Genomics 113 (2021) 3978–3988

Contents lists available at ScienceDirect

Genomics

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

Genome and transcriptome sequencing of the green bottle fly, *Lucilia sericata*, reveals underlying factors of sheep flystrike and maggot debridement therapy

Rebecca J. Davis [a,1], Esther J. Belikoff [a,1], Allison N. Dickey [b], Elizabeth H. Scholl [b], Joshua B. Benoit [c], Maxwell J. Scott [a,*]

[a] Department of Entomology and Plant Pathology, North Carolina State University, Campus Box 7613, Raleigh, NC 27695-7613, USA
[b] Bioinformatics Research Center, North Carolina State University, Campus Box 7566, Raleigh, NC 27695-7566, USA
[c] Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45211, USA

ARTICLE INFO

Keywords:
Lucilia sericata
Maggot debridement therapy
Whole genome
Salivary gland
Malpighian tubule
Heat shock

ABSTRACT

The common green bottle blow fly *Lucilia sericata* (family, Calliphoridae) is widely used for maggot debridement therapy, which involves the application of sterile maggots to wounds. The larval excretions and secretions are important for consuming necrotic tissue and inhibiting bacterial growth in wounds of patients. *Lucilia sericata* is also of importance as a pest of sheep and in forensic studies to estimate a postmortem interval. Here we report the assembly of a 565.3 Mb genome from long read PacBio DNA sequencing of genomic DNA. The genome contains 14,704 predicted protein coding genes and 1709 non-coding genes. Targeted annotation and transcriptional analyses identified genes that are highly expressed in the larval salivary glands (secretions) and Malpighian tubules (excretions) under normal growth conditions and following heat stress. The genomic resources will underpin future genetic studies and in development of engineered strains for genetic control of *L. sericata* and for biotechnology-enhanced maggot therapy.

List of abbreviations

MDT	Maggot debridement therapy
PMI	Postmortem interval
ES	maggot excretions and secretions
SG	salivary glands
MT	Malpighian tubules
FB	fat body
TPM	transcripts per million bp
WGCNA	weighted gene co-expression network analysis
FDR	false discovery rate
GO	gene ontology
Hsp	heat shock protein.

1. Introduction

Female *Lucilia sericata* lay eggs in carrion usually in large clutches of up to 200 eggs per female [1,2]. Developing larvae consume dead tissues, thus performing an important ecological role in decomposition [3,4]. In the United Kingdom, New Zealand and Australia, *L. sericata* females can also lay eggs in an open wound or moist area of the wool or skin of live sheep [5]. The developing larvae will consume living tissue and can cause death of the animal if not treated with insecticides. Thus, sheep myiasis or flystrike causes economic losses to farmers and is a significant animal welfare issue [5,6]. In New Zealand, *L. sericata* larvae were found in most flystrike samples that contained more than one species and *L. sericata* was responsible for about a quarter of the monospecific strikes [7]. *Lucilia sericata* is responsible for most of the sheep myiasis found in England and Northern Europe [5].

Maggot debridement therapy (MDT) has been approved by the FDA for the treatment for diabetic foot ulcers [8,9]. For MDT, *L. sericata* are reared under sterile conditions and first instar larvae are introduced into wounds. The larvae promote healing in part through digestion and mechanical removal of necrotic tissue. Unlike during flystrike in sheep,

* Corresponding author.
E-mail addresses: rlinger14@gmail.com (R.J. Davis), ejbeliko@ncsu.edu (E.J. Belikoff), andickey@ncsu.edu (A.N. Dickey), ehscholl@ncsu.edu (E.H. Scholl), benoitja@ucmail.uc.edu (J.B. Benoit), mjscott3@ncsu.edu (M.J. Scott).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.ygeno.2021.10.003
Received 28 June 2021; Received in revised form 3 September 2021; Accepted 1 October 2021
Available online 5 October 2021

L. sericata do not damage healthy living tissue in people. This debridement of necrotic tissue is essential for wound healing. Digestion of the tissues is accomplished by enzymes in the excretions and secretions (ES) from the *L. sericata* larvae. The enzymes include arginases, leucine aminopeptidase, collagenase, serine proteases (trypsin-like and chymotrypsin-like) and a metalloproteinase [9,10]. In addition to the ingestion and removal of necrotic tissue, secretion of ammonia inhibits bacterial growth by raising the pH of the wound environment [11]. Microbial growth is also inhibited by factors in the ES that have antibacterial and antifungal activity. These include several antimicrobial peptides (AMPs) such as a defensin called lucifensin [12] and lucimycin [13]. Larval excretions also contain allantoin [14], a metabolite of purines that is thought to promote wound healing [15]. Urate oxidase, the enzyme that converts uric acid to allantoin, is expressed mostly in the Malpighian tubules, the larval excretion organ [16]. Lastly, ES may actively promote wound repair through stimulating growth of human cells in the wound. For example, ES increased hepatocyte growth factor (HGF) synthesis in cultured cells and increased HGF levels were measured in femoral vein blood of patients during MDT [17]. Although some specific factors have been identified, extensive assessment of factors that could be involved in MDT have yet to be established.

In addition to its role in flystrike and MDT, *L. sericata* is also of importance in forensic entomology as female blow flies are early colonizers of human remains and *L. sericata* is found in most temperate and tropical areas of the world [1,2]. A postmortem interval (PMI) can be calculated based on the stage of development of the larvae and environmental conditions [18]. Consequently, there have been several studies documenting the development of *L. sericata* larvae under a variety of conditions, allowing the establishment of this species for PMI estimates [19–21].

Although *L. sericata* is important in several fields of study, there are few genomic resources available. Here we have assembled and annotated the genome from a highly inbred strain of *L. sericata*. To gain an insight into the genes that could be contributing to MDT and flystrike, we have performed RNAseq analyses of RNA from larval salivary glands and Malphigian tubules. For comparison, we have also analyzed RNA levels in larval fat body. Lastly, as maggots in wounds could be exposed to higher temperatures, specifically during their use in MDT or during flystrike, we have examined the heat shock response in the same three larval tissues. These genomics and transcriptomic resources will provide critical information on understanding this fly as a pest and as a biotherapy agent in MDT.

2. Results and discussion

2.1. Genome assembly and gene prediction

To reduce heterozygosity, the *L. sericata* MDLA strain was inbred for 10 generations of full sibling single pair matings. DNA was isolated from late stage (14–18 h) mixed sex embryos from the inbred stain. We consistently obtained higher molecular weight DNA from embryos than adults, which could be due to the presence of the adult exoskeleton causing shearing during isolation [22]. 20 kb libraries were prepared for PacBio DNA sequencing. 15 SMRT cells from a PacBio Sequel generated 6,110,102 subreads with 45.3 Gb of sequence. An initial genome was assembled using FALCON [23,24], which was polished with Arrow (B). The raw read coverage was 73× and the assembly included 11,157 primary contigs with a total size of 688.4 Mb. After removing low quality contigs and contaminants, the assembly included 6676 primary contigs with a total size of 617.3 Mb (Table 1). We next applied the reduction step from the Redundans pipeline as this can help remove duplicate contigs from genome assemblies [25]. This step reduced the number of primary contigs to 4372 with an assembly size of 565.4 Mb. The final assembly size is smaller than the measured genome sizes of 651.8 Mb for male and 701.1 Mb for female *L. sericata* [26]. This could indicate that parts of the genome are missing from the assembly.

Table 1
Assembly statistics and BUSCO scores before and after Redundans.

	Before redundans	Final
Total length	617.3 Mb	565.4 Mb
Contig count	6676	4372
Longest contig	4,338,675 bp	4,338,675 bp
N50	264,619 bp	296,078 bp
GC Content	30.8%	30.8%
BUSCO (Diptera)		
Complete-single	2512 (89.7%)	2576 (92%)
Complete-duplicated	180 (6.4%)	89 (3.2%)
Fragmented	39 (1.4%)	43 (1.5%)
Missing	68 (2.5%)	91 (3.3%)

To assess the completeness of the assembly, we searched for the presence of 2799 core Diptera genes utilizing BUSCO [27]. The results suggest a very complete set (92.0% complete and single copy) with little fragmentation (1.6%) (Table 1). The level of complete but duplicated genes (3.2%) was also low but suggests that a small fraction of single copy genes have been placed onto separate contigs, possibly due to residual heterozygosity. The application of Redundans did reduce the number of complete but duplicated genes by half suggesting duplicate contigs had been removed from the assembly. Ninety one of the 2799 genes were missing. The BUSCO scores are similar to those for other fly genomes [28–31], indicating the *L. sericata* draft genome is of comparable quality.

