Strongyloides RNA-seq Browser: a web-based software platform for on-demand bioinformatics analyses of *Strongyloides* species

Astra S. Bryant (ID),[1] Stephanie F. DeMarco (ID),[1] and Elissa A. Hallem (ID) [1,2,*]

[1]Department of Microbiology, Immunology, and Molecular Genetics, University of California, Los Angeles, Los Angeles, CA 90095, USA
[2]Molecular Biology Institute, University of California, Los Angeles, Los Angeles, CA 90095, USA

*Corresponding author: UCLA MIMG, 237 BSRB, 615 Charles E. Young Dr. S., Los Angeles, CA 90095, USA. ehallem@ucla.edu

Abstract

Soil-transmitted gastrointestinal parasitic nematodes infect approximately 1 billion people worldwide, predominantly in low-resource communities. Skin-penetrating gastrointestinal nematodes in the genus *Strongyloides* are emerging as model systems for mechanistic studies of soil-transmitted helminths due to the growing availability of functional genomics tools for these species. To facilitate future genomics studies of *Strongyloides* species, we have designed a web-based application, the *Strongyloides* RNA-seq Browser, that provides an open source, user-friendly portal for accessing and analyzing *Strongyloides* genomic expression data. Specifically, the *Strongyloides* RNA-seq Browser takes advantage of alignment-free read mapping tools and R-based transcriptomics tools to re-analyze publicly available RNA sequencing datasets from four *Strongyloides* species: *Strongyloides stercoralis*, *Strongyloides ratti*, *Strongyloides papillosus*, and *Strongyloides venezuelensis*. This application permits on-demand exploration and quantification of gene expression across life stages without requiring previous coding experience. Here, we describe this interactive application and demonstrate how it may be used by nematode researchers to conduct a standard set of bioinformatics queries.

Keywords: *Strongyloides*; parasitic nematodes; differential gene expression; functional enrichment

Introduction

Soil-transmitted gastrointestinal parasitic nematodes, including those in the genus *Strongyloides*, are a major source of disease and economic burden. *Strongyloides* species infect a range of human and animal hosts; the human parasite *Strongyloides stercoralis* is estimated to infect approximately 610 million people worldwide (Buonfrate *et al.* 2020). *Strongyloides* species have complex life cycles (Figure 1A) that include host-dwelling adults as well as multiple soil-dwelling life stages, including the developmentally arrested infective third larval stage (iL3) that actively searches for hosts to infect (Roberts and Janovy 2005; Schafer and Skopic 2006).

Ongoing advances in functional genomics techniques such as transgenesis and CRISPR/Cas9-mediated mutagenesis are positioning *Strongyloides* species as genetically tractable model systems for gastrointestinal parasitic nematodes (Gang *et al.* 2017; Lok *et al.* 2017; Castelletto *et al.* 2020). Technical advances in *Strongyloides* functional genomics have been greatly facilitated by the fully sequenced genomes of four *Strongyloides* species: *S. stercoralis*, the rodent parasites *Strongyloides ratti* and *Strongyloides venezuelensis*, and the ruminant parasite *Strongyloides papillosus* (Hunt *et al.* 2016; Howe *et al.* 2017). High-quality reference genomes permit researchers to identify homologous genes in parasitic and free-living nematodes, while transgenesis and mutagenesis protocols allow researchers to test the functional contributions of these genes to parasitic behaviors (Gang *et al.* 2017, 2020; Bryant *et al.* 2018).

In *Strongyloides* species with fully sequenced genomes, bulk RNA sequencing datasets are also publicly available, and their analysis has provided insight into the genetic and evolutionary basis of parasitism (Stoltzfus *et al.* 2012; Hunt *et al.* 2016, 2018). However, published results generally focus on only a subset of the life stages for which RNA-seq data is available (Stoltzfus *et al.* 2012; Hunt *et al.* 2016, 2018) or are not easily adapted for quantitative comparisons (Howe *et al.* 2017), although recent updates to WormBase ParaSite (circa version WBPS15) have introduced the ability to download RNA-seq data and analyses for some species. The field has lacked a user-friendly portal for accessing and visualizing quantitative *Strongyloides* gene expression data and performing custom bioinformatics queries across life stages. Here, we present a *Strongyloides* RNA-seq Browser that permits exploration of RNA expression levels in four *Strongyloides* species and features a streamlined user interface for on-demand differential gene expression (DGE) and functional enrichment analyses. We hope that this broadly accessible tool will support future studies of the genetic basis of parasitism in *Strongyloides*.

