International Journal for Parasitology 49 (2019) 911–919

Contents lists available at ScienceDirect

International Journal for Parasitology

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

De novo genome sequencing and comparative stage-specific transcriptomic analysis of *Dirofilaria repens*

Chiara Cafarelli [a,1], Giancarlo Russo [b,1], Alexander Mathis [a], Cornelia Silaghi [a,c,*]

[a] National Centre for Vector Entomology, Institute of Parasitology, Vetsuisse Faculty, University of Zürich, Winterthurerstr 266a, 8057 Zürich, Switzerland
[b] Functional Genomics Center Zürich, University of Zürich and ETH Zürich, Winterthurerstr. 190, 8057 Zürich, Switzerland
[c] Friedrich-Loeffler-Institute, Federal Research Institute for Animal Health, Südufer 10, 17493 Greifswald, Insel Riems, Germany

ARTICLE INFO

Article history:
Received 29 January 2019
Received in revised form 16 April 2019
Accepted 24 April 2019
Available online 23 September 2019

Keywords:
Dirofilaria repens
Genome
Transcriptome
Microfilariae
L3

ABSTRACT

The zoonotic mosquito-borne filarial nematode *Dirofilaria repens* causes subcutaneous and ocular infections in dogs, cats and humans. From infected vertebrate hosts, microfilariae are taken up by mosquitoes and develop into infective L3. These are transmitted to new vertebrate hosts and develop over two further moults to adult worms. The aims of the project were (i) the *de novo* sequencing and annotation of the *D. repens* genome and (ii) comparative transcriptomic analyses of the two developmental stages, mf and L3. Genomic DNA was obtained from adult male *D. repens*. RNA was extracted from mf from naturally infected dogs and from L3 produced in *Aedes aegypti* mosquitoes fed on blood spiked with mf. The 99.59 MB genome was approximately 17% larger than that of the related species *Dirofilaria immitis* (dog heartworm) and contained 8.9% fewer predicted genes (10,357). Approximately 1.8% of identified proteins (206/11,262) could not be mapped to *D. immitis*. Out of these, six (2.9%) presented an ortholog in all other considered filarial nematodes (e.g. *Loa loa*) and *Caenorhabditis elegans*. A significantly higher number of *D. repens* proteins, compared with *D. immitis*, mapped to the filarial nematode *L. loa*, reflecting the similarity in biology of *D. repens* and *L. loa*. A total of 876 genes were differentially expressed, of which 591 could be annotated in UniProtKB/Swiss-Prot. In particular, 155 genes with a UniProtKB/Swiss-Prot annotation to *C. elegans* and filarial nematodes were upregulated in the L3 and 57 in the mf stage, respectively. Fifteen Gene Ontology Biological Processes were significantly enriched for the L3 group and 12 for the mf. To our knowledge these data provide the first insight into the differential gene expression profiles of this filarial nematode and can serve future investigations of metabolic processes and stage-specific diagnostics.

1. Introduction

Dirofilaria repens is a mosquito-borne filarial nematode of the family Onchocercidae which causes subcutaneous and ocular infections in dogs and cats. The life cycle of *D. repens* contains five developmental stages and requires both an arthropod vector (mosquito species mostly of the genera *Culex, Aedes* and *Anopheles*) which are also the intermediate host (Simon et al., 2012; Silaghi et al., 2017) and a mammalian definitive host. After mating, adult female *D. repens* worms release microfilariae (mf) into the blood stream of the mammalian host. Circulating mf (L1) are taken up by a bloodsucking mosquito in which the mf invade, after approximately 24 h, the cells of the Malpighian tubules, develop into the so-called sausage form, then undergo two moults to L2 and finally to the infective L3. For successful transmission, the L3 have to migrate from the Malpighian tubules to the proboscis of the mosquito. During a subsequent blood meal, infective L3 actively leave the labiae and penetrate the host's skin (Simon et al., 2012) where they moult to L4. These stages move to the subcutaneous tissue and muscular connective fasciae where they develop after a last moult to the adult stage and reside permanently. Adult worms live up to 4 years in their natural hosts (Genchi and Kramer, 2017). Dogs are the main reservoir, while microfilariaemia is absent or low in wild carnivores or cats (Magi et al., 2008). Furthermore, humans can be accidental hosts for *D. repens*. Subcutaneous, or the often clinically more severe ocular, dirofilariosis are considered emerging zoonosis in Europe and are already well-known in eastern European countries such as Russia (the southern part) and Ukraine (Capelli et al., 2018). *Dirofilaria repens* occurs in various regions of Europe, Africa and Asia. In recent years, *D. repens* has

* Corresponding author at: Friedrich-Loeffler-Institute, Federal Research Institute for Animal Health, Südufer 10, 17493 Greifswald, Insel Riems, Germany.
E-mail address: Cornelia.Silaghi@fli.de (C. Silaghi).
[1] Both authors contributed equally.

https://doi.org/10.1016/j.ijpara.2019.04.008
0020-7519/

been expanding northwards, assumedly partly due to an increasing number of dogs travelling within Europe, and climate change (Genchi et al., 2011). As is the case in many filarial nematodes, *D. repens* harbors intracellular symbiotic bacteria of the genus *Wolbachia* which are thought to be beneficial to the nematode host (Taylor et al., 2005; Slatko et al., 2010; Godel et al., 2012).

In contrast to the related filarial nematode *Dirofilaria immitis* (heartworm), for which the genome and a few transcriptome studies have recently been published (Chaisson and Tesler, 2012; Godel et al., 2012; Luck et al., 2014), data on neither the nuclear genome nor on the transcriptome of *D. repens* is yet available. Therefore, the aim of the present study was threefold: first, we aimed to fill the gap in the genomics landscape of filarial nematodes by generating the first published comprehensive resource for the organism *D. repens;* second, we tried to investigate whether differences in the genomes and the transcriptomes of the filarial nematodes, and in the particular between the two *Dirofilaria* spp., could point to specific traits of these organisms; finally, we wanted to identify which genes contribute the most to the transcriptional activities in the different stages analysed, i.e., mf in the mammalian host and L3 in the mosquitoes. Differences in the overall transcriptional profiles between these groups are to be expected, however, the main focus was to identify stage-specific genes which would help to gain first insights into metabolic pathways of this filarial nematode.

2. Material and methods

2.1. Parasite material

The adult male *D. repens* worm used in this study for DNA sequencing originated from an experimental study in Italy (Genchi et al., 2013). It had been collected from a beagle dog (ID 9478-9800) during necropsy and was continuously stored at −20 °C.

Blood samples treated with EDTA were obtained from dogs naturally infected with *D. repens.* During routine diagnosis, microfilaraemia was detected in three dogs from a shelter in Lithuania and their blood was used for RNA sequencing, whereas mf from blood of three privately-owned dogs diagnosed in Croatia was used to produce L3 of *D. repens* in mosquitoes. The fresh blood samples from Lithuania were processed on site up to the worm homogenization step (see Section 2.3). The samples were stored at −40 °C overnight and then shipped with a courier to the Institute of Parasitology, University of Zürich, Switzerland. Immediately after arrival, they were placed at −80 °C and further processed the following day. EDTA-anticoagulated blood samples from Croatia were placed in a polystyrene box containing cooling elements to keep mf alive and sent the next day by courier to Zürich.

Vitality and number of mf were confirmed by microscopy as described in a previous study with the modification that microfilariaemia was calculated as average from four counts (Silaghi et al., 2017).