The genome was annotated by the automated NCBI eukaryotic genome annotation pipeline (NCBI *Lucilia sericata* Annotation Release 100). As part of the annotation pipeline, RNA-seq reads from adult male and female (this study) and larvae from earlier studies [32,33] were used in gene prediction and in subsequent RNA-seq analyses. The reference assembly contains 14,704 predicted protein coding genes and 1709 non-coding genes (Table 2). This number of protein coding genes is very similar to the 14,544 protein coding genes found in the initial *L. cuprina* genome assembly [31]. Principal component analyses of the RNA-seq data show that the samples cluster by stage and sex (Fig. S1). Comparisons identified female and male-biased and larval-specific gene sets (Tables S1–S3). Using Signal P [34], 3604 of the proteins were predicted to contain a secretory signal peptide indicating the protein could be secreted or retained in the plasma membrane, ER, Golgi apparatus or directed to the lysosome [34] (Table S4). Of the gene families of interest for debridement, there are 23 annotated AMP genes, including seven that encode attacin-like proteins, two defensins (one is lucifensin), three diptericin-like, ten sarcotoxin-like and one antifungal (lucimycin) (Table S5). Of the predicted proteins, there are 174 annotated peptidases, 157 proteases, 54 trypsin-like, and 17 chymotrypsin-like.

2.2. Searching for wound healing and flystrike genes: genes expressed in larval salivary glands, Malpighian tubules and fat body

To identify genes that may play important roles in MDT and flystrike, RNA was isolated from dissected larval salivary glands (SG), Malpighian tubules (MT) and fat body (FB). The SG are a major source of maggot secretions and MT are a source of maggot excretions. While the FB plays a major role in intermediary metabolism, in this study the tissue serves as a control group for pairwise comparisons. RNA-seq analyses were

Table 2
Gene annotation summary statistics.

Feature	Number
Protein-coding genes	14,704
Non-coding genes	1709
Non-transcribed pseudogenes	633
mRNAs	23,687
CDSs	23,700

undertaken to identify genes enriched in each tissue. The normalized RNA levels (TPM) for all transcripts are provided in Table S6. Two methods were used to assess significance levels in each sample, weighted gene co-expression network analysis (WGCNA, $P < 0.05$ for modules of interest) [35] and DESeq2 (FDR < 0.05) [36]. DESeq2 allows for the analysis of genes that have statistically significant difference expression between samples and WGCNA identifies genes with similar co-expression patterns between samples. From the WGCNA analysis there was a clear clustering of genes based on the level of expression in each tissue. For example, genes in the "turquoise" module show significantly higher correlation in Malpighian tubules compared to fat body and salivary glands (Fig. 1). Similarly, genes in the "blue" module show significantly higher correlation in salivary gland compared to the other tissues (Fig. 1). The transcripts in each module are listed in Table S7. The DESeq2 analysis identified genes that were more strongly expressed in the specific tissue compared to whole larvae (Fig. 1, FDR < 0.05). We then focused our efforts on genes that both approaches had identified as preferentially expressed in each tissue and which of these genes encoded predicted secreted proteins (Tables S8-S10).

The 2048 transcripts that were identified by WGCNA and DESeq2 as differentially expressed in salivary glands included many that are associated with protein metabolism, transport, exocytosis, and cell development (Table S8), which have all been associated with SG in other fly systems [29,37,38]. Three loci that were highly expressed in salivary glands encode antigen 5 like allergen proteins. The three loci are in a single cluster on one scaffold. In the stable fly *Stomoxys calcitrans*, antigen-5 like genes were very highly expressed in the SG and are thought to function as inhibitors of the classical complement system [29]. Several proteins that play important roles in protein secretion including signal peptidase complex subunits and signal recognition particle subunits show biased expression in salivary glands (Table S8), reflecting the activity of the secretory cells in the tissue. Maggot ES contain several peptidases including serine, aspartic and metallopeptidases [10,39]. Among the transcripts with biased expression in SG, were transcripts for serine proteases (snake, snake-like, furin-like, carboxypeptidase, chymotrypsin-like) aminopeptidase N, dipeptidyl peptidase 3 and aspartic protease-like from 24 transcripts. Nevertheless, this is a small fraction of the transcripts that encode annotated peptidases/proteases (24 of 419 or 5.7%). An earlier study of *L. sericata* larval transcriptomes assembled from whole body, crop, gut and salivary gland identified 185 peptidase species that were differentially expressed across larval tissues. As most of the proteases that were found in maggot ES were highly expressed in larval gut, the authors suggested that the gut is an important source of proteases found in ES, possibly due to regurgitation [39]. Two transcripts for AMPs, one for an attacin-like peptide (XM_037962634.1) and one defensin-like (XM_037971781.1) were in the blue module but were not also identified by DESeq2 as differentially expressed in salivary glands. The *Drosophila melanogaster* FlyAtlas project identified several gene families that are abundant in salivary gland transcriptomes including lipases, lyozymes and yellow-like proteins [40,41]. Transcripts for *L. sericata* orthologs of yellow (XM_037962012.1, XM_037962013.1) and yellow-like genes (XM_037962023.1) were found in the common group of 2048 transcripts expressed in salivary gland. Yellow proteins have been previously identified as components of saliva in sand flies, where these act in the modulation of immune responses during feeding [42,43]. Several lysozymes were found in the WGCNA blue cluster but were not also identified by DESeq2. The salivary gland group does include four lipases but none were highly expressed in salivary gland. The top eight transcripts with the highest expression in SG all encode predicted secreted proteins, including three transcripts for antigen-5 like allergens discussed above, two keratin-like (XM_037954133.1 and XM_037954132.1), a mucin-17-like (XM_037954131.1) and a nucleolin-like protein (XM_037964335.1) (Fig. 1). The genes for the keratin-like and mucin-17-like proteins are closely linked on a single scaffold (NW_023994267.1). BLAST searches with these proteins and the nucleolin-like protein only found significant matches with predicted uncharacterized proteins from the very closely related fly *L. cuprina*. It will be of interest to determine the importance of these very highly expressed genes for larval survival and ES activity.

There were 1746 transcripts identified by WGCNA and DESeq2 as differentially expressed in MT (Table S9). As expected, transcripts for proteins involved in purine metabolism include urate oxidase or uricase (XM_037966421.1) and xanthine dehydrogenase or *rosy* (XM_037960122.1) are highly expressed in MT. Uricase converts uric acid to allantoin, which is thought to contribute to wound healing [16]. Allantoinase converts allantoin to allantoic acid. Two *L. sericata* proteins (XP_037822102.1 and XP_037827097.1) were annotated as allantoinase or allantoinase-like enzymes but their transcripts (XM_037966174.1 and XM_037971169.1) were expressed at very low levels in all tissues. Nevertheless, it will be of interest to determine if *L. sericata* does possess allantoinase-like activity as this would reduce the amount of allantoin that is excreted. In the production of primary urine, water follows an osmotic gradient that is created by vacuolar proton ATPase and co-localized K+/H+ exchanger that facilitates a net excretion of potassium ions [44]. Other proteins important for this process are inward rectifier K+ channels, Na+/K+ ATPase and chloride channels. Transcripts for vacuolar ATPase subunits (e.g. XM_037965106.1, XM_037965125.1, XM_037968641.1, XM_037961851.1, XM_037963036.1) are highly expressed in MT as are Na+/K+ transporting ATPase subunits (XM_037952658.1, XM_037972671.1, XM_037952653.1), an inward rectifier K+ channel subunit (XM_037971714.1) and a chloride channel protein (XM_037960791.1). The most abundant transcripts in MT encode aminoacylase-1-like proteins that catalyze the conversion of N-acyl-L-amino acids to L-amino acids and carboxylate. In *D. melanogaster*, transcripts for aminoacylases are highly expressed in larval and adult MT [41]. The MT set also includes transcripts for proteins with predicted transport functions such as *white* (an ABC transporter) and sugar transport. Transcripts for detoxifying enzymes such as cytochrome P450 and glutathione-S-transferase families are abundant in the MT, as is also observed in *D. melanogaster* MT [41]. Of the predicted AMPs, four transcripts for attacin-like peptides and one for a sarcotoxin-like peptide were in the turquoise module (Table S5). However, none were also identified by DESeq2 as differentially expressed in MT. The transcripts with the highest level of expression that encode proteins with a predicted secretory signal peptide include a zinc carboxypeptidase (XP_037814177.1), two metalloprotease 1-like (XP_037824849.1, XP_037827723.1), and a venom serine protease (XP_037825788.1) (Fig. 1), which could contribute to the protease activity of maggot ES. The 364 transcripts that were identified by WGCNA and DESeq2 as differentially expressed in FB included those involved in nutrient accumulation and transport, highlighting the critical role of storing materials for subsequent development (Table S10).

Promoters from genes that are highly expressed in the SG and/or MT could be used for expression of secreted or excreted proteins for biotechnology-enhanced MDT [45]. We previously crossed a transgenic strain that expressed very high levels of the tetracycline transactivator (tTA) with a strain that carried a tTA-activated human *platelet-derived growth factor-b* (*pdgf-b*) gene. In the absence of tetracycline, tTA activated transcription of *pdgf-b* in most tissues at all stages. Some PDGF-B protein was detected in maggot ES [45]. However, the strain could not be maintained as high constitutive expression of tTA negatively impacted strain fitness. A tissue-specific gene promoter would be advantageous as it would restrict tTA expression to the larval tissues of interest for MDT while potentially having little impact on the overall fitness of the strain.

