Molecular and Cellular Probes 31 (2017) 85–90

Contents lists available at ScienceDirect

Molecular and Cellular Probes

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

Original research article

Transcriptome of larvae representing the *Rhipicephalus sanguineus* complex

Leone De Marco [a, b], Sara Epis [c], Francesco Comandatore [a, c], Daniele Porretta [d], Claudia Cafarchia [e], Valentina Mastrantonio [d], Filipe Dantas-Torres [e, f], Domenico Otranto [e], Sandra Urbanelli [d], Claudio Bandi [c], Davide Sassera [a, *]

[a] Department of Biology and Biotechnology, University of Pavia, Pavia, Italy
[b] School of Bioscience and Veterinary Medicine, University of Camerino, Camerino, Italy
[c] Department of Veterinary Science and Public Health, University of Milan, Milan, Italy
[d] Department of Environmental Biology, University of Rome 'La Sapienza', Rome, Italy
[e] Department of Veterinary Medicine, University 'Aldo Moro' of Bari, Bari, Italy
[f] Department of Immunology, Aggeu Magalhães Research Center, Oswaldo Cruz Foundation, Recife, Brazil

ARTICLE INFO

Article history:
Received 23 December 2015
Received in revised form
4 February 2016
Accepted 7 February 2016
Available online 24 February 2016

Keywords:
Rhipicephalus sanguineus
Larvae
Transcriptome
ABC transporters
Microsatellites

ABSTRACT

Rhipicephalus sanguineus sensu lato (Ixodida: Ixodidae) is possibly the most widespread tick species worldwide, responsible for transmitting several vector-borne pathogens of medical and veterinary importance. Here, we explore the transcriptome of *R. sanguineus* s.l. larvae (Putignano strain). We sequenced total RNA from *R. sanguineus* s.l. larvae. A total of 15,566,986 short paired-end reads were *de novo*-assembled into 33,396 transcripts and then annotated and analyzed. Particular attention was paid to transcripts putatively encoding ATP-binding proteins, due to their importance as mechanisms of detoxification and acaricide resistance. Additionally, microsatellite loci were investigated, as these are useful markers for population genetic studies. The present data and analyses provide a comprehensive transcriptomic resource for *R. sanguineus*. The results presented here will aid further genetic and genomic studies of this important tick species.

1. Introduction

Next-generation sequencing technologies and advances in assembly algorithms have greatly assisted genetic and genomic studies of arthropod vectors. In this context, RNA-sequencing (RNA-Seq) approaches allow whole-transcriptome analyses, even without the aid of a reference genome [1,2]. RNA-Seq has been applied to several arthropod vector species and has provided genetic resources for functional studies of pathogen-vector interactions and insecticide/acaricide resistance [3,4]. Despite ticks being amongst the most important arthropod vectors for infectious diseases of humans and animals [5], transcriptomic studies are only available for a small number of species [4,6,7].

The brown dog tick, *Rhipicephalus sanguineus* sensu lato (s.l), is possibly the most widespread hard tick species worldwide. Due to its capability of transmitting bacterial, viral, and protozoan pathogens, this species is of major public health relevance and economic impact [8]. *R. sanguineus* s.l. can act as a vector of, for example, *Babesia vogeli*, *Ehrlichia canis* (the agents of canine babesiosis and ehrlichiosis, respectively), and *Rickettsia* spp. (causing spotted fever in humans) [8]. The taxonomic status of *R. sanguineus* sensu stricto is the subject of ongoing debate, and morphological, genetic and crossbreeding studies have showed that ticks identified as '*R. sanguineus*' represent a complex of species (reviewed in Refs. [9,10]); henceforth, we will refer to these ticks as *R. sanguineus* sensu lato (s.l).

The only previous work focused on the sialotranscriptome of *R. sanguineus* used expressed sequence tags (ESTs), a low throughput technique, which provided a partial view of the transcriptome [7,11]. Here, we aimed to construct a reference transcriptome for larvae representing *R. sanguineus*, to provide data that will aid to future genetic, genomic and functional studies. We performed a whole-transcriptome analysis, starting from total RNA from larvae of *R. sanguineus* s.l.. We characterized all larval transcripts, and studied simple sequence repeats (SSRs) and sequences

* Corresponding author. Department of Biology and Biotechnology, University of Pavia, Via Ferrata 9, 27100, Pavia, Italy.
E-mail address: davide.sassera@unipv.it (D. Sassera).

http://dx.doi.org/10.1016/j.mcp.2016.02.006
0890-8508/

of ATP-binding cassette (ABC) transporters.

2. Materials and methods

2.1. Ethics statement

Adult *R. sanguineus* were collected from clinically healthy dogs in Putignano (province of Bari, Italy; 40 510N, 1770E) from private owners according to the principles of good clinical practice (VICH GL9 GCP, 2000 http://www.emea.eu.int/pdfs/vet/vich/059598en.pdf), upon agreement with the owners of the dogs. All procedures for tick breeding were carried out in strict accordance with principles of the 105 3Rs European directive (2010/63/EU), National Animal Testing Rules (D.Lgs 116/92) and all efforts were made to minimize animal suffering. All procedures were approved by the University of Bari (protocol no. 9/12).