The infections with *D. repens* and the absence of *D. immitis* were confirmed by traditional PCR as described previously (Rishniw et al., 2006). PCRs were run on the PTC-200 Peltier thermal cycler (Bio-Rad, Reinach, Switzerland), and the amplicons examined on a 1.5% agarose gel under UV light.

2.2. Mosquitoes and production of L3

A laboratory colony of *Aedes aegypti* was maintained in a climate chamber in an insectarium under standard laboratory conditions at a temperature of 27 °C, a relative humidity (rh) of 85% and a light:dark cycle of 16:8h including dusk/dawn phases of 1 h. Mosquitoes were kept in purpose-built plastic boxes (dimensions: 32.7 cm × 16.3 cm × 22.7 cm, volume 8.2 L) containing racks of plates to increase the surface. On average, each box contained 4000 mosquitoes. Sugar cubes and water were provided ad libitum. For reproduction, the mosquitoes were fed with EDTA-anticoagulated bovine blood from the local slaughterhouse using a standard artificial feeding system (Hemotek, Hemotek Ltd, Lancashire, UK).

Mosquitoes were fed at room temperature for 2 h through stretched Parafilm membranes in a Hemotek system on 3 ml of spiked blood at 37 °C (Silaghi et al., 2017). Three biological replicates were done with blood samples containing 4675, 4000 or 150 mf/ml, respectively. After exposure to the blood meal, all mosquitoes (fed and unfed) were kept under the standard conditions mentioned above. After 14 days, the mosquitoes were anaesthetized with diethylether (≥95.5%) and gently ground with a mortar and pestle in order to enable the L3 to escape the mosquitoes' bodies without damaging them. The material was then rinsed onto a 100 µm mesh sieve placed in a glass bowl (22 cm diameter) containing Hank's balanced salt solution (HBBS) (Thermo Fisher Scientific, Waltham, USA) for 2.5 h at 27 °C and 85% rh. The L3 migrated through the sieve into the warm HBSS to the bottom of the glass bowl. The L3 were then filtered through two sieves (mesh sizes 32 µm, 22 µm) and single L3 were picked using a binocular and put into RNAlater (Thermo Fisher), then stored at −20 °C overnight.

2.3. Nucleic acid extraction

The *D. repens* worms were placed in a 2 ml Eppendorf tube with 500 µl of Tris-EDTA buffer and disrupted with a Tissue Lyser II (Qiagen, Hilden, Germany) with one 5 mm stainless steel bead at 30 beats/s for 1 min. This step was repeated twice. In between the disruption processes, the sample was centrifuged at room temperature for 30 s (10,000g). Genomic DNA was extracted using the Gentra Purogene Tissue Kit (Qiagen) according to the manufacturer's instructions, using the tissue protocol for processing 50–100 mg tissue. The sample was stored at −20 °C until further processing.

A total of 4–6 ml of mf-containing blood was drawn into a 10 ml syringe and immediately diluted and lysed in the syringe with diethyl pyrocarbonate (DEPC)-treated distilled water (prewarmed to 37 °C). The solution was injected through a Difil-filter system (mesh size 5 µm) (Evsco, Buena, USA) and thoroughly washed with DEPC-treated water (37 °C). The filter was transferred to a 2 ml Eppendorf tube filled with 1 ml of Trizol reagent (Invitrogen, Carlsbad, USA) and cut into small pieces with sterile scissors. Five cycles of flash freezing in liquid N_2 and crushing with plastic pestles were performed to obtain homogeneous worm extracts according to a method described previously (Ballesteros et al., 2016). RNA was isolated by extraction with the Trizol LS kit (Invitrogen), followed by column purification using the RNeasy mini kit (Qiagen) according to the manufacturer's instructions. Briefly, the homogenized samples were incubated for 5 min at room temperature to permit the complete dissociation of nucleoprotein complexes. Two hundred µl of chloroform were added to each tube and the samples were vortexed and incubated for 3 min at room temperature. The samples were then centrifuged at 12,000g at 4 °C for 15 min. The aqueous phase was transferred to fresh tubes and mixed with an equal volume of cold ethanol (70%). The tubes were vortexed for 5 s and the mixture transferred to an RNA binding spin column from the RNeasy kit (Qiagen). In order to eliminate genomic DNA contamination, the samples were treated with DNase using an RNase-Free DNase Set DNA-free Kit (Qiagen).

L3 were concentrated by filtration (Difil test) and washed with DEPC-treated distilled water. The filters were transferred to 2 ml Eppendorf tubes containing 1 ml of TRIZOL and cut into small

pieces with sterile scissors. The samples were disrupted in a Tissue Lyser II (Qiagen) with 15–20 glass beads (ø 2.85–3.45 mm) at 30 beats/s for 1 min. RNA was isolated as described for microfilariae.

2.4. DNA sequencing

Single-molecule real-time sequencing (SMRT) bell templates were produced using the DNA Template Prep Kit 1.0 (Pacific Biosciences, Menlo Park, USA) (p/n 100-259-100). The input genomic DNA concentration was measured using a Qubit Fluorometer double-stranded (ds)DNA Broad Range assay (Thermo Fisher Scientific) (p/n 32850), revealing a concentration of 88 ng/µl (total amount 25.2 µg). A total of 10 µg of genomic (g)DNA was mechanically sheared to an average size distribution of 15 kb, using a Megaruptor Device (Diagenod, Seraing, Belgium). A Bioanalyzer 2100 12 K DNA Chip assay (Agilent, Santa Clara, USA) (p/n 5067-1508) was used to assess the fragment size distribution. Five µg of sheared gDNA were DNA-damage repaired and end-repaired using polishing enzymes. A blunt-end ligation reaction followed by exonuclease treatment was performed to create the SMRT bell template. A Blue Pippin device (Sage Science, Beverly, USA) was used to size select the SMRT bell template and enrich large fragments >7 kb. The sized selected library was quality inspected and quantified on the Agilent Bioanalyzer 12 kb DNA Chip and on a Qubit Fluorimeter, respectively. A ready to sequence SMRT bell-Polymerase Complex was created using the P6 DNA/Polymerase binding kit 2.0 (Pacific Biosciences) (p/n 100-236-500) according to the manufacturer's instructions. The Pacific Biosciences RS2 instrument was programmed to load and sequence the sample(s) on nine SMRT cells v3.0 (Pacific Biosciences) (p/n100-171-800), taking one movie of 240 min each per SMRT cell. A MagBead loading (Pacific Biosciences) (p/n 100-133-600) method was chosen in order to improve the enrichment the longer fragments.

2.5. RNA sequencing

The quality of the isolated RNA was determined with a Qubit (1.0) Fluorometer and a Bioanalyzer 2100. Only those samples with a 260 nm/280 nm ratio between 1.8–2.1 and a 28S/18S ratio within 1.5–2 were further processed. The TruSeq RNA Sample Prep Kit v2 (Illumina Incorporation, San Diego, USA) was used in the succeeding steps. Briefly, total RNA samples (100–1000 ng) were poly A enriched and then reverse-transcribed into ds cDNA. The cDNA samples were fragmented, end-repaired and polyadenylated before ligation of TruSeq adapters containing the index for multiplexing fragments containing TruSeq adapters on both ends were selectively enriched with PCR. The quality and quantity of the enriched libraries were validated using Qubit (1.0) Fluorometer and the Caliper GX LabChip GX (Caliper Life Sciences Incorporation, Waltham, USA). The product is a smear with an average fragment size of approximately 260 bp. The libraries were normalized to 10 nM in Tris-Cl 10 mM, pH 8.5 with 0.1% Tween 20.