2.3. Genes induced by heat shock in larval salivary glands, Malpighian tubules and fat body

As maggots in a wound are likely exposed to higher temperatures than in the laboratory, we next isolated FB, MT and SG RNA from larvae that were exposed to a heat shock of 37 °C. We anticipated that transcripts for canonical heat shock proteins (Hsps) would be elevated in

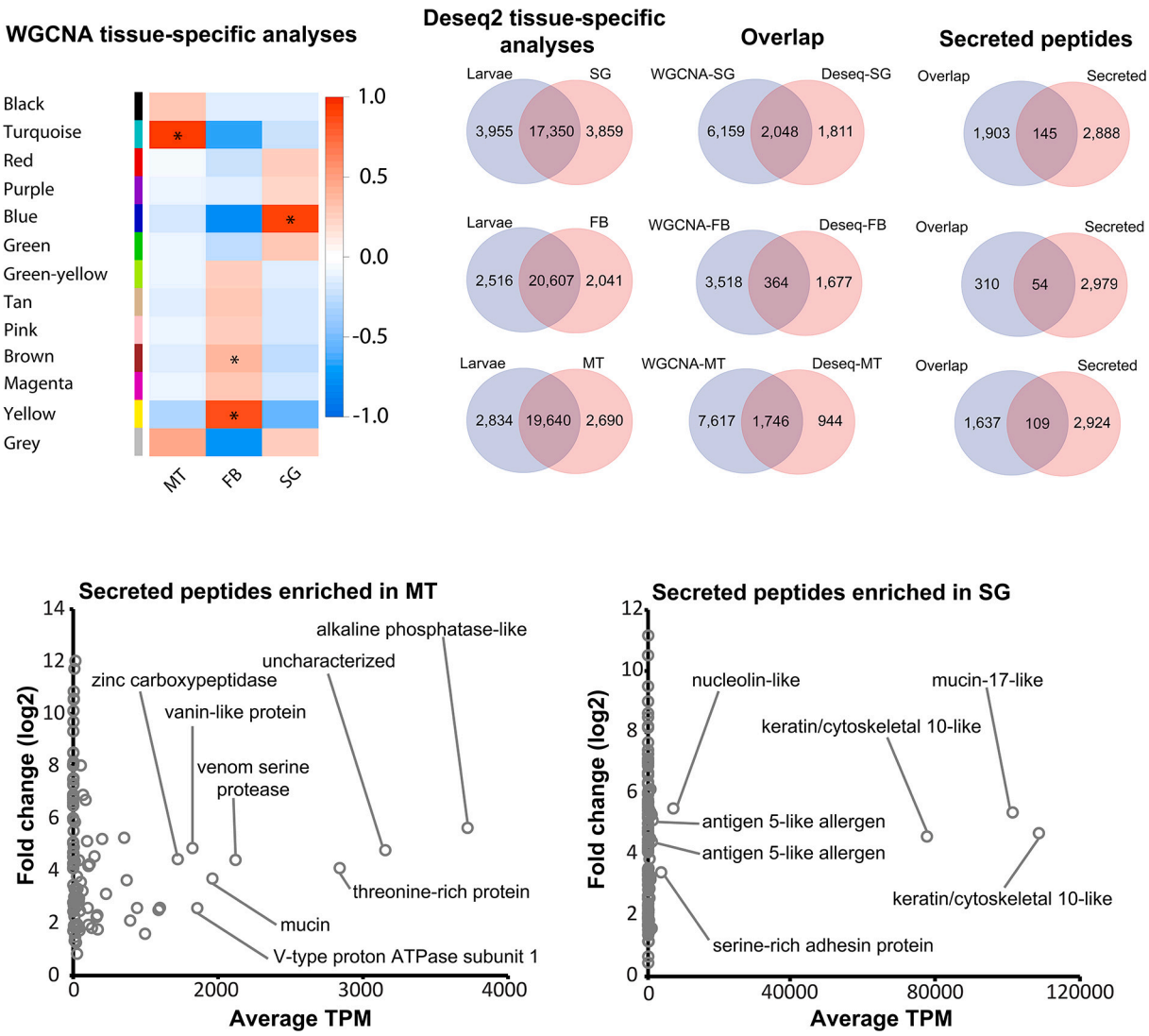


Fig. 1. Combined DESeq2 and WGCNA analyses to identify tissue specific genes set that are secreted. (Top left) WGNCA analysis to identify genes with correlated expression patterns in specific tissues. Red is indicator of increasing positive correlation and blue is increasing negative correlation. Specific groups are of co-expressed genes are listed by the color of each module. Asterisks represent values with a significant positive correlation for based on a regression-based *p*-value for assessing the statistical significance $P < 0.05$) between a tissue and module [76]. Turquoise module contains genes with correlated with the Malpighian tubules (MT), blue contains genes with correlated with the salivary glands (SG), and brown and magenta contains genes with correlated with the fat body (FB). (Top right) Comparative overlap between genes with enriched expression in specific tissues based on DESeq2 (FDR, $P < 0.05$) and correlated expression based on WGCNA following by identification of putative signal peptides to identify secreted proteins enriched in specific tissue. (Bottom) Secreted peptides with high expression levels and enrichment in the salivary glands (SG) and Malpighian tubules (MT). Base mean represents normalized reads counts over all samples determined by DESeq2 as representation of relative level of expression. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

some or all of the tissues. Of interest were any changes in expression of genes that may participate in MDT and if there was any tissue-specificity in the heat shock response.

Thirty-seven of the predicted proteins in the genome were annotated as Hsps, including 15 in the Hsp70 family (Hsp68-like, Hsp70-like), 11 small heat shock proteins (Hsp23-like, Hsp27-like, Hsp67B-like), two in the hsp60 family (Hsp60A-like) and a single Hsp83. Seven of the *hsp68/hsp70* genes (LOC119606578…LOC119606584) were in a single cluster on one scaffold, NW_023994610.1. Similarly, there were two clusters of three small heat shock protein *hsp23/hsp27* genes on scaffolds NW_023996390.1 and NW_023997200.1. The heat shock response was quite distinct for each tissue (Fig. 2), highlighting the divergent functions of each tissue. The 127 transcripts that showed increased expression in all tissues encoded proteins that were mainly associated with the heat shock response (Table S11). There was an enrichment for GO terms such as heat response, chaperone-activity and protein folding (Fig. S2). Of the *hsp* genes, *hsp83* and *hsp60A* showed a high basal level of expression and a modest heat shock response (Fig. 3 A,B), similar to what we had previously reported for *hsp83* in the closely related blowfly *L. cuprina* [47]. Four of *hsp23/hsp27* genes showed a very robust induction following heat shock, particularly in MT (Fig. 3 A,B). Six of the *hsp23/hsp27/hsp67B* genes show little expression in any tissue. The seven *hsp68/hsp70* genes that are in a single cluster all show a strong heat induction of transcription in all tissues (Fig. 3 C,D). Similarly, a *hsp70* and two *hsp23*-like genes showed a strong heat shock response in *L. cuprina* [47]. Twenty transcripts showed a decrease in level after heat shock (Table S12). A strongly heat inducible gene promoter such as from a *hsp23/hsp27* gene could be useful for biotechnology-enhanced MDT, as it could provide high levels of expression of a gene of interest (e.g. *pdgf-b*) in a wound environment [45]. A heat-inducible gene promoter could

Fig. 2. Change in transcript expression in larval tissues during the response to heat treatment mimicking wound treatment. Venn diagrams are for transcript with increased (top) or decreased (bottom) during heat exposure. Treemaps indicate gene ontology terms that are enriched following heat in all tissues and specific tissues. Full treemaps with all subcategories labeled as available as Fig. S4-S6. Full list of genes that show an increase or decrease in RNA level in all tissues are available in Tables S11, S12.

also be used to develop a strain for genetic control of *L. sericata* (Section 2.4). For example, a heat-inducible Cas9 transgene with sgRNAs targeting gene(s) required for female development or fertility [46].

The tissue-specific responses are more specialized. To help visualize the tissue-specific response we made REVIGO treemaps of the gene ontology (GO) terms that were enriched in each tissue (Fig. 2, Fig. S3-S5). The transcripts that increased in fat body after heat shock are mostly involved in metabolic processes including transcription and intracellular signaling (Fig. 2, Fig. S3). The MT show increased expression of genes involved in transcription, development and cell communication (Fig. S4). The SG have altered levels in protein synthesis and vesicle transport and endomembrane system organization (Fig. S5). Interestingly, MT show increased expression of genes involved in the positive regulation of cellular processes whereas SG are enriched for negative regulation of cellular processes. There is little overlap in the transcripts that show decreased expression after heat shock with no specific GO enrichment. Most of the SG-enriched transcripts that encode proteins with a predicted secretory signal peptide (e.g. antigen 5-like, lucifensin-like) show little change in average RNA level after a heat shock. One exception was a transcript for a predicted Kunitz-like serine protease inhibitor (XM_037971815.1), which increased by over a 100-fold after heat shock (average 0.6 without shock, 105.8 after shock, difference not significant due to wide variation in heat shock response, $p = 0.052$). The RNA level was low (<1) in the FB and MT with or without heat shock. A transcript for a lysine oxidase (XM_037970107.1) also showed an increase in level following heat shock (0.08 without shock, 1.53 with shock, $p < 0.01$). Lysine oxidase is an extracellular enzyme that plays an important role in repair of the extracellular matrix [48]. As for SG, most of the highly expressed MT-enriched transcripts that encode proteins with a predicted secretory signal peptide (e.g. peptidases) show little change in expression following heat shock. Of note are transcripts that encode annotated secreted phosphatases (XM_037951621.1 and XM_037952823.1) that are increased over 3-fold after heat shock. The largest increase in expression following heat shock of the potentially secreted proteins is for the transcript XM_037967996.1 (0.11 average before shock, 3.17 after shock, a 27-fold increase), which encodes a lectizyme (XP_037823924.1) that shows highest similarity to lectizymes found in *Glossina* species. In *Glossina*, lectizyme has both lectin and serine protease activities and is thought to play a role in the establishment of trypanosome infections [49].

