Ticks and Tick-borne Diseases xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

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

A foreleg transcriptome for *Ixodes scapularis* ticks: Candidates for chemoreceptors and binding proteins that might be expressed in the sensory Haller's organ

Tanya Josek*, Kimberly K.O. Walden, Brian F. Allan, Marianne Alleyne, Hugh M. Robertson

Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL, 61801, United States

ARTICLE INFO

Keywords:
Ixodes scapularis
Ionotropic receptors
Gustatory receptors
Haller's organ

ABSTRACT

Little is known about the molecular basis for the olfactory capabilities of the sensory Haller's organ on the forelegs of ticks. We first expanded the known repertoire of Ionotropic Receptors (IRs), a variant lineage of the ionotropic glutamate receptors, encoded by the black-legged *Ixodes scapularis* genome from 15 to 125. We then undertook a transcriptome study of fore- and hind-legs of this tick in an effort to identify candidate chemoreceptors differentially expressed in forelegs as likely to be involved in Haller's organ functions. We primarily identified members of the IR family, specifically Ir25a and Ir93a, as highly and differentially expressed in forelegs. Several other IRs, as well as a few members of the gustatory receptor family, were expressed at low levels in forelegs and might contribute to the sensory function of Haller's organ. In addition, we identified eight small families of secreted proteins, with sets of conserved cysteines, which might function as binding proteins. The genes encoding these Microplusin-Like proteins and two previously described Odorant Binding Protein-Like proteins share a common exon-intron structure, suggesting that they all evolved from a common ancestor and represent an independent origin of binding proteins with potential roles comparable to the ChemoSensory Proteins and Odorant Binding Proteins of insects. We also found two Niemann-Pick Type C2 proteins with foreleg-biased expression, however we were unable to detect foreleg-biased expression of a G-Protein-Coupled pathway previously proposed to mediate olfaction in the tick Haller's organ.

1. Introduction

Hematophagous arthropods exploit a wide array of physiological mechanisms to find hosts. Insects, such as mosquitoes, find their hosts by utilizing their antennae, maxillary palps, and proboscis (Maekawa et al., 2011). However, ticks have neither a proboscis nor antennae to aid in host finding, instead they have a specialized sensory organ found on the tarsi of the forelegs called the Haller's organ (Haller, 1881). The Haller's organ is unique to the Order Ixodida and varies greatly in morphology between tick families (Keirans et al., 1976; Klompen and Oliver, 1993), however its overall role in sensory biology is shared among species within the Order (Hindle and Merriman, 1912; Sonenshine and Roe, 2013). Use of the organ for sensory purposes can be seen when ticks lift up and wave their forelegs, similarly to how insects move their antennae (Nuttall et al., 1908). The Haller's organ itself consists of a pit containing a central cluster of sensory hairs and a group of sensilla within a capsule aperture (Josek et al., 2017). There are additional sensory hairs surrounding the structure as well. Although not all functions of the Haller's organ are well understood, the organ's sensilla are involved in mechanosensation, olfaction, and humidity detection (Foelix and Axtell, 1972; Soares and Borges, 2012). The Haller's organ is primarily responsible for recognizing chemicals such as carbon dioxide, but is capable of recognizing other host odorants (Sonenshine, 2004). Additionally, studies have concluded that this organ plays a role in sensing tick pheromones (Carr et al., 2017; Sonenshine, 2004; Rechav et al., 1977). If the Haller's organ is removed or lost before a blood meal, the tick loses its ability to find a specific host, but can still locate a proper spot to feed if placed on a host (Hindle and Merriman, 1912), indicating that the Haller's organ is not involved in final feeding site location.