The TruSeq PE Cluster Kit HS4000 or TruSeq SR Cluster Kit HS4000 (Illumina) was used for cluster generation using 10 pM of pooled normalized libraries on the cBOT sequencing system (Illumina). Sequencing was performed on the Illumina HiSeq 2000 paired end at 2 X101 bp or single end 100 bp using the TruSeq SBS Kit HS4000 (Illumina).

2.6. Data analyses

A detailed description of all the steps in the data analysis, and of the methods used, is available in Supplementary Data S1. Briefly, long fragments sequenced on the Pacific Bioscience RSII were trimmed and filtered. The final draft of the genome was obtained using Falcon and the annotation was performed using the Maker2 pipeline (English et al., 2012).

De novo transcriptome assembly was performed using the Trinity software suite v.2.2.0 (Grabherr et al., 2011) using default parameters followed by CDS prediction with Transdecoder (http://transdecoder.sourceforge.net/) and cd-hit-est (http://weizhongli-lab.org/cd-hit/).

The RNA sequences were aligned to our annotated genome using STAR (Dobin et al., 2013); expression levels were quantified using the R package GenomicRanges (Lawrence et al., 2013) and differential expressions were tested for using the R package edgeR (Robinson et al., 2010).

Finally, GO over-representation analyses were performed using PANTHER (Mi et al., 2017).

3. Results and discussion

3.1. De novo genome assembly and annotation

The use of long reads to assemble the genome of *D. repens* resulted in 916 total contigs, representing an over 30-fold increase in compactness compared with the latest release of *D. immitis* (www.nematodes.org). Similarly, a N50 of 584,065 bp equates to an over 50-fold increase in contiguity. The size of the assembled *D. repens* genome was just below 100 MB, approximately 17% larger than that of *D. immitis*, and their GC contents differed by only 0.7% (Table 1). In order for these comparisons to be appropriate, all the metrics in Table 1 refer to contigs. The annotation of the de novo genome assembly predicted 10,357 genes. This gene

Table 1
Summary of some relevant metrics associated to the assembly and annotation of the *Dirofilaria repens* and other filarial nematodes.

Metric	*Dirofilaria repens*[a]	*Dirofilaria immitis*[b]	*Brugia malayi*[b]	*Loa loa*[b]	*Caenorhabditis elegans*[c]
Assembly size (Mb)	99.59	78.16	87.95	96.36	100.27
Number of contigs	916	11,654	205	2183	6
Contig N50 (bp)	584,065	15,962	10,384,967	180,288	17,493,829
Largest contig (bp)	4,800,701	195,591	23,878,448	1,570,872	20,924,180
Protein-coding gene models	10,357	11,375	10,959	12,473	20,362
Predicted proteins	11,262	12,344	15,393	12,473	32,061
Protein-coding sequence (%)	15.5	18	13.8	5.4	25.4
Annotated genes	6467	8113	NA	8592	10,190
Annotated transcripts	7044	8113	NA	8592	17,862
Overall GC content (%)	27.6	28.3	28.5	30.8	35.4
Median exons per gene	7	5	5	9	6
Median exon size	136	142	139	138	147

[a] Source: present study.
[b] Source: https://parasite.wormbase.org.
[c] Source: https://wormbase.org.

number is approximately 8.9% fewer than those predicted in *D. immitis*, and this difference is consistent at the transcript level with 11,262 transcripts predicted, approximately 8.8% fewer than those predicted in *D. immitis*. On average (median), *D. repens* shows a larger number of exons per gene (seven versus five). At the same time, we report the exons being slightly shorter (136 versus 142 bp, median size, Table 1). Such differences in the annotation could be partially reflecting the biological difference between the two nematodes and could partially be a consequence of the fact that the *D. repens* transcriptome has been assembled from RNA samples extracted from microfilariae and L3 stages whereas that of *D. immitis* is based on male and female adult samples (Godel et al., 2012). Moreover, the fact that the annotation are less divergent than the assembly suggests that the most likely reason for such a difference in genome size lies in the repetitive regions that the all-short-reads assembly of *D. immitis* has not been able to resolve. By looking at some of the common quality metrics for a genome annotation, we find that more than 95% of the transcripts have an Annotation Edit Distance (AED, the main transcript quality score produced by MAKER2) below 0.5 (Supplementary Fig. S1) and over 60% of the transcripts were annotated to UniProtKB/Swiss-Prot. By comparing these values with the MAKER2 benchmarks (Campbell et al., 2014), we can state that the overall quality of our *D. repens* annotation is definitely high.

Of the 206 proteins that did not have an ortholog in the *D. immitis* transcriptome, 141 presented a homologous sequence in the genome of *D. immitis*, suggesting that they are indeed present in *D. immitis*, but that the current *D. immitis* annotation has not been able to produce a gene model for them. Of the remaining 65, 16 could be annotated to UniProtKB/Swiss-Prot (Supplementary Table S1). Of particular interest could be the only protein out of the aforementioned 65 (augustus-000063F-processed-gene-0.33-mRNA-1) that, despite showing no ortholog either in the *D. immitis* transcriptome or genome, was found in all other considered filarial nematodes (*Brugia malayi, Brugia pahangi, D. repens, Loa loa, Onchocerca volvulus, Wuchereria bancrofti*) and *Caenorhabditis elegans*. A homolog search using BLAST in UniProtKB/Swiss-Prot revealed that the only match is in the organism *Plasmodium reichenowi* with the circumsporozoite protein, the immunodominant surface antigen on the sporozoite (the infective stage of the malaria parasite that is transmitted from the mosquito to the vertebrate host).

A putative set of 834 filaria-specific proteins has been reported (i.e. present in *D. immitis* and *B. malayi*, but absent in *C. elegans*, *Trichinella spiralis* and *Ascaris suum*) (Luck et al., 2014). Among