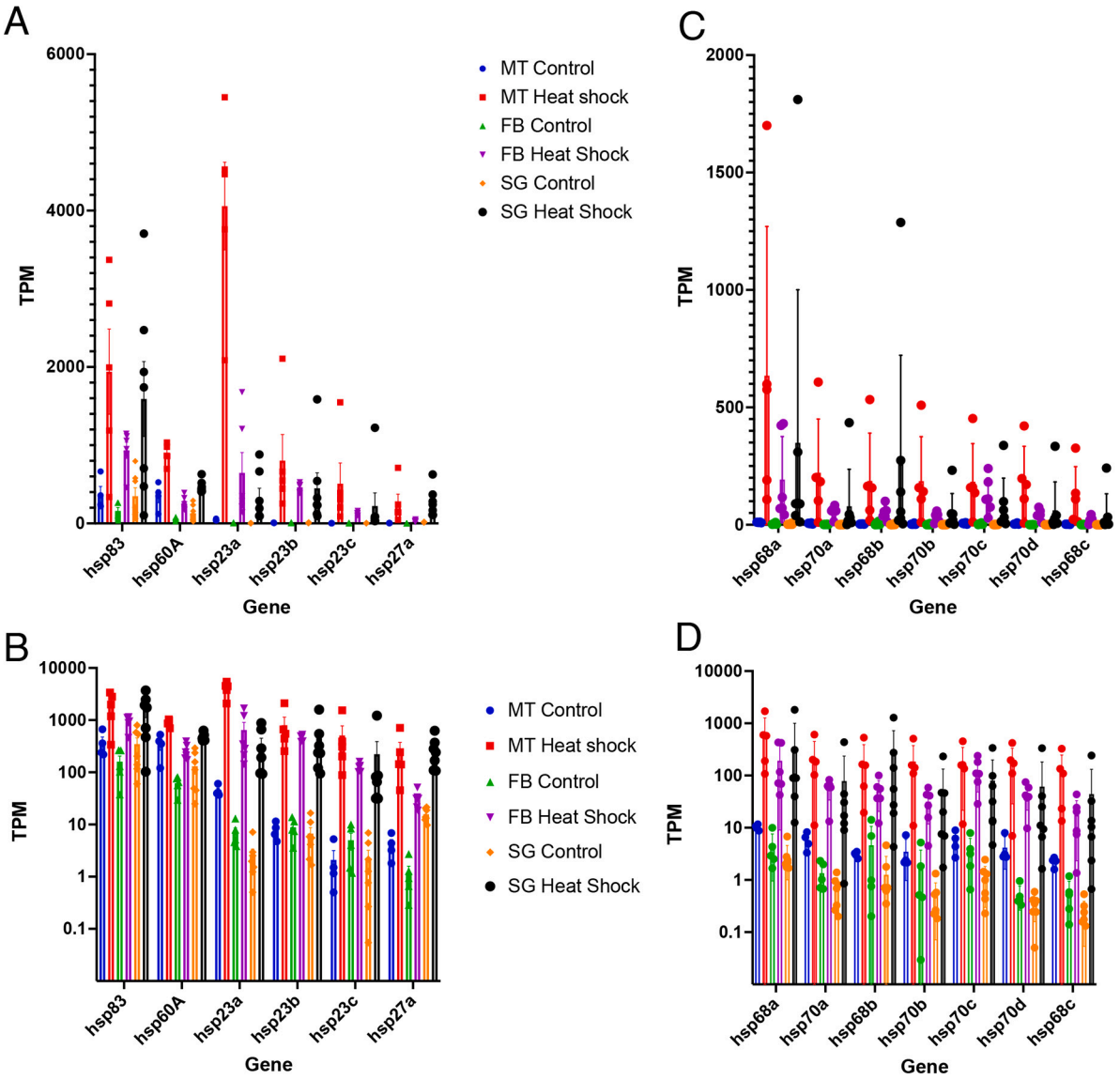
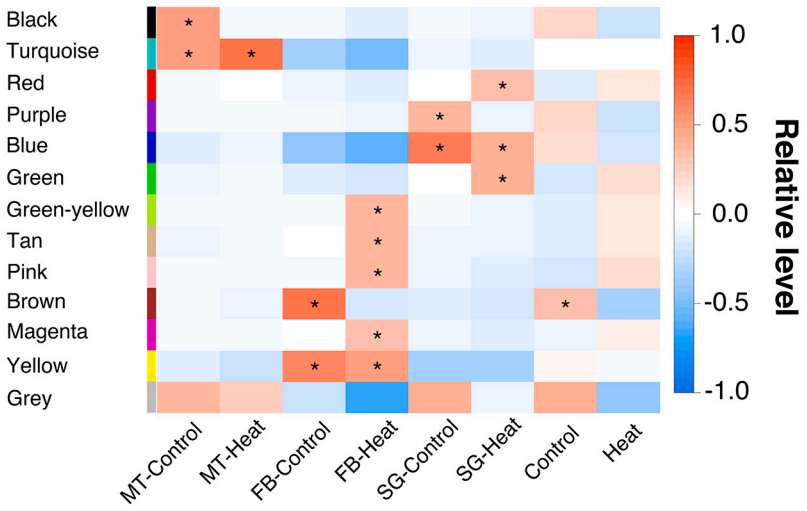


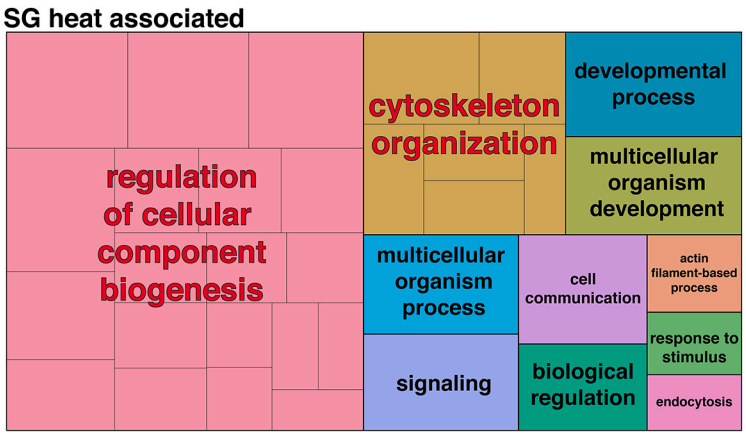
Fig. 3. Basal expression and heat induction of *hsp* gene transcripts in three larval tissues. Normalized transcripts levels (TPM) for selected *hsp* genes in tissues from control and heat shocked larvae. To highlight the basal level of expression, values were also plotted on a log10 scale for the Y-axis (B,D). All data points, mean and s.e.m. are shown.

We next undertook a WGCNA analysis of the heat shock response in the three tissues. The analysis failed to reveal correlated expressional patterns for heat stress among tissues. This lack of a correlated response in relation to heat is likely since tissues-specific impacts are more significant. For MT, no modules other than the previously identified turquoise module were found to be significantly associated with the tissue in relation to heat stress (Fig. 4). For SG, the green, red and previously identified blue module were found to be associated with the tissue after heat shock (Fig. 4, Fig. S6, Table S13). For FB, there were four new modules, green-yellow, tan, pink and magenta that were now found to be associated with the tissue and heat stress (Fig. 4, Fig. S7, Table S14). However, the brown module was no longer significantly associated with FB in the heat shocked samples. Treemaps were made of the GO terms for the proteins in the modules that were significantly associated with heat shocked SG and FB, but not the control tissue. For SG, regulation of cellular biogenesis including transcription and signaling were found as identified in the above differential expression analysis (Fig. S6). In addition, several terms related to cytoskeletal organization were identified. GO terms linked to cytoskeletal organization were also identified in the FB modules (Fig. S7). Also common were terms linked to development and regulation of gene expression including translation. These analyses highlight the profound changes in gene expression that occur in these tissues following a heat shock that are different in relation to the tissue-specific genes. The enrichment for terms associated with the cytoskeletal organization suggests an involvement of the cytoskeleton in the heat shock response. Small Hsps are known to interact with cytoskeletal proteins, which may be important for maintaining cell integrity during cellular stresses [50].

