebmann C. G protein-coupled receptors and their signaling pathways: classical therapeutical targets susceptible to novel therapeutic concepts. Curr. Pharm. Des. 2004; 10:1937–1958. [PubMed: 15180530]

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005; 437:376–380. [PubMed: 16056220]

Meruelo AD, Han SK, Kim S, Bowie JU. Structural differences between thermophilic and mesophilic membrane proteins. Protein Sci. 2012; 21:1746–1753. [PubMed: 23001966]

Neupert S, Russell WK, Predel R, Russell DH, Strey OF, Teel PD, Nachman RJ. The neuropeptidomics of *Ixodes scapularis* synganglion. J. Proteomics. 2009; 72:1040–1045. [PubMed: 19540946]

Nordström KJV, Sällman Almén M, Edstram MM, Fredriksson R, Schiöth HB. Independent HH search, Needleman–Wunsch-based, and motif analyses reveal the overall hierarchy for most of the G protein-coupled receptor families. Mol. Biol. Evol. 2011; 28:2471–2480. [PubMed: 21402729]

Prullage JB, Pound JM, Meola SM. Synganglion morphology and neurosecretory centres of adult *Amblyomma americanum* (L.) (Acari: Ixodidae). J. Med. Entomol. 1992; 29:1023–1034. [PubMed: 1460618]

Richards SA, Stutzer C, Bosman A-M, Maritz-Olivier C. Transmembrane proteins—mining the cattle tick transcriptome. Ticks Tick-Borne Dis. 2015; 6:695–710. [PubMed: 26096851]

Šimo L, Slovak M, Park Y, Žitňan D. Identification of a complex peptidergic neuroendocrine network in the hard tick, *Rhipicephalus appendiculatus*. Cell Tissue Res. 2009; 335:639–655. [PubMed: 19082627]

Šimo, L.; Sonenshine, DE.; Park, Y.; Žitňan, D. Nervous and sensory systems:structure, function, genomics, and proteomics. In: Sonenshine, DE.; Roe, RM., editors. Biology of Ticks. Vol. 1. Oxford: Oxford University Press; 2014. p. 309-367.

Stewart NP, Callow LL, Duncalfe F. Biological comparisons between a laboratory-maintained and a recently isolated field strain of *Boophilus microplus*. J. Parasitol. 1982; 68:691–694. [PubMed: 7119993]

Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef:comprehensive and non-redundant UniProt reference clusters. Bioinformatics. 2007; 23:1282–1288. [PubMed: 17379688]

Szlendak E, Oliver JH. Anatomy of synganglia, including their neurosecretory regions, in unfed, virgin female *Ixodes scapularis* (Acari:Ixodidae). J. Morphol. 1992; 213:349–364. [PubMed: 1404406]

van Heel M. A new family of powerful multivariate statistical sequence analysis techniques. J. Mol. Biol. 1991; 220:877–887. [PubMed: 1880802]

Wang M, Guerrero FD, Pertea G, Nene VM. Global comparative analysis of ESTs from the southern cattle tick, *Rhipicephalus (Boophilus) microplus*. BMC Genomics. 2007; 8:368. [PubMed: 17935616]

Zamanian M, Kimber MJ, McVeigh P, Carlson SA, Maule AG, Day TA. The repertoire of G protein-coupled receptors in the human parasite *Schistosoma mansoni* and the model organism *Schmidtea mediterranea*. BMC Genomics. 2011; 12:596. [PubMed: 22145649]