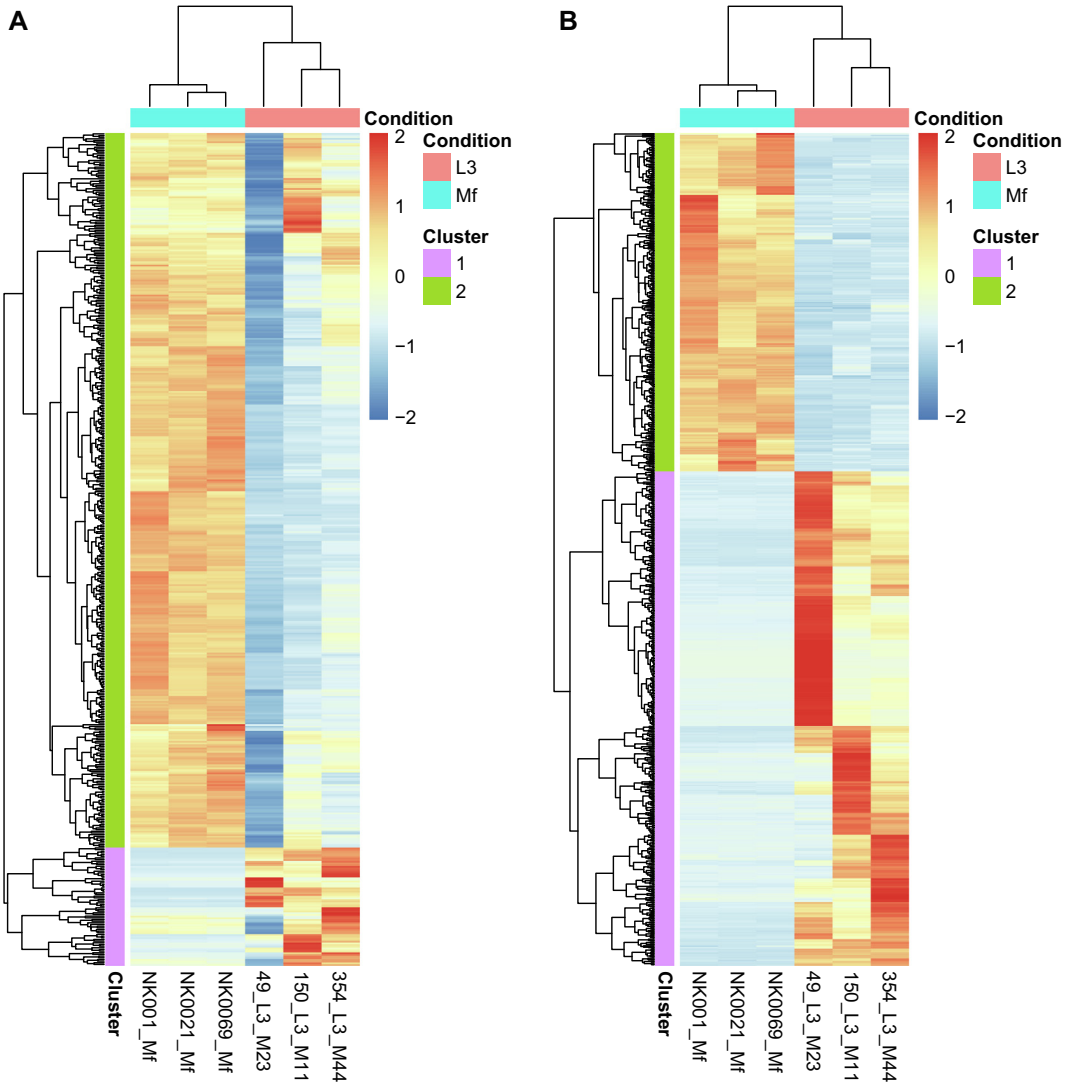

Fig. 1. Hierarchical clustering of the *Dirofilaria repens* genes and samples based on gene expression data using the R function *hclust*. (A) Unsupervised clustering based on the top 500 genes as ranked by variance across all the samples. (B) Clustering based on the genes with false discovery rate < 0.05 in the differential expression analysis. Mf, microfilariae.

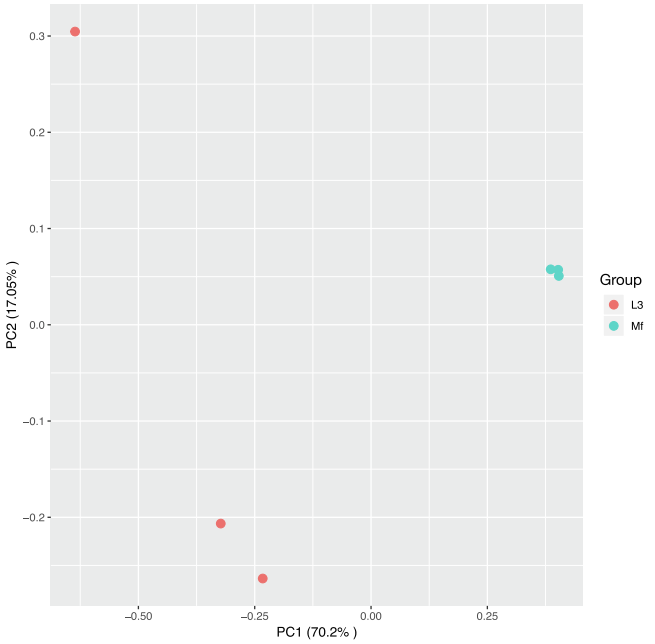

Fig. 2. Principal component analysis based on the top 500 *Dirofilaria repens* genes as ranked by variance across all the samples. The number on the axis represents the percentage of variation explained by the two main components (PC1 and PC2). Mf, microfilariae.

the proteins identified in our annotation, 712 overlap with the aforementioned set, 448 of which are annotated to UniProtKB/Swiss-Prot. Of the remaining 122, 50 could be annotated to UniProtKB/Swiss-Prot (Supplementary Table S2). It is possible that the difference in the qualities of the assemblies plays a role, however, in both cases the N50 is well above the average gene length, and therefore the discrepancies at the annotation level, as expected, are much less severe. We would rather argue that this is simply a consequence of the fact that our *D. repens* annotation has fewer transcripts predicted overall than that of *D. immitis*.

Since the biology of *L. loa* is more similar to that of *D. repens* than it is to *D. immitis*, we checked whether some differences could be identified at the molecular level. More precisely, we compared the proportion of ortholog transcripts between these species. Since we appreciate that the sets of predicted proteins of the current annotations of these organisms might be incomplete, we first aligned the six-frame translated genomes to the *L. loa* proteins, and this resulted in only approximately 15% of open reading frames (ORFs) finding a match. We then also used Blat to align the transcripts against the genome. The latter resulted in over 80% of the transcripts finding a match. Interestingly, in both cases, we found that the *L. loa* proteome shows a significant enrichment ($P < 0.01$ and $P < 0.00005$, respectively) for *D.repens* proteins compared with *D.immitis* proteins. This similarity resembles the macroscopic traits associated with these nematodes. In particular, *L. loa* is endemic to central Africa and causes ocular and systemic symptoms in humans. Similar to *D. repens*, adult *L. loa* are located in loose connective tissue beneath the skin and between the fascial layers on top of somatic muscles (Whittaker et al., 2018).

Supplementary Fig. S2 shows the Gene Ontology Biological Processes (GO BP) which are shared by *D. repens* and *L. loa*, but not present in *D. immitis*. Interestingly, many of the significant GO_BP are motility-related (locomotion, regulation of locomotion) and feeding-related (regulation of pharyngeal pumping).