2.4. Genes for building transgenic strains for genetic control of L. sericata

Like the related blow fly *L. cuprina*, *L. sericata* is a pest of sheep in the U.K., New Zealand and other countries [5,6]. For genetic control of *L. cuprina*, we have made conditional female lethal strains that produce only males on diet that does not contain tetracycline [51–53]. Population suppression could be achieved through multiple releases of





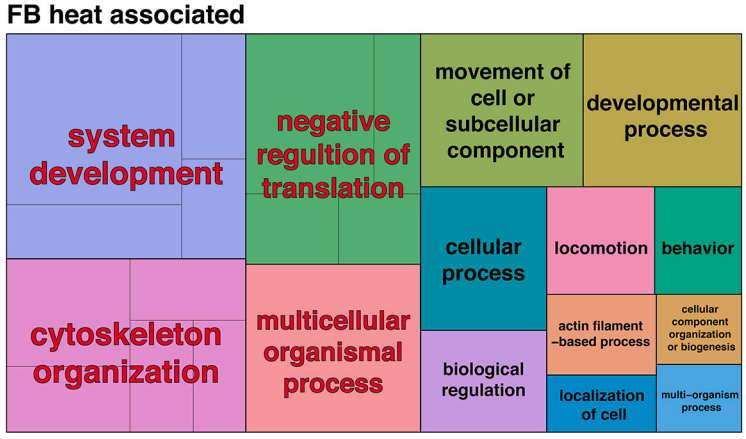


Fig. 4. Weighted Correlation Network Analysis (WCGNA) of heat stress in specific tissues. Top, modules associated with specific tissues and heat stress. Asterisks represent values with a significant positive correlation for based on a regression-based *p*-value for assessing the statistical significance ($P < 0.05$) between a tissue and module [76].Treemaps denote module gene ontology terms associated with heat stress in FB and SG. Complete treemaps are shown in Figs. S6 and S7. Expressional sheets for modules associated with heat stress are shown in Tables S13, S14.

transgenic males as matings with wild type females would only produce viable male offspring [54,55]. Many of the genes required to build these strains such as genes active at cellular blastoderm, proapoptotic and sex determination genes have been previously identified in *L. sericata* [56,57]. For monitoring mating success, it can be beneficial if the released males carry fluorescent protein genes that are expressed in sperm. The male-biased gene set (Table S2) includes genes that in *Drosophila* are mostly expressed in testes such as *beta 2-tubulin* (LOC119611024), which can provide gene promoters for driving fluorescent protein marker genes in developing sperm [58,59].

Cas9-based homing gene drives targeting genes required for female development or fertility could potentially be very efficient for population suppression [54,60]. To make these strains requires identification of genes required for female development and fertility, promoters from genes active in the germline for driving Cas9 expression and U6 snRNA gene promoters for expression of single guide RNAs. The female and male-biased gene sets identified in our analyses of RNA-seq data (Fig. S1, Tables S1 and S2) are a useful resource for such genes. For example, the female-biased gene set (Table S1) contains several maternally expressed genes that are important for fertility such as *zero population growth* (*zpg*) or embryonic development in *Drosophila* including *vasa*, *bicoid*, *nanos*, and *bicaudal* (Table S1). Promoters from the *vasa* and *nanos* genes have

been used to express Cas9 in the germline in mosquitoes [61–63]. Further the *bicoid* gene promoter was used to express a maternal toxin in a synthetic *Medea* toxin-antidote gene drive system [64]. The organization of the *L. sericata vasa* gene (LOC119612105) is very similar to that in *Cochliomyia hominivorax* with two transcripts (XM_037967836.1 and XM_037967837.1) that differ only in the first upstream exon that is noncoding. The transcripts encode identical proteins (XP_037823764.1 and XP_037823765.1). As in *C. hominivorax*, the *nanos* gene (LOC119605916) is relatively compact (less than 2.5 kb) with 4 exons and is very closely linked to the ortholog of the *D. melanogaster CG11779* gene (LOC119605921). The *L. sericata zpg* gene (LOC119604767, annotated as innexin 5) is also relatively compact (less than 2 kb) with 3 exons. As in *C. hominivorax*, the *zpg* gene is very closely linked to the *nudel* gene (LOC119604763) but, unlike in *C. hominivorax*, the genes are transcribed in the same direction. The simple organization of the *L. sericata nanos* and *zpg* genes should facilitate isolation of promoters active in the germline. The *nudel* gene is also of interest as it is essential for female fertility and was the target of a population suppression drive in *Anopheles* mosquitoes [65]. There are 11 annotated *U6* spliceosomal RNA genes in the *L. sericata* genome in five contigs (Table S15). There are three clusters with two or more *U6* genes with one cluster containing four *U6* genes.

3. Conclusions

The whole genome sequence of *L. sericata* reported in this study will provide a reference for future studies on genes that may play important roles in MDT and in parasitism of sheep. Further the genome should facilitate use of gene expression for estimation of PMI [18]. The RNA-seq analyses of larval tissues identified many genes that are expressed in salivary glands or Malphigian tubules that encode potentially secreted proteins that could contribute to maggot ES activity and flystrike. Promoters from these genes could be used to express wound healing factors for enhanced MDT. The identification of germline and U6 genes could provide a source of gene promoters for construction of homing gene drive strains. The identification of *hsp* genes that are strongly activated by heat shock should facilitate the isolation of heat inducible promoters for building temperature-regulated genetic systems for biotechnology enhanced MDT [45] or sterile insect technique [66].

4. Methods

4.1. Lucilia sericata rearing

The MDLA wild type strain of *L. sericata* was reared as described previously [51,57]. In brief, larvae were raised on 93% ground beef at room temperature (approximately 23 °C). Adults were kept in cages and fed water, sugar and given a protein-rich wafer. To obtain the highly inbred line, 21 crosses were performed between single males and single virgin females. Eight of these initial crosses were fertile. From the offspring of each cross, a single male and single virgin female were randomly selected and crossed. In the initial generations up to four single pair crosses were set for each line. If more than one cross was fertile, the progeny from a single cross was selected at random as a source of flies for the next generation. After five generations of inbreeding three lines remained and the number of single pair crosses per line was increased to ten. After seven generations of inbreeding, two lines remained and the number of single pair crosses per line was increased to fifteen. One line, Ls3, remained after 10 generations of single pair full sibling crosses and was the source of genomic DNA for the PacBio library preparation. Attempts to continue inbreeding beyond 10 generations were unsuccessful. The expectation was that the inbred line, Ls3–8 SP2, was approximately 85–90% homozygous across the genome [67]. Approximately 200–250 developing late-stage embryos (14-18 h) were collected from clutches laid by multiple females for genomic DNA isolation. To obtain tissues, early-stage 3rd instar larvae from the highly inbred line were removed from diet, transferred to Ringer's solution and dissected with fine forceps. For heat shock treatment, early-stage 3rd instar larvae were transferred to containers and placed in an air incubator at 37 °C for one hour. Larvae were then immediately dissected, tissues collected in tubes then quick frozen in liquid nitrogen. Four to seven independent samples were collected for each tissue.

4.2. Nucleic acid isolation and sequencing

High molecular weight DNA was isolated from mixed sex late-stage embryos of the inbred strain Ls3–8 SP2 using procedures described previously [68]. The size of the DNA was estimated by using agarose gel electrophoresis with a pippin pulse power supply (Sage Science) following conditions recommended by the manufacturer. DNA that was greater >200 kb in size was used for preparation of 20 kb DNA libraries following instructions from the manufacturer (Pacific Biosciences). Eighteen SMRT Cells were run from one of the 20 kb libraries at the NC State University Genomic Sciences Laboratory.

Total RNA was extracted from frozen samples using Trizol as described previously [68]. Purification of mRNA and cDNA library preparation for Illumina DNA sequencing was as previously described [68]. The final quantified libraries were pooled in equimolar amounts for clustering and sequencing on an Illumina NextSeq 500 DNA sequencer. A total of 781,664,668 paired-end 150 bp reads were produced for the early stage 3rd instar larval tissue samples.

4.3. Genome assembly and gene prediction

FALCON (v. 0.3.0) was used to assemble the PacBio data [23] with assembly parameters including a genome_size = 620 Mb, seed_coverage = 30× and a length_cutoff_pr = 5 kb. Fifteen of the 18 SMRT Cells were used in the assembly, where the 3 SMRT Cells with the smallest average subread length were not used. The assembly, including primary and associate contigs, was polished with the subreads from the 15 SMRT Cells using Arrow from the GenomicConsensus package (v2.2.1) from SMRT Link v5.0.1 (https://github.com/PacificBiosciences/GenomicConsensus). Low quality contigs were removed from the polished assembly using the *trimLowercaseContigs.py* script from the FALCONAssemblyTools (https://github.com/PacificBiosciences/apps-scripts/tree/master/FALCONAssemblyTools). A BLAST search against the nucleotide database was used to determine if any of the contigs were contaminants. Contaminants were identified as being Algae, Archaea, Bacteria, Fungi (Ascomycetes), Slime Molds, *Homo sapiens*, Vectors and Viruses. The reduction step of the Redundans pipeline (v0.14a) was run with the identity parameter = 0.90 and the overlap parameter = 0.80 (default value) on the primary contigs (27). Four additional contigs were removed for hits to Bos in addition to a partial Mitochondrial contig. For the final Mitochondrial contig, only the first 15,961 bp were saved as an inverted duplication existed in the second half of this contig. The final assembly stats were evaluated using QUAST [69] and shown in Table 1. Annotation was done by the NCBI Eukaryotic Genome Annotation, an automated pipeline that annotates genes, transcripts and proteins on finished genome assemblies. RNA-seq reads from this study and earlier studies [32,33] (and accession PRJNA417347) were used for gene prediction. Proteins that contained predicted signal peptides for secretion were identified using SignalP [34].