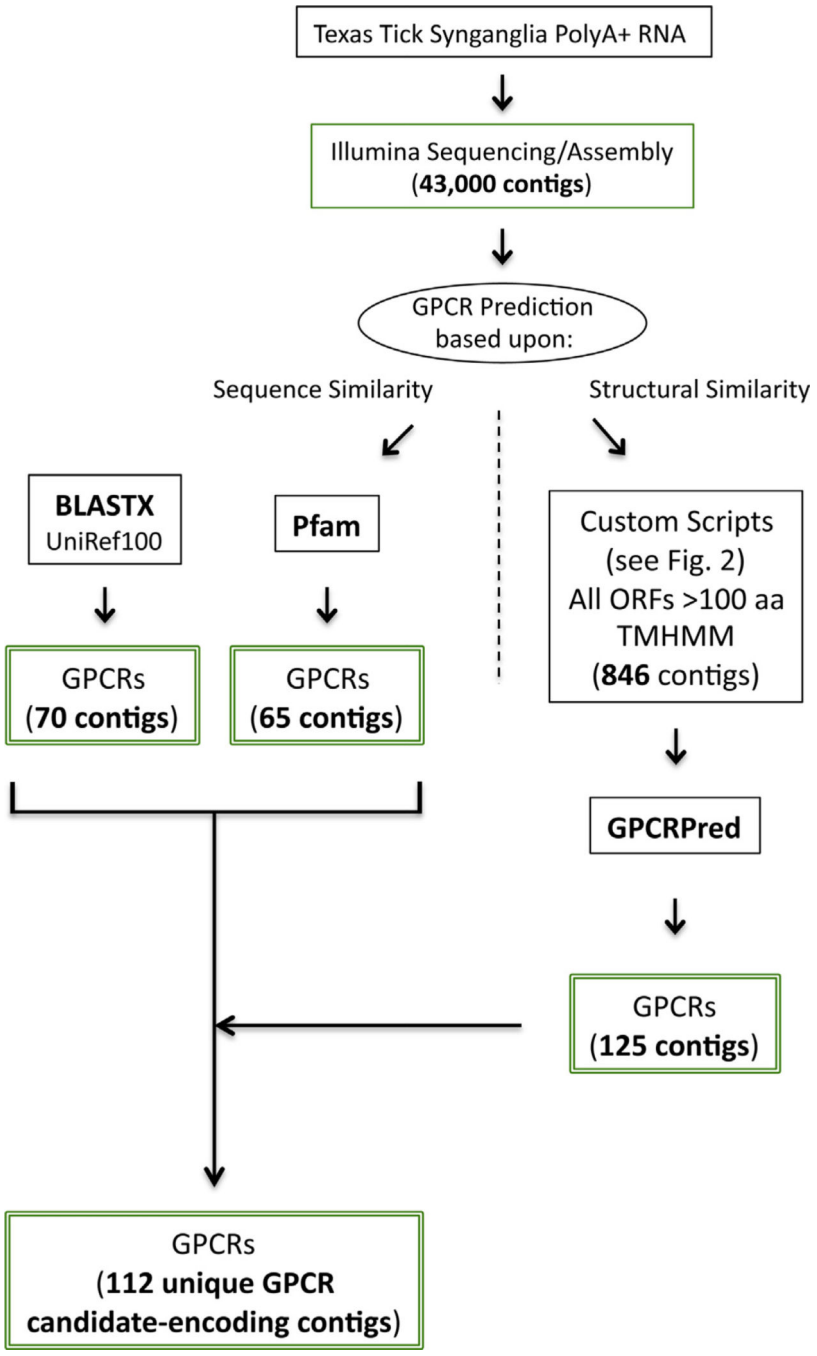

Fig. 1.

GPCR analysis flow diagram showing the three concurrent analytical approaches to GPCR prediction. Each of the 43,468 Texas cattle tick synganglion transcript contigs were independently evaluated by BLASTX, Pfam, and TMHMM/GPCRPred analysis. Each method predicted different numbers of GPCRs, with some overlap in predicted contigs. The final non-redundant set of candidate GPCR-encoding transcripts contained 112 sequences.