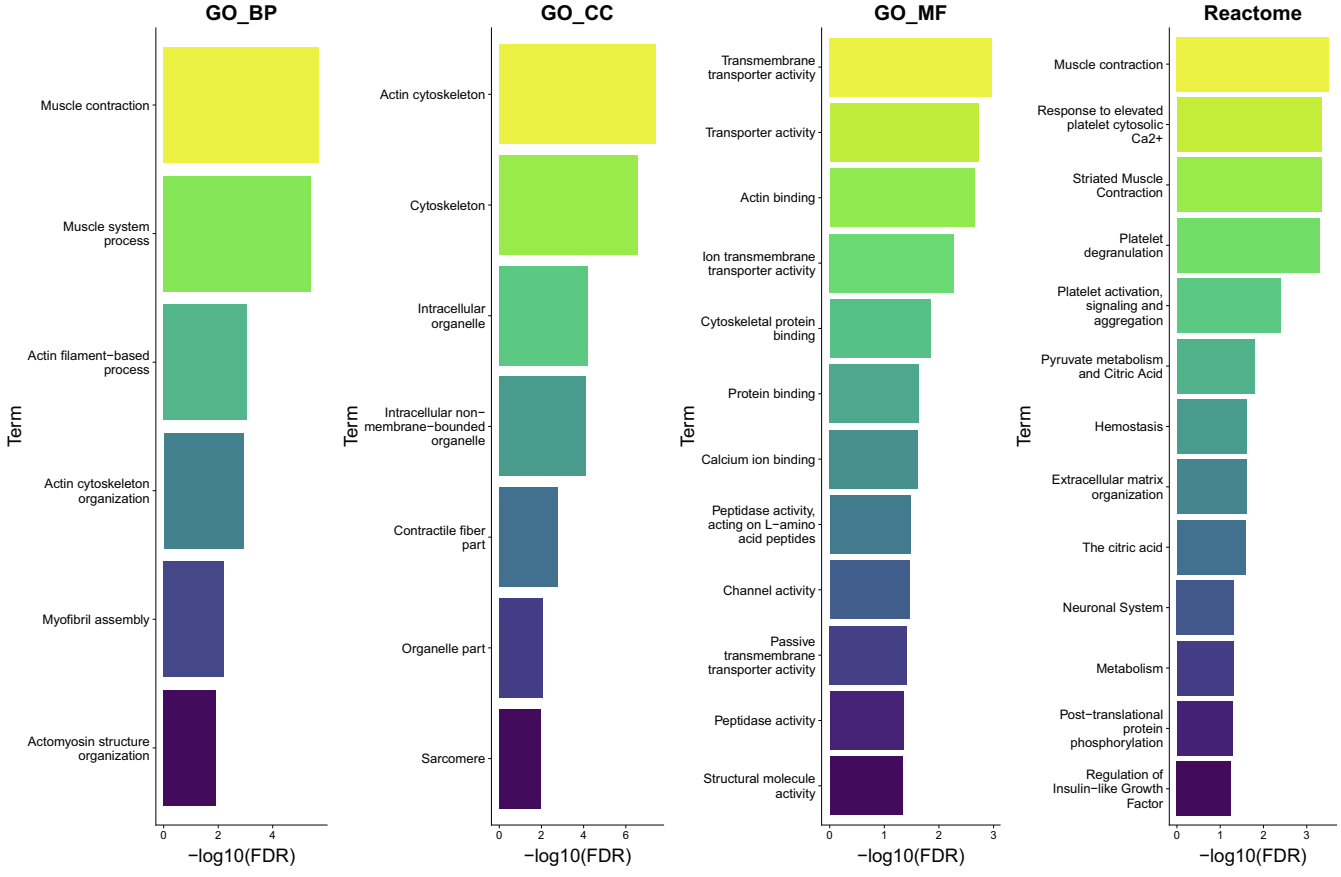

Fig. 3. List of Gene Ontology databases and Reactome pathway terms with an enrichment analysis false discovery rate < 0.05 based on the genes upregulated in the *Dirofilaria repens* L3 stage. GO_BP, Gene Ontology Biological Processes; GO_CC, Gene Ontology Cellular Component; GO_MF, Gene Ontology Molecular Function.

3.2. Gene expression analysis

The analysis of the RNA sequencing data of mf and L3 revealed a clear separation between the two groups upon unsupervised clustering (Fig. 1A) and Principal component analysis (PCA) (Fig. 2). By scrutinising these results, one also notices that the intra-group analysis of the RNA sequencing data revealed a lower variance among the mf samples than among the L3 samples. However, since the first principal component (inter-groups) accounts for a percentage of variance more than four times that of the second component (intra-groups), this effect can be considered small compared with the intra-group variance. Moreover, by looking at the intra-group scatter plots, also showing a higher overall homogeneity of the group Mf (smaller dispersion), no particular sample is an outlier (Supplementary Fig. S3). Substantial differences between the groups were confirmed by the differential expression analysis, with 876 genes reported to be differentially expressed at thresholds of 0.05 for the false discovery rate (FDR). Of these, 591 could be annotated to UniProtKB/Swiss-Prot. Cluster signatures for the upregulated genes in the two groups can be readily recognized when plotting the expression profiles in a heatmap (Fig. 1B).

For further discussion, only genes with an UniProt annotation to the filarial nematodes *D. immitis, B. malayi, B. pahangi, O. volvulus,* the model organism *C. elegans* or a UniProtKB/Swiss-Prot annotation to the endosymbiont bacteria *Wolbachia* sp. and *Wolbachia pipientis* have been considered.

In the L3 stage, 358 genes were upregulated compared with the mf stage. Out of these upregulated genes, 155 genes had a UniProtKB/Swiss-Prot annotation to *D. immitis, C. elegans, B. malayi, B. pahangi, O. volvulus,* and none to *Wolbachia* spp. Many of the highly transcribed genes were described as genes that are part of structural components (e.g. cuticle collagens (P34687, P18835, P18833), are involved in muscle development (O01761), muscle contraction and locomotion (P02566, Q11176, P19625, Q11176) or are genes that code for proteins used for hydrolysis and proteases (Q04457, O44451).

The identification of genes that are uniquely upregulated in the L3 stage provide potential targets for the development of improved diagnostic tools to screen mosquitoes for infective stages by detecting these over-expressed genes based on cDNA. One gene (GenBank accession number P21249) that was upregulated in the *D. repens* L3 stage is described as "major antigen" or "myosin-like antigen" (OVT1) in *O. volvulus*, a parasitic filarial nematode (https://www.uniprot.org/uniprot/P21249). There are two potential homologues of the OVT1 gene (nDi.2.2.2.t09053-RA and nDi.2.2.2.t01724-RA) in the heartworm *D. immitis*, which both are upregulated in the *D. immitis* L3 and the L4 stage compared with the mf (Luck et al., 2014). One of these genes (nDi.2.2.2.t01724-RA) is also upregulated in the L4 stage compared with the L3 stage, which indicates that these genes are likely to be stage markers. Thus, these genes seem to be promising targets for diagnostic PCR of infective L3 stages in mosquitoes. However, final conclusions on the suitability of this target cannot be made yet, as deeper stage-specific analyses of the further mosquito-associated stages (sausage stage L1, L2) of *D. repens* are required. These stages are immobile, in contrast to the highly motile L3, and cannot easily be recognized under the binocular after grinding the mosquitoes, and cannot be separated from the Malpighian tubules. Although high quantities of RNA were extracted from samples containing sausage stage L1 or L2 at days 5 and 8 p.i. respectively, the quality was very low, probably consisting mainly of degraded mosquito RNA, and therefore could not be used for further analysis (data not shown).