4.4. Gene expression analyses

Trimming of adapter and for quality was done using Trimmomatic version 0.23 [70] with a sliding window quality cut-off of 15 and a minimum length of 36 required to keep a read. FastQC was utilized to identify quality of each sample set. RNA-seq analyses were conducted according to Attardo et al. [71] and Scott et al. [28]. Two pipelines were used. First, reads were mapped to the predicted genes with the use of CLC Genomics (Qiagen), allowed two mismatches with over 90%

similarity. Expression values were converted to transcript per million mapped (TPM). EdgeR [72] was utilized for statistical comparison under standard setting with a false detection rate (FDR) at 0.05. Second, reads were mapped to the predicted transcripts from the *L. sericata* genome (GCA_015586225.1) with Sailfish [73] under recommended setting and this was followed by DEseq2 [36] following an FDR at 0.05. Both methods were compared for validation and expression levels had a high Pearson correlation coefficient (0.945) between each pipeline for analysis. As such further analyses were based upon the Sailfish/DESeq2 analysis. For transcriptional analyses of the secreted peptides, 3604 sequences were reduced to 3033 to remove splice variants that are nearly identical and cannot be separated by nucleotide sequence.

Genes with over two-fold enrichment or reduction were used to establish pairwise comparisons. Genes that were enriched in a single stage compared to other sex or developmental stage were deemed stage- or sex-specific. Tissue specific analyses were conducted similarly, which was followed by analyses to establish tissue-specific gene sets. This was then followed by assessing heat-induced transcriptional changes. GO analyses were conducted by examining the *D. melanogaster* orthologs with g:Profiler [74] and visualization of the GO results through REVIGO [75].

Along with our comparative analysis between each developmental stage and tissue type, a gene co-expression network was created utilizing the WGCNA [76] R software package (https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/). This approach was used to correlate genes that are similarly expressed across developmental stages sexes, tissues and in relation to thermal stress, then place the genes into modules, and relate the resulting modules to the samples. To prepare our expression data for WGCNA, genes of zero variance were removed (e.g. expression values of 0 for all sets) from the dataset. A soft thresholding power of 10 was chosen based on the scale-free topology fit index curve made prior to network construction. The minimum module size for this network was set to 12. To indicate identity of each module to the developmental stages and tissues, the developmental stages, sexes, tissues, and heat treatments were used as input traits during network construction. The modules exhibiting the greatest significance to the trait data (<0.05) were further analyzed to determine function and relationship to the developmental stages, sexes, tissues, and heat stress condition relative to GO analyses were conducted as before by examining the *D. melanogaster* orthologs with g:Profiler [74].

4.5. Statistics and reproducibility

For the gene expression analyses, Malpighian tubules (4 samples), fat body (5 samples) and salivary glands (7 samples) were collected from the control larvae. From the heat shocked larvae we collected 5 samples of Malpighian tubules, 6 samples of fat body and 7 samples of salivary glands. JMP15 was used for comparison of means for individual transcript TPM values from heat shocked and control samples. The test statistics for DESeq2 and WGCNA are noted in the text. The datasets have been made public on the NCBI Sequence Read Archive under BioProject PRJNA602506.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.10.003.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The assembly has been submitted to NCBI, biosample SAMN13896454, Bioproject, PRJNA680594, assembly GCA_015586225.1.

The PacBio sequencing reads were deposited in the Sequence Read Archive (SRA) under accession numbers SRR10951024, SRR10951025, SRR10951026, SRR10951027, SRR10951028, SRR10951029, SRR10951030, SRR10951031, SRR10951032, SRR10951033, SRR10951034, SRR10951035, SRR10951036, SRR10951037, SRR10951038.

The RNA-seq reads included 3 samples from adult females (SRR11817803, SRR11817804, SRR11817805) and 3 samples from adult males (SRR11817806, SRR11817807, SRR11817808). In addition, there are RNA-seq samples that were isolated from tissues that were dissected from early feeding third instar larvae that included samples from salivary glands (7 control samples: SRR11814661, SRR11814662, SRR14882999, SRR14883000, SRR14883001, SRR14883003, SRR14883004; 7 heat shock samples: SRR11814663, SRR11814664, SRR11814666, SR118 14667, SRR11814672, SRR11814683, SRR11814684), fat bodies (5 control samples: SRR11814676, SRR14882995, SRR14882996, SRR14882997, SRR14882998; 6 heat shock samples: SRR11814677, SRR11814678, SRR11814679, SRR11814680, SRR11814681, SRR1181 4682) and Malpighian tubules (4 control samples: SRR11814668, SRR11814669, SRR14882994, SRR14883002; 5 heat shock samples: SRR11814670, SRR11814671, SRR11814673, SRR11814674, SRR1814675).

The RNA-seq larvae data included in Fig. S1, Table S4 came from SRA: ERS569670, SRS254714, SRS254713.

The datasets supporting the conclusions of this article are included within the article and its additional files.

Funding

This work was supported by the Defense Advanced Research Projects Agency (cooperative agreement number D17AC00007). The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

CRediT authorship contribution statement

Rebecca Davis: Methodology, Investigation, Project administration, Writing - Review & Editing Formal analysis, Data Curation. *Esther Belikoff:* Methodology, Investigation, Formal analysis, Data Curation, Writing - Review & Editing. *Allison Dickey:* Software, Formal analysis, Data Curation, Writing - Review & Editing. *Elizabeth Scholl:* Software, Formal analysis, Data Curation, Writing - Review & Editing. *Joshua Benoit:* Formal analysis, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. *Maxwell Scott:* Conceptualization, Formal analysis, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration, Funding acquisition.

Authors' contributions

RJD isolated high molecular weight DNA and RNA from different stages and dissected tissues. EJB performed crosses to make the highly inbred line, performed timed larval and adult collections and dissected larval tissues. Differential RNA expression analyses, gene predictions and BUSCO analyses were performed by EHS. JBB performed the comparative RNA-seq studies and WGCNA. Assembly of the genome from PacBio reads was performed by AD. MJS designed research, analyzed data, wrote the paper and obtained funding for this project. All authors read and approved the final manuscript.

Declaration of Competing Interest

The authors declare they have no competing interests.

Acknowledgements

We are grateful for technical assistance from Mary Hester and thank Samuel T. Bailey for help running the initial part of the WGCNA analysis.