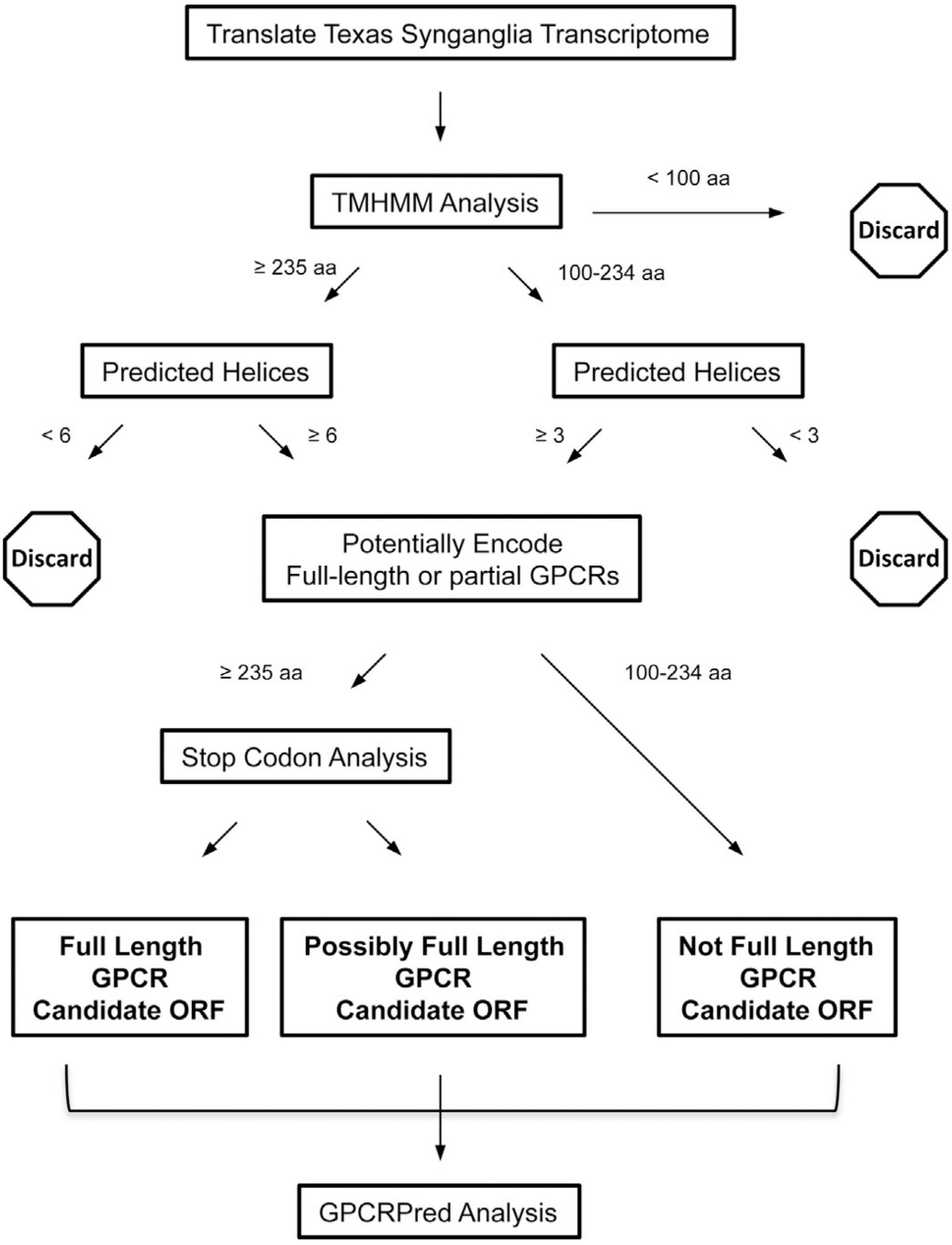

Fig. 2.

Flow chart of ORF prediction process whereby the translated Texas synganglion transcriptome ORFs are routed through TMHMM and GPCRPred to identify candidate GPCR-encoding sequences. The transcriptome sequences were translated into all 6 possible ORFs and those containing <100 amino acids were omitted from further analysis. All remaining ORFs were analyzed by TMHMM to predict and tally transmembrane helix regions. Following TMHMM, ORFs containing <100 aminoacids, ORFs containing 100–234 amino acids and <3 predicted helices, or ORFs containing ≥235 amino acids and <6

helices were omitted from further analysis. The corresponding nucleotide sequences from the remaining ORFs were examined for stop codons in the presumptive 5′ and 3′ untranslated regions to predict if the ORF encodes a full length protein. ORFs containing ≥2 stop codons prior to the putative initiator methionine codon and ≥2 stop codons at the 3′ end of the ORF were parsed as “Full Length GPCR Candidate ORF”. ORFs with only one stop codon both before and after the protein sequence were parsed as “Possibly Full Length GPCR Candidate ORF”. ORFs not meeting these criteria were parsed as “Not Full Length GPCR Candidate ORF”. ORFs in each set were analyzed by GPCRPred.