Chemoperception in arthropods depends on several large and a few small gene families (Benton, 2015; Joseph and Carlson, 2015; Ray et al., 2014; Suh et al., 2014). While the odorant receptor (OR) family of seven-transmembrane-domain proteins is the best understood of these in insects and mediates much of their diverse olfactory abilities, it is a relatively young expansion of this kind of chemoreceptor, probably originating within the Insecta (Ioannidis et al., 2017; Missbach et al., 2014; Robertson et al., 2003). It is not present in the non-insect

* Corresponding author.
E-mail address: josek1@illinois.edu (T. Josek).

https://doi.org/10.1016/j.ttbdis.2018.05.013
Received 29 December 2017; Received in revised form 24 May 2018; Accepted 25 May 2018

arthropods for which genome sequences are available like the crustacean *Daphnia pulex* (Peñalva-Arana et al., 2009), the centipede *Strigamia maritima* (Chipman et al., 2014), or the mite *Metaseiulus occidentalis* (Hoy et al., 2016), and the recently published genome of *Ixodes scapularis* similarly contains no OR genes (Gulia-Nuss et al., 2016). Instead, non-insect arthropods appear to rely on two chemoreceptor gene families also present in insects. The gustatory receptor (GR) family of seven-transmembrane proteins has highly diverse members in insects that mediate perception of diverse stimuli, mostly gustatory like sugars and bitter tastants, but also some odorants like carbon dioxide (Benton, 2015). It is an ancient gene family present in basal animals (Eyun et al., 2017; Robertson, 2015; Saina et al., 2015;), and perhaps beyond (Benton, 2015). Like the ORs they range in size from around 350–450 amino acids. The OR and GR families are clearly related to each other and together are known as the insect chemoreceptor superfamily (Robertson et al., 2003), with the OR family being an insect-specific expansion of a single lineage (Missbach et al., 2014; Ioannidis et al., 2017). A structure is not yet known for any member of this superfamily, but a likely model for the ORs that surely also applies to the GRs has been inferred (Hopf et al., 2015). It consists of the expected seven transmembrane domains arranged in the opposite polarity of the G-Protein-Coupled Receptor (GPCR) superfamily to which most nematode and vertebrate chemoreceptors belong, that is, with the N-terminus intracellular and the C-terminus extracellular. This structure and other features make it unlikely that these ORs and GRs function via a G-protein intermediate, but rather as ligand-gated ion channels (Joseph and Carlson, 2015). While the ligand-binding site is not known, it is likely to involve a combination of parts of some of the transmembrane domains as well as the three extra-cellular loops that connect them (Hopf et al., 2015).

The second large family of chemoreceptors in insects and other arthropods is known as the Ionotropic Receptor (IR) family and is a variant lineage of the ionotropic glutamate receptors widely present in animals and beyond (Benton et al., 2009; Rytz et al., 2013; Rimal and Lee, 2018). These proteins consist of an extracellular ligand-binding domain supported by three transmembrane domains and are a completely independent evolutionary origin of chemoreceptors in arthropods and other protostomes (Croset et al., 2010; Eyun et al., 2017). While some are known to be involved in olfaction (Rytz et al., 2013), most are likely to have gustatory roles (Koh et al., 2014; Stewart et al., 2015), while some even mediate perception of temperature and humidity (Enjin et al., 2016; Knecht et al., 2017, 2016; Ni et al., 2016). There are two highly conserved co-receptors in the IR family known for their names in *Drosophila melanogaster* of Ir8a and Ir25a that form dimers or higher multimers with the other IR proteins. These two proteins, as well as the Ir93a protein involved in perception of humidity in *D. melanogaster*, are comparable in length to the ionotropic glutamate receptors at around 940 amino acids. Most of the other IRs in insects and other arthropods are much shorter at their N-terminus and in ticks consist of around 430 amino acids.

The *I. scapularis* genome paper reported 62 GR genes but only 15 IR genes (Gulia-Nuss et al., 2016). Using transcriptomes from the legs of adult ticks, we first identified many additional IR genes in this genome and refined a few GR models. We then compared expression levels of GRs and IRs in fore- and hind-legs seeking chemoreceptors differentially expressed in the forelegs that might be involved in chemosensation in the Haller's organ. We also examined our transcriptome for additional proteins expressed highly and differentially in forelegs, and describe several families of candidate binding proteins that might be involved in olfaction, while also examining the expression of members of the Niemann-Pick type C2 family proposed to be binding proteins important to tick olfaction (Iovinella et al., 2016; Pelosi et al., 2014), and a GPCR pathway recently proposed as an alternative molecular mechanism for olfaction in ticks (Carr et al., 2017).