References

[1] J.H. Byrd, J.L. Castner, Forensic Entomology: The Utility of Arthropods in Legal Investigations, Taylor & Francis, Second Edition, 2009.
[2] D.B. Rivers, G.A. Dahlem, The Science of Forensic Entomology, John Wiley & Sons, Oxford, U.K., 2014.
[3] G.S. Anderson, Comparison of decomposition rates and faunal colonization of carrion in indoor and outdoor environments, J. Forensic Sci. 56 (2011) 136–142.
[4] A. Cruise, D.W. Watson, C. Schal, Ecological succession of adult necrophilous insects on neonate *Sus scrofa domesticus* in central North Carolina, PLoS One 13 (2018), e0195785.
[5] M.J. Hall, R.L. Wall, J.R. Stevens, Traumatic myiasis: a neglected disease in a changing world, Annu. Rev. Entomol. 61 (2016) 159–176.
[6] R.M. Sandeman, G.W. Levot, A.C. Heath, P.J. James, J.C. Greeff, M.J. Scott, P. Batterham, V.M. Bowles, Control of the sheep blowfly in Australia and New Zealand—are we there yet? Int. J. Parasitol. 44 (2014) 879–891.
[7] A.C. Heath, D.M. Bishop, Flystrike in New Zealand: an overview based on a 16-year study, following the introduction and dispersal of the Australian sheep blowfly, *Lucilia cuprina* Wiedemann (Diptera: Calliphoridae), Vet. Parasitol. 137 (2006) 333–344.
[8] R.A. Sherman, M.J. Hall, S. Thomas, Medicinal maggots: an ancient remedy for some contemporary afflictions, Annu. Rev. Entomol. 45 (2000) 55–81.
[9] D. Bazalinski, M. Kozka, M. Karnas, P. Wiech, Effectiveness of chronic wound debridement with the use of larvae of *Lucilia sericata*, J. Clin. Med. 8 (2019).
[10] L. Chambers, S. Woodrow, A.P. Brown, P.D. Harris, D. Phillips, M. Hall, J. C. Church, D.I. Pritchard, Degradation of extracellular matrix components by defined proteinases from the greenbottle larva *Lucilia sericata* used for the clinical debridement of non-healing wounds, Br. J. Dermatol. 148 (2003) 14–23.
[11] Y. Nigam, A. Bexfield, S. Thomas, N.A. Ratcliffe, Maggot therapy: the science and implication for CAM part II-maggots combat infection, Evid. Based Complement. Alternat. Med. 3 (2006) 303–308.
[12] V. Cerovsky, J. Slaninova, V. Fucik, L. Monincova, L. Bednarova, P. Malon, J. Stokrova, Lucifensin, a novel insect defensin of medicinal maggots: synthesis and structural study, ChemBioChem 12 (2011) 1352–1361.
[13] A.K. Poppel, A. Koch, K.H. Kogel, H. Vogel, C. Kollewe, J. Wiesner, A. Vilcinskas, Lucimycin, an antifungal peptide from the therapeutic maggot of the common green bottle fly *Lucilia sericata*, Biol. Chem. 395 (2014) 649–656.
[14] A.W. Brown, The nitrogen metabolism of an insect (*Lucilia sericata* Mg.): uric acid, allantoin and uricase, Biochem. J. 32 (1938) 895–902.
[15] H.W. Margraf, T.H. Covey Jr., A trial of silver-zinc-allantoinate in the treatment of leg ulcers, Arch. Surg. 112 (1977) 699–704.
[16] A. Baumann, M. Skaljac, R. Lehmann, A. Vilcinskas, Z. Franta, Urate oxidase produced by *Lucilia sericata* medical maggots is localized in Malpighian tubes and facilitates allantoin production, Insect Biochem. Mol. Biol. 83 (2017) 44–53.
[17] K. Honda, K. Okamoto, Y. Mochida, K. Ishioka, M. Oka, K. Maesato, R. Ikee, H. Moriya, S. Hidaka, T. Ohtake, K. Doi, T. Fujita, S. Kobayashi, E. Noiri, A novel mechanism in maggot debridement therapy: protease in excretion/secretion promotes hepatocyte growth factor production, Am. J. Physiol. Cell Physiol. 301 (2011) C1423–C1430.
[18] A.M. Tarone, D.R. Foran, Gene expression during blow fly development: improving the precision of age estimates in forensic entomology, J. Forensic Sci. 56 (Suppl. 1) (2011) S112–S122.
[19] A.M. Tarone, C.J. Picard, C. Spiegelman, D.R. Foran, Population and temperature effects on Lucilia sericata (Diptera: Calliphoridae) body size and minimum development time, J. Med. Entomol. 48 (2011) 1062–1068.
[20] L. Cervantes, L. Dourel, E. Gaudry, T. Pasquerault, B. Vincent, Effect of low temperature in the development cycle of *Lucilia sericata* (Meigen) (Diptera, Calliphoridae): implications for the minimum postmortem interval estimation, Forensic Sci. Res. 3 (2018) 52–59.
[21] M. Wang, Y. Wang, G. Hu, Y. Wang, W. Xu, M. Wu, J. Wang, Development of *Lucilia sericata* (Diptera: Calliphoridae) under constant temperatures and its significance for the estimation of time of death, J. Med. Entomol. 57 (2020) 1373–1381.
[22] L. Rabinow, A. Barrera, J. Hyde, A simple method for isolation of very high molecular weight DNA from *Drosophila* embryos, Nucleic Acids Res. 21 (1993) 4985–4986.
[23] C.S. Chin, P. Peluso, F.J. Sedlazeck, M. Nattestad, G.T. Concepcion, A. Clum, C. Dunn, R. O'Malley, R. Figueroa-Balderas, A. Morales-Cruz, G.R. Cramer, M. Delledonne, C. Luo, J.R. Ecker, D. Cantu, D.R. Rank, M.C. Schatz, Phased diploid genome assembly with single-molecule real-time sequencing, Nat. Methods 13 (2016) 1050–1054.
[24] C.S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E.E. Eichler, S.W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data, Nat. Methods 10 (2013) 563–569.
[25] L.P. Pryszcz, T. Gabaldon, Redundans: an assembly pipeline for highly heterozygous genomes, Nucleic Acids Res. 44 (2016), e113.
[26] C.J. Picard, J.S. Johnston, A.M. Tarone, Genome sizes of forensically relevant Diptera, J. Med. Entomol. 49 (2012) 192–197.
[27] M. Seppey, M. Manni, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness, Methods Mol. Biol. 1962 (2019) 227–245.
[28] M.J. Scott, J.B. Benoit, R.J. Davis, S.T. Bailey, V. Varga, E.O. Martinson, P. V. Hickner, Z. Syed, G.A. Cardoso, T.T. Torres, M.T. Weirauch, E.H. Scholl, A. M. Phillippy, A. Sagel, M. Vasquez, G. Quintero, S.R. Skoda, Genomic analyses of a livestock pest, the New World screwworm, find potential targets for genetic control programs, Commun. Biol. 3 (2020) 424.
[29] P.U. Olafson, S. Aksoy, G.M. Attardo, G. Buckmeier, X. Chen, C.J. Coates, M. Davis, J. Dykema, S.J. Emrich, M. Friedrich, C.J. Holmes, P. Ioannidis, E.N. Jansen, E. C. Jennings, D. Lawson, E.O. Martinson, G.L. Maslen, R.P. Meisel, T.D. Murphy, D. Nayduch, D.R. Nelson, K.J. Oyen, T.J. Raszick, J.M.C. Ribeiro, H.M. Robertson, A.J. Rosendale, T.B. Sackton, P. Saelao, S.L. Swiger, S.H. Sze, A.M. Tarone, D. B. Taylor, W.C. Warren, R.M. Waterhouse, M.T. Weirauch, J.H. Werren, R. K. Wilson, E.M. Zdobnov, J.B. Benoit, The genome of the stable fly, *Stomoxys calcitrans*, reveals potential mechanisms underlying reproduction, host interactions, and novel targets for pest control, BMC Biol. 19 (2021) 41.
[30] S.B. Sim, S.M. Geib, A chromosome-scale assembly of the *Bactrocera cucurbitae* genome provides insight to the genetic basis of white pupae, G3 (Bethesda) 7 (2017) 1927–1940.
[31] C.A. Anstead, P.K. Korhonen, N.D. Young, R.S. Hall, A.R. Jex, S.C. Murali, D. S. Hughes, S.F. Lee, T. Perry, A.J. Stroehlein, B.R. Ansell, B. Breugelmans, A. Hofmann, J. Qu, S. Dugan, S.L. Lee, H. Chao, H. Dinh, Y. Han, H. V. Doddapaneni, K.C. Worley, D.M. Muzny, P. Ioannidis, R.M. Waterhouse, E. M. Zdobnov, P.J. James, N.H. Bagnall, A.C. Kotze, R.A. Gibbs, S. Richards, P. Batterham, R.B. Gasser, *Lucilia cuprina* genome unlocks parasitic fly biology to underpin future interventions, Nat. Commun. 6 (2015) 7344.
[32] S.H. Sze, J.P. Dunham, B. Carey, P.L. Chang, F. Li, R.M. Edman, C. Fjeldsted, M. J. Scott, S.V. Nuzhdin, A.M. Tarone, A *de novo* transcriptome assembly of *Lucilia sericata* (Diptera: Calliphoridae) with predicted alternative splices, single nucleotide polymorphisms and transcript expression estimates, Insect Mol. Biol. 21 (2012) 205–221.
[33] A.K. Poppel, H. Vogel, J. Wiesner, A. Vilcinskas, Antimicrobial peptides expressed in medicinal maggots of the blow fly *Lucilia sericata* show combinatorial activity against bacteria, Antimicrob. Agents Chemother. 59 (2015) 2508–2514.
[34] J.J. Almagro Armenteros, K.D. Tsirigos, C.K. Sonderby, T.N. Petersen, O. Winther, S. Brunak, G. von Heijne, H. Nielsen, SignalP 5.0 improves signal peptide predictions using deep neural networks, Nat. Biotechnol. 37 (2019) 420–423.
[35] B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis, Stat. Appl. Genet. Mol. Biol. 4 (2005), 17.
[36] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.
[37] J. Alves-Silva, J.M. Ribeiro, J. Van Den Abbeele, G. Attardo, Z. Hao, L.R. Haines, M. B. Soares, M. Berriman, S. Aksoy, M.J. Lehane, An insight into the sialome of *Glossina morsitans morsitans*, BMC Genomics 11 (2010) 213.
[38] E.L. Telleria, J.B. Benoit, X. Zhao, A.F. Savage, S. Regmi, T.L. Alves e Silva, M. O'Neill, S. Aksoy, Insights into the trypanosome-host interactions revealed through transcriptomic analysis of parasitized tsetse fly salivary glands, PLoS Negl. Trop. Dis. 8 (2014) e2649.
[39] Z. Franta, H. Vogel, R. Lehmann, O. Rupp, A. Goesmann, A. Vilcinskas, Next generation sequencing identifies five major classes of potentially therapeutic enzymes secreted by *Lucilia sericata* medical maggots, Biomed. Res. Int. 2016 (2016) 8285428.
[40] V.R. Chintapalli, J. Wang, P. Herzyk, S.A. Davies, J.A. Dow, Data-mining the FlyAtlas online resource to identify core functional motifs across transporting epithelia, BMC Genomics 14 (2013) 518.
[41] S.W. Robinson, P. Herzyk, J.A. Dow, D.P. Leader, FlyAtlas: database of gene expression in the tissues of *Drosophila melanogaster*, Nucleic Acids Res. 41 (2013) D744–D750.
[42] X. Xu, F. Oliveira, B.W. Chang, N. Collin, R. Gomes, C. Teixeira, D. Reynoso, V. My Pham, D.E. Elnaiem, S. Kamhawi, J.M. Ribeiro, J.G. Valenzuela, J.F. Andersen, Structure and function of a "yellow" protein from saliva of the sand fly Lutzomyia longipalpis that confers protective immunity against Leishmania major infection, J. Biol. Chem. 286 (2011) 32383–32393.
[43] A.B. Guimaraes-Costa, J.P. Shannon, I. Waclawiak, J. Oliveira, C. Meneses, W. de Castro, X. Wen, J. Brzostowski, T.D. Serafim, J.F. Andersen, H.D. Hickman, S. Kamhawi, J.G. Valenzuela, F. Oliveira, A sand fly salivary protein acts as a neutrophil chemoattractant, Nat. Commun. 12 (2021) 3213.
[44] E. Cohen, J.K. Sawyer, N.G. Peterson, J.A.T. Dow, D.T. Fox, Physiology, development, and disease modeling in the *Drosophila* excretory system, Genetics 214 (2020) 235–264.
[45] R.J. Linger, E.J. Belikoff, Y. Yan, F. Li, H.A. Wantuch, H.L. Fitzsimons, M.J. Scott, Towards next generation maggot debridement therapy: transgenic *Lucilia sericata* larvae that produce and secrete a human growth factor, BMC Biotechnol. 16 (2016) 30.
[46] Y. Yan, Y. Kobayashi, C. Huang, B. Liu, W. Qian, F. Wan, M.F. Schetelig, Highly efficient temperature inducible CRISPR-Cas9 gene targeting in *Drosophila suzukii*, Int. J. Mol. Sci. 22 (2021).
[47] C. Concha, R.M. Edman, E.J. Belikoff, A.H. Schiemann, B. Carey, M.J. Scott, Organization and expression of the Australian sheep blowfly (*Lucilia cuprina*) *hsp23*, *hsp24*, *hsp70* and *hsp83* genes, Insect Mol. Biol. 21 (2012) 169–180.
[48] S. Kumari, T.K. Panda, T. Pradhan, Lysyl oxidase: its diversity in health and diseases, Indian J. Clin. Biochem. 32 (2017) 134–141.
[49] L.U. Abubakar, W.D. Bulimo, F.J. Mulaa, E.O. Osir, Molecular characterization of a tsetse fly midgut proteolytic lectin that mediates differentiation of African trypanosomes, Insect Biochem. Mol. Biol. 36 (2006) 344–352.