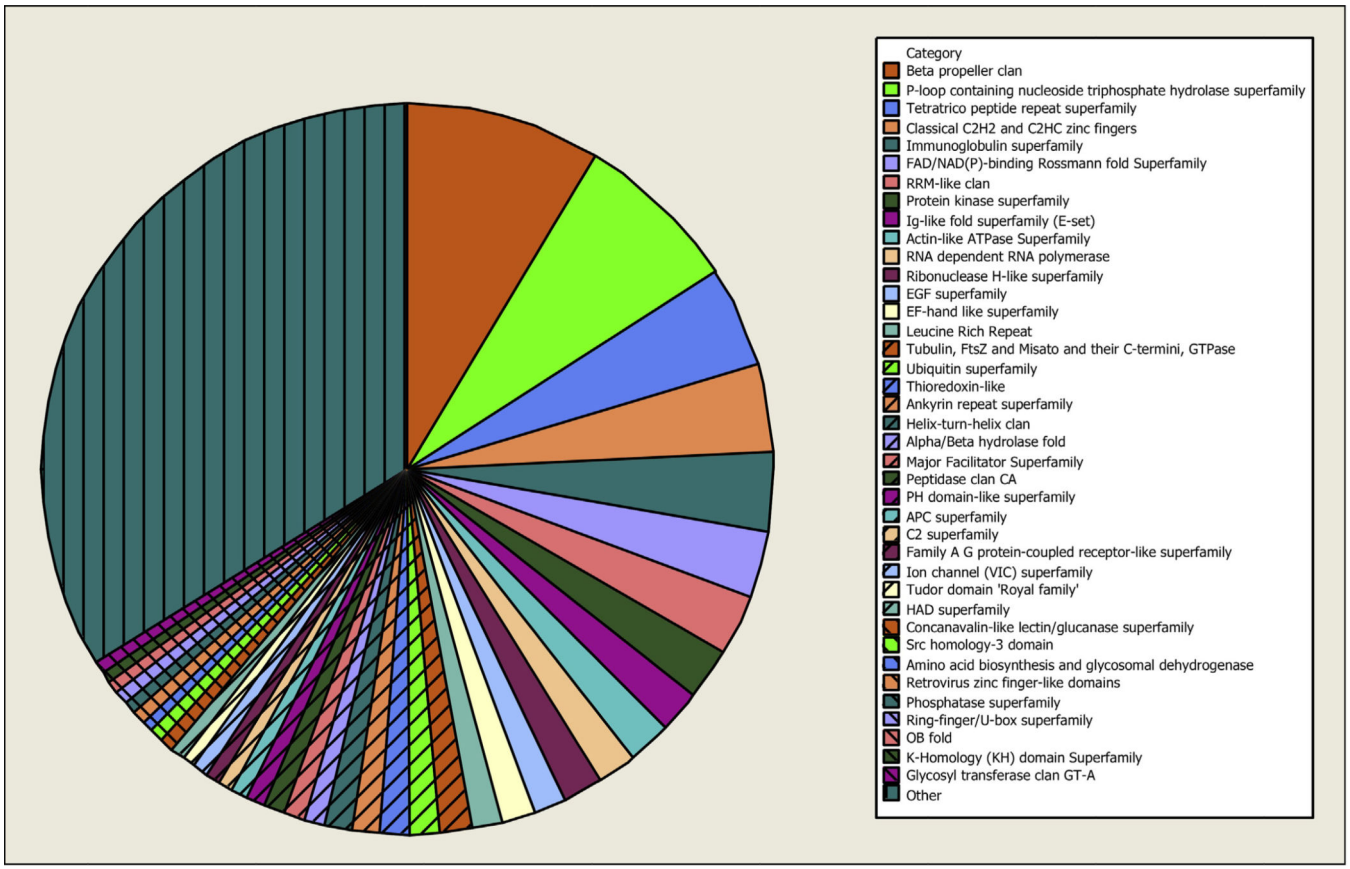

Fig. 3.

Chart showing the most common Pfam clans identified in the cattle tick synganglion transcriptome. Each contig of the Texas cattle tick synganglion transcriptome was translated in all 6 reading frames. ORFs of ≥50 amino acids were submitted for Pfam analysis and the resulting clans are shown in the pie chart. Pie slice sizes represent the number of times that a specific clan occurs in the population of all clan occurrences. Clans with occurrences <0.5% of the total were grouped into the “Other” category.