2. Methods

2.1. Annotation of chemoreceptors

IR genes were identified and modeled using methods similar to those in Terrapon et al. (2014) and Hoy et al. (2016). Briefly, iterative TBLASTN searches of the genome assembly with known tick, mite, and other arthropod IRs identified regions of scaffolds. Gene models were built manually in TEXTWRANGLER and relevant repairs were made to the genome assembly using raw genome reads from the Trace Archive at NCBI. The transcriptome assembly described below was employed to assist in building models when possible, as were raw RNAseq reads from multiple other experiments deposited in the Sequence Read Archive at NCBI. Pseudogenes were translated as best possible to provide a protein that could be aligned (using Z for stop codons and X for frameshifts and other obvious pseudogenizing mutations). Protein sequences were aligned in CLUSTALX v2.0 (Larkin et al., 2007) using default settings and gene models refined in light of these alignments. For phylogenetic analysis, the final alignment was trimmed using TRIMAL v1.4 (Capella-Gutiérrez et al., 2009), using the "gappyout" option for the GRs that are all of approximately the same length, and the "strict" option for the IRs, which effectively removes the variable length N-termini with highly divergent sequences from this family. Maximum likelihood phylogenetic analysis was performed using PHYML v3.0 (Guindon et al., 2010) using default settings, and tree figures were prepared in FIGTREE v1.4 (http://tree.bio.ed.ac.uk/software/figtree/). A similar updating effort and phylogenetic analysis was performed for the GR family, including the mite proteins from Hoy et al. (2016), along with representatives of the three most conserved GR lineages in insects, the carbon dioxide, sugar, and fructose receptors. Models for genes encoding candidate binding proteins and GPCR pathway proteins were similarly manually built.

2.2. Tick collection, processing, and sequencing

Ixodes scapularis ticks were collected from three locations: Allerton Park (39°59′39.7"N 88°38′46.5"W, County, Illinois), Danada House (41°49′16.5"N 88°06′21.7"W, DuPage County, Illinois) and Waterfall Glen County Forest Preserve, (41°42′19.8"N 88°00′12.6"W, DuPage County, Illinois) using dragging and flagging methods. After collection, ticks were brought back to the lab and identified to species (Keirans and Litwak, 1989). Forelegs and hindlegs of *I. scapularis* ticks were removed and placed in separate RNase-free tubes embedded in pelleted dry-ice and stored at −80° C. A total of 99 males (198 forelegs/198 hindlegs) and 100 females (200 forelegs/200 hindlegs) were used for RNA isolation. Legs were manually ground in 1 ml Trizol (Invitrogen) in glass tissue grinders, then filtered over a Qiashredder column (Qiagen). The homogenate was extracted with chloroform, and the RNA was precipitated with LPA (10 mg/mL) and isopropanol. RNA pellets were washed with 75% ethanol and resuspended in RNase-free water. RNA was quantified with a Qubit RNA Broad Range Assay Kit on a Qubit fluorometer (Life Technologies). Small samples of total RNA from the four leg samples were visualized using ethidium bromide on a 1.0% agarose gel to evaluate quality. The RNAseq libraries were prepared from an average cDNA fragment size of 250 bp using the TruSeq Stranded RNAseq Sample Prep kit from Illumina. The four libraries were individually barcoded and quantitated using qPCR before pooling and sequencing from both ends with the TruSeq SBS Sequencing Kit v3 for 100 cycles on a single lane of a HiSeq2500 instrument. Our data were processed with Casava 1.8.2 (Illumina) before conversion into FASTQ files. Using the FASTX-Toolkit software (http://hannonlab.cshl.edu/fastx_toolkit/) the reads were trimmed at the 5′ end and the 3′ end to remove low quality bases (-t 20). The resulting trimmed reads from all four leg samples were assembled together with SOAPdenovo-Trans-127mer v1.02 and Trinity (Release 2014-04-13) (Haas et al., 2013; Luo et al., 2012). The raw RNAseq reads have been submitted to the