2.2. Tick rearing

Adult *R. sanguineus* s.l. were collected from clinically healthy dogs in Putignano (province of Bari, Italy; 40 510N, 1770E) and identified based on morphological and genetics as the operational taxonomic unit *Rhipicephalus* sp. I, which belongs to the *R. sanguineus* s.l. complex [12]. The specimens were used to establish a laboratory colony at the Department of Veterinary Medicine, University of Bari (Italy) in the absence of acaricide exposure for multiple generations under controlled conditions of temperature (27 ± 1 °C), relative humidity (RH 80 ± 5%) and photoperiod (12 h light: 12 h dark). For this study, larvae were collected 7 days after hatching; a total of 300 larvae were pooled and stored in 15 ml of RNA-Later stabilization solution (Qiagen, Hilden, Germany) at −80 °C.

2.3. RNA extraction and RNA-Seq

RNA from the pool of 300 *R. sanguineus* larvae was sequenced by a company (GATC Biotech AG, Constance, Germany) in one run of 2 × 250 paired-ends reads on a MiSeq platform (Illumina). Reads are available in the EBI Short Read Archive under the accession number ERS687377 (http://www.ebi.ac.uk/ena/data/view/ERS687377). The assembled transcriptome is available via EBI TSA under accession numbers HACW01000001-HACW01033611 (http://www.ebi.ac.uk/ena/data/view/PRJEB8914&portal=sequence_update). The whole dataset is available as a Study entry at http://www.ebi.ac.uk/ena/data/view/PRJEB8914.

2.4. Reads assembly and filtering

After an assessment of the quality of the reads using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), the paired end reads were assembled using the Trinity de novo-assembler for short reads [13], setting all parameters at default values, filtering out sequences of less than 200 nucleotides. In order to obtain a final curated transcriptome, we curated and filtered this draft transcriptome as follows. We used RSEM [14] to estimate levels of transcription (fragments per kilobase of transcript per million mapped reads, FPKM), a relative measure of transcript abundance. RSEM specifies estimates for both isoforms and genes; estimates for the genes are obtained by collapsing all isoform estimates into the corresponding genes. The filtering was performed at the gene level. Conservatively, all putative transcripts corresponding to each gene were retained only if FPKM was >1. Subsequently, a second filtering step was performed using TransDecoder, to identify probable candidate coding regions within transcript sequences. Finally, redundant duplicate transcripts were discarded with an in-house Python script.

2.5. Annotation

We designated genes using a BLASTP search of the transcriptome against the RefSeq protein database [15], setting the e-value at 10^{-5}. The BlastP results were used as the input for Blast2GO [16], to associate Gene Onthology terms [17], EC numbers and KEGG pathways. GO annotations were reduced using a combination of the Blast2GO online GUI version and in-house Python scripts, in order to obtain level 4 GO terms belonging to the three main categories for easier visualisation. A second BLASTP search was executed on the Cluster of Orthologous Groups of proteins (COG) database [18], using an e-value of $<10^{-5}$. Transcripts assigned to multiple COG pathways were considered as belonging to each of the assigned pathways. All transcripts were additionally annotated using InterProScan5 [19], a program that uses a multitude of databases, such as PANTHER, Pfam, ProSite, PRINTS, SMART and Superfamily. The SignalP software [20] was used locally, with default settings, in order to predict whether transcripts could be encode proteins that are excretory/secretory. In order to find simple sequence repeats (SSRs), the whole tick transcriptome was screened using the software MISA [21] using default parameters.

2.6. ABC family phylogeny

We retrieved from the Swissprot database [22] sequences encoding ABC transporters belonging to the Metazoa, to be used as a BLAST database to detect potential ABC genes in the transcriptome of *R. sanguineus* s.l.. After a filtering step (e-value $< 10^{-10}$), the one isoform scoring the lowest e-value for each gene was retained. The resultant sequences, along with the ABC Swissprot database previously introduced, were multi-aligned using Clustal omega [23]. The multi-alignment was used for a maximum likelihood phylogenetic analysis using RAxML [24] using 100 bootstraps.

3. Results

3.1. Sequencing and assembly

In order to obtain a snapshot of the larval transcriptome of *R. sanguineus* s.l., whole-transcriptome sequencing of RNA was performed; 15,566,986 high-quality paired-end short reads were obtained. De novo transcriptome assembly was performed, resulting in a total of 208,926 putative transcripts (N50 = 1228), clustered into 142,934 putative genes. We filtered this dataset, retaining only genes with a baseline level of transcription (FPKM > 1), obtaining 99,224 transcripts inferred to belong to 43,594 unique genes. An additional filtering step was then performed running TransDecoder, yielding 43,204 transcripts belonging to 16,555 unique genes. A final count of 33,396 transcripts (clustered into 16,432 unique genes) was obtained after the final filtering step consisting of the removal of redundant identical sequences. Based on the workflow described here, we consider this a reliable and conservative dataset. Refer to Table 1 for a complete summary of the assembly characteristics.

3.2. Annotation

The BLASTP search of the transcriptome against RefSeq was performed, returning significant matches for 27,565 sequences of a total of 33,396; 73% of transcripts matched to sequences belonging to organisms of the Parasitiformes, in particular to *Ixodes scapularis* (64%) and *Metaseiulus occidentalis* (Mesostigmata: Phytoseiidae) (9%). Transcript length distributions were significantly different