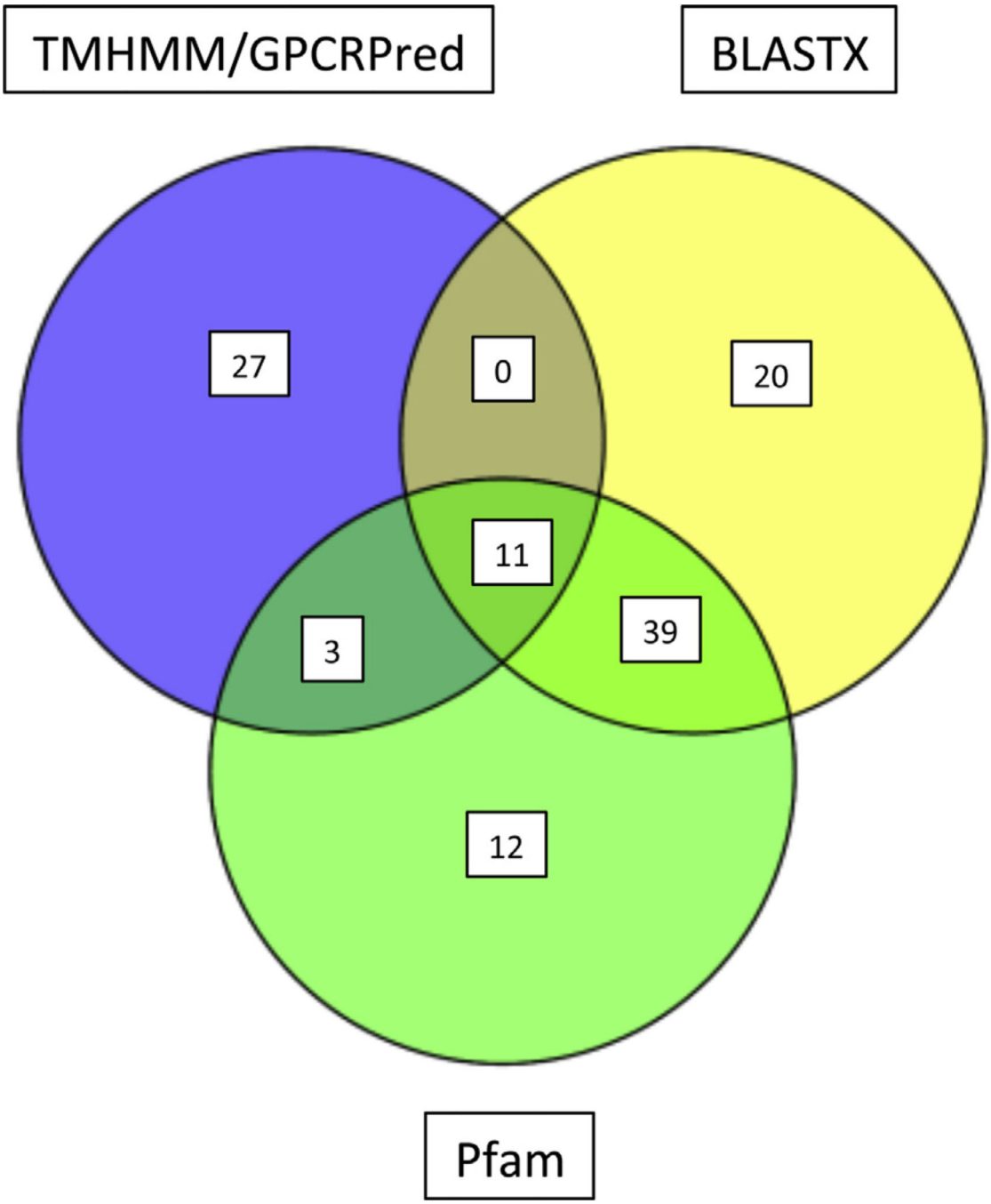

Fig. 4.
Venn diagram depicting the overlaps between the GPCR candidate predictions of the BLASTX-, Pfam-, and TMHMM/GPCRPred-based methodologies.

