Ticks and Tick-borne Diseases 7 (2016) 670–677

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

journal homepage: www.elsevier.com/locate/ttbdis

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], 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*

ARTICLE INFO

Article history:
Received 23 September 2015
Received in revised form 17 February 2016
Accepted 18 February 2016
Available online 22 February 2016

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

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

Published by Elsevier GmbH.

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).

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

http://dx.doi.org/10.1016/j.ttbdis.2016.02.014
1877-959X/Published by Elsevier GmbH.

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, Šimo et 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 targets 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