In the mf stage, a total of 235 genes were upregulated compared with the L3 stage. Out of these, 57 genes had a UniProtKB/Swiss-Prot annotation to the same nematodes mentioned above and eight to *Wolbachia* spp. Many of these upregulated genes are located in the cell nucleus and concern DNA (P04255) and RNA binding (Q09524), transcription (Q9NAL4), chemotaxis (Q8I7F8), and some genes are described as stress response genes (P29778). In total, seven *Wolbachia* proteins were upregulated in the mf stage, including elongation factors (Q5GSU1, Q73H58) and chaperon proteins (B3CNB5, Q73I71), indicating high protein synthesis activity. Corresponding to the findings with *D. immitis*, upregulation of *Wolbachia* genes was only detected for mf stages (Luck et al., 2014). By mapping our transcripts to those of the study on *D. immitis* (Luck et al., 2014), we identified 59 and three genes reported to be significantly upregulated in both studies for the L3 and mf groups, respectively (Supplementary Table S3) and the effect sizes for these genes are remarkably comparable (Supplementary Fig. S4). Many more such genes were identified (254) in that study, however a non-negligible part of those probably represent false

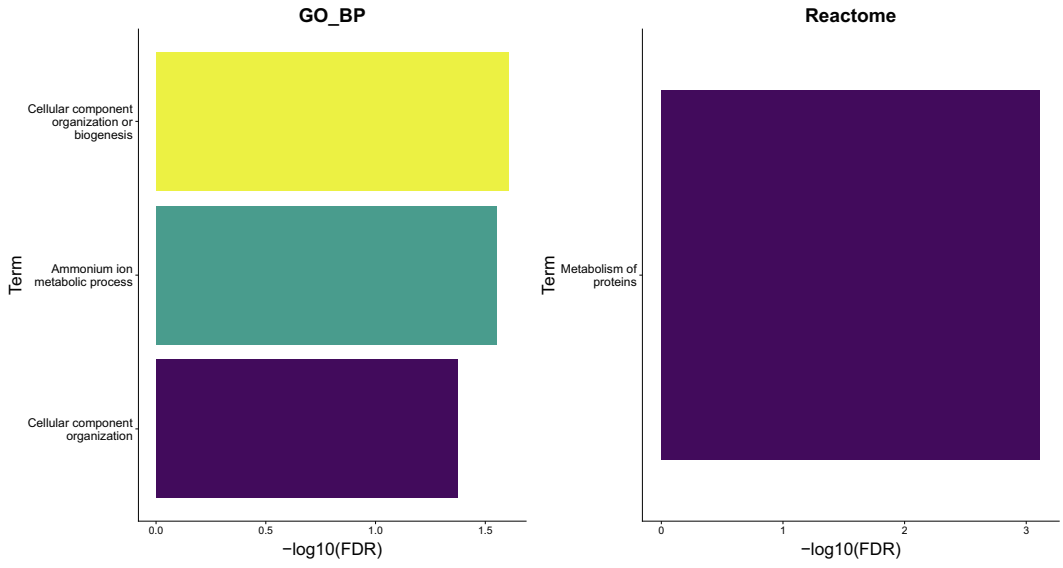

Fig. 4. List of Gene Ontology databases and Reactome pathway terms with an enrichment analysis false discovery rate < 0.05 based on the genes upregulated in the *Dirofilaria repens* microfilariae stage. GO_BP, Gene Ontology Biological Processes.

positives associated with the absence of biological replicates ($n = 1$).

In order to perform exploratory gene set analyses, we used a subset of 153 genes which were upregulated in the L3 group and mapped to the model organism *C. elegans* and the subset of 57 genes which were upregulated in the mf group and mapped to *C. elegans* to interrogate the GO databases (sub-ontologies: Biological Processes (BP), Cellular Components (CC), Molecular Function (MF)) and the Reactome Pathway database using PANTHER. After a gene enrichment analysis, several significantly enriched terms were found in all the databases for the genes upregulated in the L3 group (Fig. 3), while only a handful of GO_BPs and one generic pathway were reported for the mf stage unregulated genes (Fig. 4). The significantly enriched GO_BP for the L3 group could be classified into six terms (Fig. 3). These findings match with the biology of the L3 of *D. repens*, which have to migrate from the Malpighian tubules to the mosquito vector's mouthparts (proboscis) and penetrate the host's skin at the bite site (Simon et al., 2012). Soon after the infection, the L3 moult to L4. For this, the L3 have to be motile and contract their muscles, as well as change their anatomical structure (prepare for moulting and morphogenesis of the anatomical structures). This is supported by the fact that also the GO_CC and GO_MF as well as the reactome which are significantly enriched are involved in muscle contractions and related metabolism.

For the mf group, only three significantly enriched GO_BP could be identified with the gene enrichment analysis (Fig. 4), which are basically associated with cellular component organization or biogenesis and ammonium ion metabolic process. Microfilariae represent the first developmental stage of *D. repens* preparing for life in a

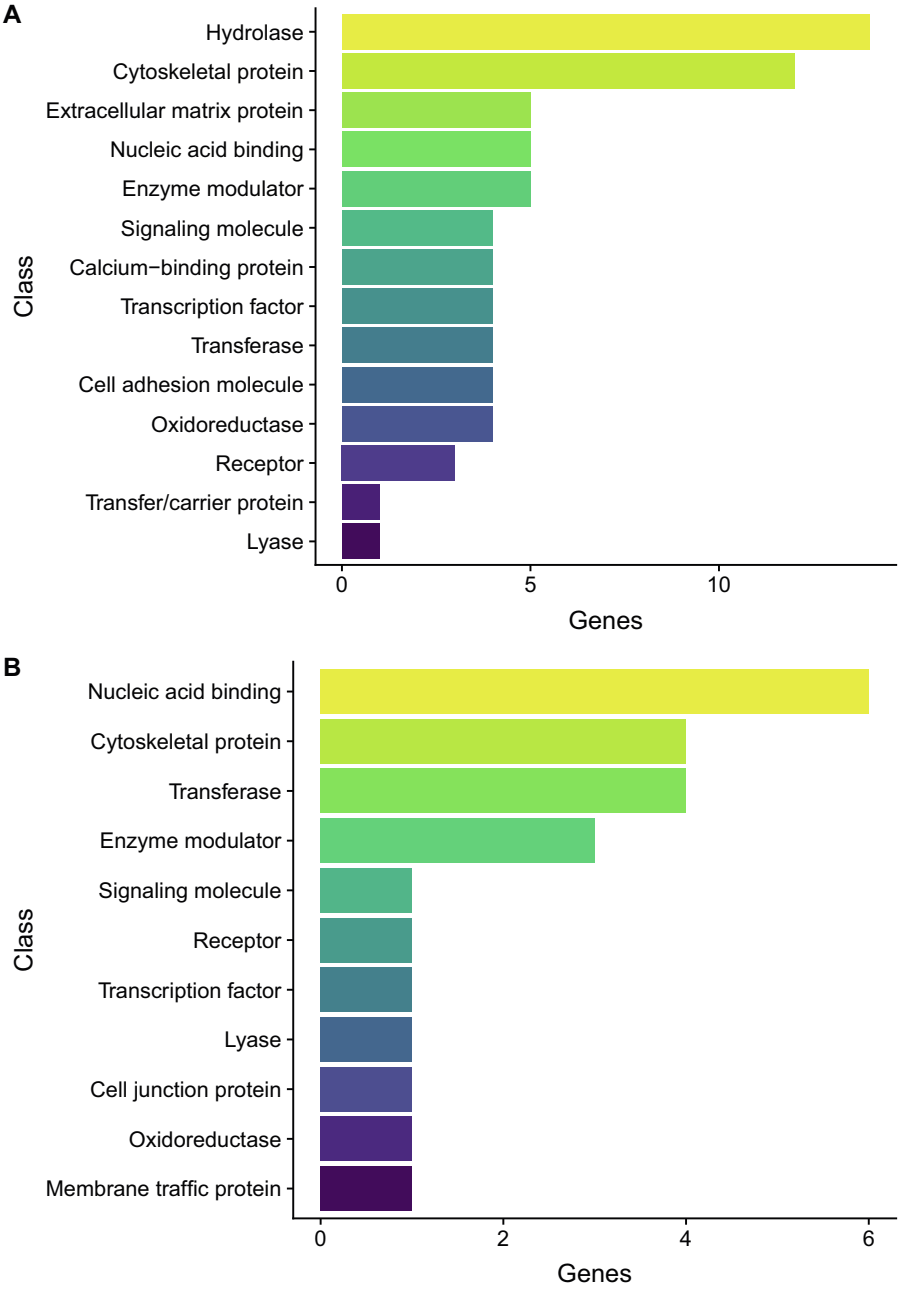

Fig. 5. Distribution of the differentially regulated genes in the various PANTHER protein classes. (A) Genes upregulated in the *Dirofilaria repens* L3 stage. (B) Genes upregulated in the microfilariae stage.