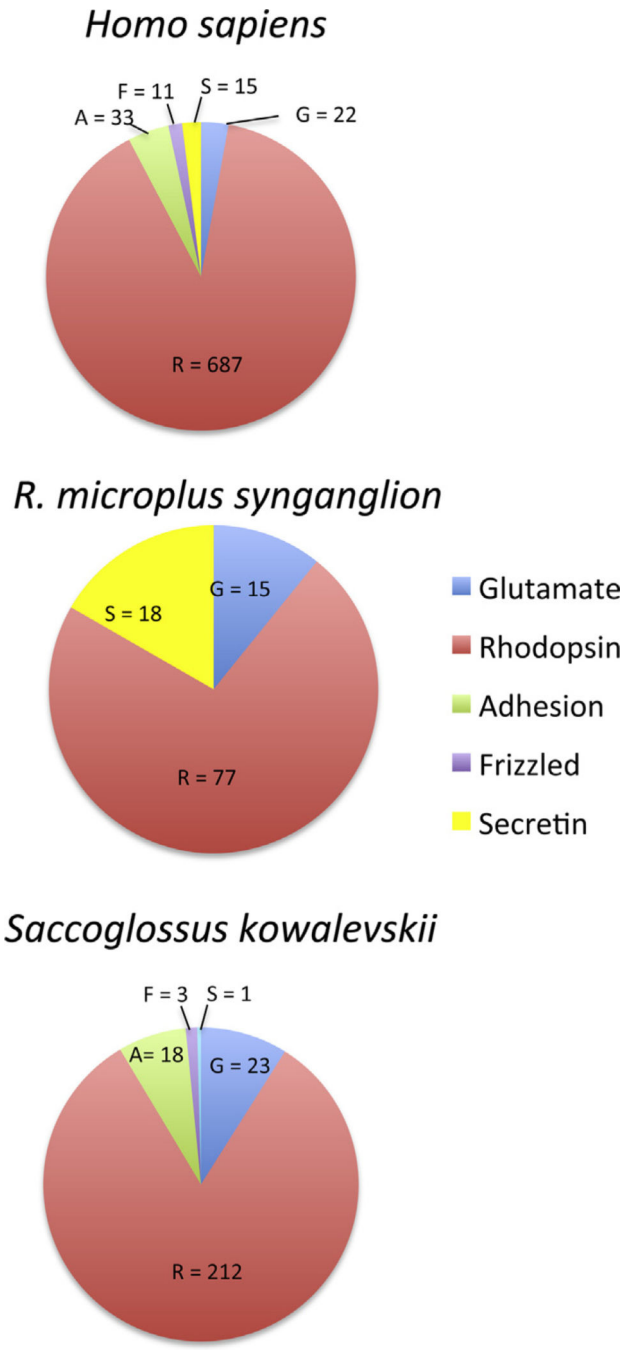

Fig. 5.

Distribution of GPCRs in the GRAFS classification system. The number of GPCRs in each of the Glutamate (blue), Rhodopsin (red), Adhesion (green), Frizzled (purple) and Secretin (yellow) families from *H. sapiens*, *S. kowalevskii*, and *R.microplus* synganglia are noted. The number of GPCRs from each family are shown in the relevant sections of the pie charts. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Cumulative listing of predicted GPCRs and proteins related to the GPCR signaling cascade from the Texas cattle tick synganglion transcriptome.

Predicted protein name	Candidate Unigene IDs[a]
GPCR Rhodopsin family	
5-Hydroxytryptamine receptor	MTc32715
5-Hydroxytryptamine type 7 receptor	MTc22509
Acetylcholine receptor, G-protein-linked	MTc5003
Acetylcholine receptor gar-2A, G-protein-linked	MTc22266
Acetylcholine receptor, Muscarinic	MTc2211, MTc40499
Allatostatin receptor	MTc6228
Allatostatin-A receptor	MTc39630
Capa receptor	MTc21993
Dopamine D1/beta receptor	MTc1524
Dopamine type 2 receptor	MTc15956, MTc24736
Leucine-rich repeat-containing G-protein coupled receptor	**MTc20805**
Leucine-rich repeat-containing G-protein coupled receptor 5	**MTc25248**
Myoinhibitory peptide receptor	**MTc27624**
Neuropeptide FF receptor	MTc27095
Neuropeptide receptor 15	MTc21990
Neuropeptide receptor A31	MTc35823
Neuropeptide Y receptor	MTc10061, MTc41765
Octopamine receptor, Alpha 2 adrenergic-like	MTc34982
Octopamine receptor	**MTc1903**, **MTc3394**, MTc5330, MTc24553, MTc28003, MTc41218
Pyrokinin receptor	MTc15995, MTc24861
Pyrokinin-like receptor	MTc104
Relaxin receptor 2	MTc37046
Serotonin receptor	MTc4134, MTc4625, MTc22214
Substance-K receptor	MTc8234, MTc29471
Sulfakinin receptor	MTc8208
Vesicular amine transporter	MTc422
Predicted Rhodopsin family but no candidate identity	MTc287, MTc925, MTc1475, MTc3264, **MTc3696**, MTc4791, MTc5115, **MTc5617**, MTc5741, MTc6796, MTc7698, MTc7846, MTc8067, MTc9908, MTc11315, MTc14173, MTc14673, MTc15391, MTc15557, MTc16269, MTc17406, MTc17511, MTc21160, MTc21447, MTc22106, MTc22868, **MTc24520**, MTc25152, **MTc27818**, **MTc32073**, **MTc32084**, MTc32224, MTc35522, MTc35757, MTc39286, MTc40170,