[50] G. Wettstein, P.S. Bellaye, O. Micheau, P. Bonniaud, Small heat shock proteins and the cytoskeleton: an essential interplay for cell integrity? Int. J. Biochem. Cell Biol. 44 (2012) 1680–1686.
[51] F. Li, H.A. Wantuch, R.J. Linger, E.J. Belikoff, M.J. Scott, Transgenic sexing system for genetic control of the Australian sheep blow fly *Lucilia cuprina*, Insect Biochem. Mol. Biol. 51 (2014) 80–88.
[52] Y. Yan, M.J. Scott, A transgenic embryonic sexing system for the Australian sheep blow fly *Lucilia cuprina*, Sci. Rep. 5 (2015) 16090.
[53] Y. Yan, M.J. Scott, Building a transgenic sexing strain for genetic control of the Australian sheep blow fly *Lucilia cuprina* using two lethal effectors, BMC Genet. 21 (2020) 141.
[54] L. Alphey, Genetic control of mosquitoes, Annu. Rev. Entomol. 59 (2014) 205–224.
[55] C. Concha, A. Palavesam, F.D. Guerrero, A. Sagel, F. Li, J.A. Osborne, Y. Hernandez, T. Pardo, G. Quintero, M. Vasquez, G.P. Keller, P.L. Phillips, J. B. Welch, W.O. McMillan, S.R. Skoda, M.J. Scott, A transgenic male-only strain of the New World screwworm for an improved control program using the sterile insect technique, BMC Biol. 14 (2016) 72.
[56] R.M. Edman, R.J. Linger, E.J. Belikoff, F. Li, S.H. Sze, A.M. Tarone, M.J. Scott, Functional characterization of calliphorid cell death genes and cellularization gene promoters for controlling gene expression and cell viability in early embryos, Insect Mol. Biol. 24 (2015) 58–70.
[57] F. Li, S.P. Vensko 2nd, E.J. Belikoff, M.J. Scott, Conservation and sex-specific splicing of the *transformer* gene in the Calliphorids *Cochliomyia hominivorax*, *Cochliomyia macellaria* and *Lucilia sericata*, PLoS One 8 (2013), e56303.
[58] F. Scolari, M.F. Schetelig, S. Bertin, A.R. Malacrida, G. Gasperi, E.A. Wimmer, Fluorescent sperm marking to improve the fight against the pest insect *Ceratitis capitata* (Wiedemann; Diptera: Tephritidae), New Biotechnol. 25 (2008) 76–84.
[59] Y. Yan, J. Schwirz, M.F. Schetelig, Characterization of the *Drosophila suzukii beta2-tubulin* gene and the utilization of its promoter to monitor sex separation and insemination, Gene 771 (2021) 145366.
[60] M.J. Scott, F. Gould, M.D. Lorenzen, N. Grubbs, O.R. Edwards, D.A. O'Brochta, Agricultural production: assessment of the potential use of Cas9-mediated gene drive systems for agricultural pest control, J. Resp. Innov. 5 (2018) S98–S120.
[61] V.M. Gantz, N. Jasinskiene, O. Tatarenkova, A. Fazekas, V.M. Macias, E. Bier, A. A. James, Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E6736–E6743.
[62] K. Kyrou, A.M. Hammond, R. Galizi, N. Kranjc, A. Burt, A.K. Beaghton, T. Nolan, A. Crisanti, A CRISPR-Cas9 gene drive targeting *doublesex* causes complete population suppression in caged *Anopheles gambiae* mosquitoes, Nat. Biotechnol. 36 (2018) 1062–1066.
[63] A. Hammond, X. Karlsson, I. Morianou, K. Kyrou, A. Beaghton, M. Gribble, N. Kranjc, R. Galizi, A. Burt, A. Crisanti, T. Nolan, Regulating the expression of gene drives is key to increasing their invasive potential and the mitigation of resistance, PLoS Genet. 17 (2021), e1009321.
[64] C.H. Chen, H. Huang, C.M. Ward, J.T. Su, L.V. Schaeffer, M. Guo, B.A. Hay, A synthetic maternal-effect selfish genetic element drives population replacement in *Drosophila*, Science 316 (2007) 597–600.
[65] A. Hammond, R. Galizi, K. Kyrou, A. Simoni, C. Siniscalchi, D. Katsanos, M. Gribble, D. Baker, E. Marois, S. Russell, A. Burt, N. Windbichler, A. Crisanti, T. Nolan, A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*, Nat. Biotechnol. 34 (2016) 78–83.
[66] C. Concha, Y. Yan, A. Arp, E. Quilarque, A. Sagel, A.P. de Leon, W.O. McMillan, S. Skoda, M.J. Scott, An early female lethal system of the New World screwworm, *Cochliomyia hominivorax*, for biotechnology-enhanced SIT, BMC Genet. 21 (2020) 143.
[67] S. Wright, Systems of Mating. II. The effects of inbreeding on the genetic composition of a population, Genetics 6 (1921) 124–143.
[68] R.J. Linger, E.J. Belikoff, M.J. Scott, Dosage compensation of X-linked Muller element F genes but not X-linked transgenes in the Australian sheep blowfly, PLoS One 10 (2015), e0141544.
[69] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, Bioinformatics 29 (2013) 1072–1075.
[70] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120.
[71] G.M. Attardo, A.M.M. Abd-Alla, A. Acosta-Serrano, J.E. Allen, R. Bateta, J. B. Benoit, K. Bourtzis, J. Caers, G. Caljon, M.B. Christensen, D.W. Farrow, M. Friedrich, A. Hua-Van, E.C. Jennings, D.M. Larkin, D. Lawson, M.J. Lehane, V. P. Lenis, E. Lowy-Gallego, R.W. Macharia, A.R. Malacrida, H.G. Marco, D. Masiga, G.L. Maslen, I. Matetovici, R.P. Meisel, I. Meki, V. Michalkova, W.J. Miller, P. Minx, P.O. Mireji, L. Ometto, A.G. Parker, R. Rio, C. Rose, A.J. Rosendale, O. Rota-Stabelli, G. Savini, L. Schoofs, F. Scolari, M.T. Swain, P. Takac, C. Tomlinson, G. Tsiamis, J. Van Den Abbeele, A. Vigneron, J. Wang, W.C. Warren, R.M. Waterhouse, M.T. Weirauch, B.L. Weiss, R.K. Wilson, X. Zhao, S. Aksoy, Comparative genomic analysis of six Glossina genomes, vectors of African trypanosomes, Genome Biol. 20 (2019) 187.
[72] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[73] R. Patro, S.M. Mount, C. Kingsford, Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms, Nat. Biotechnol. 32 (2014) 462–464.
[74] U. Raudvere, L. Kolberg, I. Kuzmin, T. Arak, P. Adler, H. Peterson, J. Vilo, g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update), Nucleic Acids Res. 47 (2019) W191–W198.
[75] F. Supek, M. Bosnjak, N. Skunca, T. Smuc, REVIGO summarizes and visualizes long lists of gene ontology terms, PLoS One 6 (2011), e21800.
[76] P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis, BMC Bioinform. 9 (2008) 559.