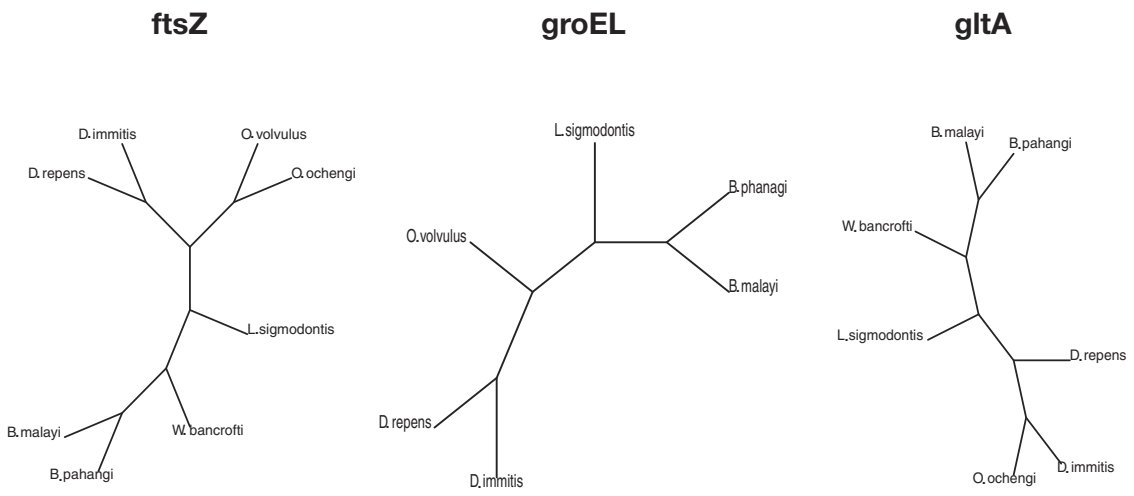

Fig. 6. Phylogenetic trees of the filarial *Wolbachia* based on the sequences of the ftsZ (cell division protein), groEL (heat-shock protein 60) and gltA (citrate synthase) genes. *Wolbachia* sequences were identified from *Dirofilaria repens Dirofilaria immitis Brugia malayi, Brugia pahangi, Onchocerca volvulus, Onchocerca ochengi, Wuchereria bancrofti* and *Litomosoides sigmodontis*.

different host (invertebrate versus vertebrate), a phase which requires many metabolic changes such as the development of internal organs. This may explain the upregulation of many GO terms involving metabolic processes in this life stage.

In order to add further depth to the potentially limited statistical power of gene set analysis when short input gene lists are used, we also looked at the distribution of the genes across the PANTHER protein classes (Fig. 5). For the L3 stage, the largest number of genes that was upregulated belonged to the class of hydrolases and cytoskeletal proteins (Fig. 5A), and in the mf stage to genes encoding proteins for nucleic acid binding, but also for cytoskeletal proteins which can be explained as both larval stages undergo morphological changes and are under development.

In the present study, parasite material from three different European countries was used (adult worm from Italy for de novo genome sequencing, mf from Lithuania, L3 produced with material from Croatia). However, these different origins of the samples should not have biased our results, as an earlier study had revealed only very small genetic differenced in *D. repens* isolates from Europe (Yilmaz et al., 2016).

3.3. Wolbachia

The assembly of the endosymbiont *Wolbachia* (see Supplementary Data S1 for details) resulted in 44 contigs, predicting a total size for the bacterial genome of 818,820 bp which is a similar size range as the genome size of *Wolbachia* of *D. immitis* (0.92 MB) and *B. malayi* (1.08 Mb) (Godel et al., 2012). *Loa loa* does not have *Wolbachia* endosymbionts (McGarry et al., 2003). *Wolbachia* are classified into supergroups based on three genes, ftsZ (cell division protein), groEL (heat-shock protein 60) and gltA (citrate synthase) (Casiraghi et al., 2005). *Wolbachia* from *D. immitis* and *O. volvulus* for example belong to the supergroup C, while those from *B. malayi* and *W. bancrofti* belong to the supergroup D.

The phylogenetic trees based on the homologues of these genes in our *D. repens* assembly are in full agreement with those in a previous study (Casiraghi et al., 2005), i.e. *Wolbachia* from *D. repens* clusters with that from *D. immitis* (Fig. 6).

3.4. Conclusions

In this study, we present the first known genome assembly draft and annotation of *D. repens*, a filarial nematode causing subcutaneous and ocular infections in dogs and cats. By sequencing Pacbio long reads at high coverage, we have been able to generate very compact and contiguous assemblies for both *D. repens* and its endosymbiont *Wolbachia*, for which a full genome has also been missing to date. When compared with the only other *Dirofilaria* for which such data are available (*D. immitis*), the quality metric of our assembly are orders of magnitudes higher, making our dataset a valuable resource for more specific, in depth studies of *Dirofilaria*.

Moreover, the comparative analysis of the mf and L3 groups provides a snapshot of the molecular differences between these two developmental stages and a first step to better understand the mechanisms behind the specific abilities of these filarial nematodes to interact in different ways with different hosts and to ensure the completion of their life cycles.

Genomic and transcriptomic data on nematodes were exploited previously to identify novel intervention methods (drug and vaccine targets) and host-parasite interactions of filarial nematodes (McNulty et al., 2016; Gasser et al., 2017; Grote et al., 2017; Bennuru et al., 2018). At the same time, it is important to remain critical with the interpretation of newly gathered information about protein-coding genes. The advancements in genome sequencing have also led to an accumulation of hypothetical proteins, and results could be compared with possibly miss-annotated genes since many helminth annotations are still based on "primary-sequence-level search protocols" and, additionally, different methods are used to deal with hypothetical proteins (Palevich et al., 2018).

Acknowledgments

We thank Relja Beck (Croatian Veterinary Institute, Zagreb, Croatia) and Mindaugas Šarkūnas (Lithuanian University of Health Sciences, Kaunas, Lithuania) for providing parasite material; Mathieu Blaser (formerly Novartis Animal Health, St. Aubin, Switzerland) for his advice and help regarding the mosquito laboratory colony; Chandra Ramakrishnan, Lucienne Tritten and Saša Štefanić (Institute of Parasitology, University of Zürich, Switzerland) for invaluable help in laboratory and data analysis work; Sirisha Aluri (Functional Genomics Center Zurich, Zürich) for advice on RNA extraction and transcriptome sequencing. The financial contributions of the Federal Food Safety and Veterinary Office, Switzerland (as sponsor of the National Centre for Vector Entomology) and Bayer Animal Health (Germany) are gratefully acknowledged.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijpara.2019.04.008.

References

Ballesteros, C., Tritten, L., O'Neill, M., Burkman, E., Zaky, W.I., Xia, J., Moorhead, A., Williams, S.A., Geary, T.G., 2016. The effect of in vitro cultivation on the transcriptome of adult *Brugia malayi*. PLOS Neglect. Trop. Dis. 10, e0004311.