Predicted protein name	Candidate Unigene IDs[a]
	MTc41195, MTc42769, MTc43929
GPCR Secretin family	
Calcitonin receptor	MTc5685, MTc19512, MTc39975
Corticotropin-releasing factor receptor type	MTc20694, MTc25752
DH31 receptor	MTc20395
Latrophilin	**MTc14709**, MTc25802, MTc44419
Methuselah-like 3	**MTc1232**
Parathyroid hormone/parathyroid hormone-related peptide receptor	MTc17148
Predicted Secretin Family but no Candidate Identity	**MTc19630**, **MTc20723**, **MTc32602**, **MTc39823**, **MTc39868**, **MTc40313**, **MTc42573**
GPCR glutamate family	
Metabotropic GABA-B receptor subtype	MTc42297
Metabotropic gamma-aminobutyric acid receptor	MTc840, MTc2654, MTc39275, MTc43975
Metabotropic glutamate receptor	MTc17814
Metabotropic glutamate receptor 1	MTc1831, MTc37251
metabotropic glutamate receptor 2	MTc7139
Metabotropic glutamate receptor 4, 6, 7	MTc44180
Pheromone and odorant receptor	MTc20045, MTc21997, MTc22950, MTc23544
Predicted Glutamate Family but no Candidate Identity	MTc12
GPCR predicted but unclassifiable to family	
MTc37686	
MTc27554	
G proteins	
Guanine nucleotide binding protein beta subunit	MTc8797, MTc3426, MTc10031, MTc10239, MTc10469, MTc10616, MTc10744, MTc11060, MTc12068, MTc13386, MTc14081, MTc14129, MTc14435, MTc14489, MTc15183, MTc18040, MTc24941, MTc26235, MTc26826, MTc29590, MTc31138, MTc31358, MTc34616, MTc37393, MTc37592, MTc39119, MTc40432, MTc41893, MTc42759, MTc44290
G protein beta subunit-like protein	MTc31519, MTc31755, MTc39402
GTP-binding protein (Q) alpha-11 subunit, gna11	MTc6574
GTP-binding protein (I) alpha subunit, gnai	MTc43652
Guanine nucleotide-binding protein G(O) subunit alpha	MTc896
Guanine nucleotide-binding protein subunit gamma	MTc9439, MTc25067
G protein-coupled receptor kinase	

Predicted protein name	**Candidate Unigene IDs**[a]
G-protein coupled receptor kinase 2/3	MTc18044
Adenylyl cyclase	
Adenylyl cyclase	MTc1356, MTc17511, MTc19476, MTc21334, MTc27478, MTc27533, MTc31965, MTc33919, MTc34460, MTc36772, MTc36830, MTc38293, MTc39594, MTc41584, MTc41766
Adenylyl cyclase type	MTc6353, MTc17129
Adenylyl cyclase type 2	MTc3632
Adenylyl cyclase-associated protein	MTc2672, MTc6656, MTc7841
Inositol triphosphate receptor	
Type 3 inositol 1,4,5-trisphosphate receptor	MTc21140, MTc37353
Arrestin	
Arrestin domain-containing protein	MTc297, MTc18392
Beta-arrestin 1	MTc6527

[a]See Supplemental files for detailed information about each contig. The 22 Texas tick contigs in bold text were noted in the initial BLASTX search of the Uniref100 database (Supplemental file 4) with top hit description term of “G-protein coupled receptor”.