Bennuru, S., O'Connell, E.M., Drame, P.M., Nutman, T.B., 2018. Mining filarial genomes for diagnostics and therapeutic targets. Trends Parasitol. 34, 80–90.

Campbell, M.S., Holt, C., Moore, B., Yandell, M., 2014. Genome annotation and curation using MAKER and MAKER-P. Curr. Protoc. Bioinformatics 48, 4.11.1–4.11.39.

Capelli, G., Genchi, C., Baneth, G., Bourdeau, P., Brianti, E., Cardoso, L., Danesi, P., Fuehrer, H.-P., Gianelli, A., Ionica, A.M., Maia, C., Modry, D., Montarsi, F., Krücken, J., Papadopoulos, E., Petric, D., Pfeffer, M., Savic, S., Otranto, D., Poppert, S., Silaghi, C., 2018. Recent advances on *Dirofilaria repens* in dogs and humans in Europe. Parasite Vector 11, 663.

Casiraghi, M., Bordenstein, S.R., Baldo, L., Lo, N., Beninati, T., Wernegreen, J.J., Werren, J.H., Bandi, C., 2005. Phylogeny of *Wolbachia pipientis* based on gltA, groEL and ftsZ gene sequences: clustering of arthropod and nematode symbionts in the F supergroup, and evidence for further diversity in the *Wolbachia* tree. Microbiology 151, 4015–4022.

Chaisson, M.J., Tesler, G., 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinform. 13, 238.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

English, A.C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., Qin, X., Muzny, D.M., Reid, J.G., Worley, K.C., Gibbs, R.A., 2012. Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. PLOS ONE 7, e47768.

Gasser, R.B., Tan, P., Teh, B.T., Wongkham, S., Young, N.D., 2017. Genomics of worms, with an emphasis on *Opisthorchis viverrini* – opportunities for fundamental discovery and biomedical outcomes. Parasitol. Int. 66, 341–345.

Genchi, C., Genchi, M., Petry, G., Kruedewagen, E.M., Schaper, R., 2013. Evaluation of the efficacy of imidacloprid 10% / moxidectin 2.5% (Advocate®, Advantage® Multi, Bayer) for the prevention of *Dirofilaria repens* infection in dogs. Parasitol. Res. 112 (Suppl 1), 81–89.

Genchi, C., Kramer, L., 2017. Subcutaneous dirofilariosis (*Dirofilaria repens*): an infection spreading throughout the old world. Parasite Vector 10, 517.

Genchi, C., Mortarino, M., Rinaldi, L., Cringoli, G., Traldi, G., Genchi, M., 2011. Changing climate and changing vector-borne disease distribution: the example of *Dirofilaria* in Europe. Vet. Parasitol. 176, 295–299.

Godel, C., Kumar, S., Koutsovoulos, G., Ludin, P., Nilsson, D., Comandatore, F., Wrobel, N., Thompson, M., Schmid, C.D., Goto, S., Bringaud, F., Wolstenholme, A., Bandi, C., Epe, C., Kaminsky, R., Blaxter, M., Maser, P., 2012. The genome of the heartworm, *Dirofilaria immitis*, reveals drug and vaccine targets. FASEB J. 26, 4650–4661.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652.

Grote, A., Lustigman, S., Ghedin, E., 2017. Lessons from the genomes and transcriptomes of filarial nematodes. Mol. Biochem. Parasitol. 215, 23–29.

Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., Carey, V.J., 2013. Software for computing and annotating genomic ranges. PLOS Comput. Biol. 9, e1003118.

Luck, A.N., Evans, C.C., Riggs, M.D., Foster, J.M., Moorhead, A.R., Slatko, B.E., Michalski, M.L., 2014. Concurrent transcriptional profiling of *Dirofilaria immitis* and its *Wolbachia* endosymbiont throughout the nematode life cycle reveals coordinated gene expression. BMC Genomics 15, 1041.

Magi, M., Calderini, P., Gabrielli, S., Dell'Omodarme, M., Macchioni, F., Prati, M.C., Cancrini, G., 2008. *Vulpes vulpes*: a possible wild reservoir for zoonotic filariae. Vector Borne Zoonot. Dis. 8, 249–252.

McGarry, H.F., Pfarr, K., Egerton, G., Hoerauf, A., Akue, J.-P., Enyong, P., Wanji, S., Klager, S.L., Bianco, A.E., Beeching, N.J., Taylor, M.J., 2003. Evidence against *Wolbachia* symbiosis in *Loa loa*. Filaria J. 2, 9.

McNulty, S.N., Strube, C., Rosa, B.A., Martin, J.C., Tyagi, R., Choi, Y.J., Wang, Q., Hallsworth Pepin, K., Zhang, X., Ozersky, P., Wilson, R.K., Sternberg, P.W., Gasser, R.B., Mitreva, M., 2016. *Dictyocaulus viviparus* genome, variome and transcriptome elucidate lungworm biology and support future intervention. Sci. Rep. 6, 20316.

Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., Thomas, P.D., 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res. 45, D183–D189.

Palevich, N., Britton, C., Kamenetzky, L., Mitreva, M., de Moraes Mourao, M., Bennuru, S., Quack, T., Scholte, L.L.S., Tyagi, R., Slatko, B.E., 2018. Tackling hypotheticals in helminth genomes. Trends Parasitol. 34, 179–183.

Rishniw, M., Barr, S.C., Simpson, K.W., Frongillo, M.F., Franz, M., Dominguez Alpizar, J.L., 2006. Discrimination between six species of canine microfilariae by a single polymerase chain reaction. Vet. Parasitol. 135, 303–314.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Silaghi, C., Beck, R., Capelli, G., Montarsi, F., Mathis, A., 2017. Development of *Dirofilaria immitis* and *Dirofilaria repens* in *Aedes japonicus* and *Aedes geniculatus*. Parasite Vector 10, 94.

Simon, F., Siles-Lucas, M., Morchon, R., Gonzalez-Miguel, J., Mellado, I., Carreton, E., Montoya-Alonso, J.A., 2012. Human and animal dirofilariasis: the emergence of a zoonotic mosaic. Clin. Microbiol. Rev. 25, 507–544.

Slatko, B.E., Taylor, M.J., Foster, J.M., 2010. The *Wolbachia* endosymbiont as an anti-filarial nematode target. Symbiosis 51, 55–65.

Taylor, M.J., Bandi, C., Hoerauf, A., 2005. *Wolbachia* bacterial endosymbionts of filarial nematodes. Adv. Parasitol. 60, 245–284.

Whittaker, C., Walker, M., Pion, S.D.S., Chesnais, C.B., Boussinesq, M., Basanez, M.-G., 2018. The population biology and transmission dynamics of *Loa loa*. Trends Parasitol. 34, 335–350.

Yilmaz, E., Fritzenwanker, M., Pantchev, N., Lendner, M., Wongkamchai, S., Otranto, D., Kroidl, I., Dennebaum, M., Le, T.H., Anh Le, T., Ramunke, S., Schaper, R., von Samson-Himmelstjerna, G., Poppert, S., Krucken, J., 2016. The mitochondrial genomes of the zoonotic canine filarial parasites *Dirofilaria (Nochtiella) repens* and *Candidatus* Dirofilaria (Nochtiella) Hongkongensis provide evidence for presence of cryptic species. PLOS Negl. Trop. Dis. 10, e0005028.