A

B

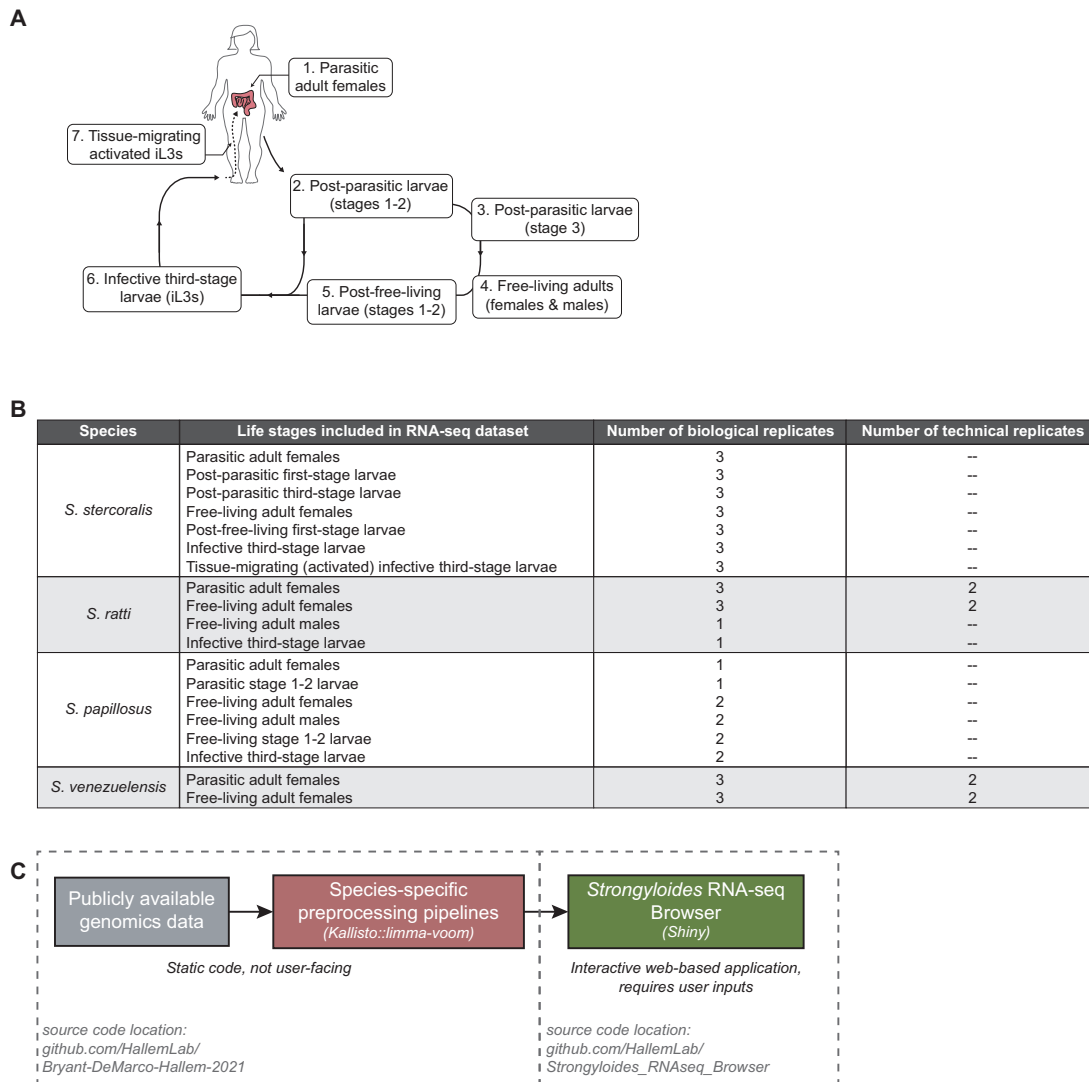| Species	Life stages included in RNA-seq dataset	Number of biological replicates	Number of technical replicates
S. stercoralis	Parasitic adult females	3	--
	Post-parasitic first-stage larvae	3	--
	Post-parasitic third-stage larvae	3	--
	Free-living adult females	3	--
	Post-free-living first-stage larvae	3	--
	Infective third-stage larvae	3	--
	Tissue-migrating (activated) infective third-stage larvae	3	--
S. ratti	Parasitic adult females	3	2
	Free-living adult females	3	2
	Free-living adult males	1	--
	Infective third-stage larvae	1	--
S. papillosus	Parasitic adult females	1	--
	Parasitic stage 1-2 larvae	1	--
	Free-living adult females	2	--
	Free-living adult males	2	--
	Free-living stage 1-2 larvae	2	--
	Infective third-stage larvae	2	--
S. venezuelensis	Parasitic adult females	3	2
	Free-living adult females	3	2

C

Figure 1 Life cycle of *Strongyloides* species. (A) Schematic of the developmental life cycle of *Strongyloides* species, including: stages that reside within host animals (parasitic adult females and tissue-migrating, activated iL3s); free-living adults and larvae that primarily reside on feces excreted from host animals (post-parasitic larvae, free-living adult males and females, and post-free-living larvae); and developmentally arrested third-stage larvae (iL3s) that migrate away from host feces to actively search the environment for host animals. (B) Life stages that are included in the *Strongyloides* RNA-seq Browser, by species. The number of biological and technical replicates are listed for each life stage, when available. For *S. stercoralis*, *S. ratti*, and *S. papillosus*, all publicly available samples are included in the app. For *S. venezuelensis*, only data from free-living adult females and parasitic adult females are included. Analyses and discussion related to this decision are available at https://github.com/HallemLab/Bryant-DeMarco-Hallem-2021. (C) Diagram of data flow, from publicly available genomics data (including raw RNA-seq read files and reference transcriptomes) to a Kallisto::limma-voom preprocessing pipeline that generates static RNA-seq expression files, which are accessed and analyzed via an interactive web-based Shiny application, the *Strongyloides* RNA-seq Browser.

Materials and methods
Application overview
The *Strongyloides* RNA-seq Browser (Supplementary File S1) enables users to browse publicly available *Strongyloides* bulk RNA-seq datasets and perform on-demand analyses, including differential expression and gene set enrichment.

Data content
The *Strongyloides* RNA-seq Browser contains bulk RNA sequencing data from four *Strongyloides* species: *S. stercoralis*, *S. ratti*, *S. papillosus*, and *S. venezuelensis*. These four species can be divided into two evolutionarily divergent subclades featuring pairs of more closely related species: *S. stercoralis—S. ratti* and *S. papillosus—S. venezuelensis* (Hunt *et al.* 2016, 2018). For each species, the number and type of

developmental life stages included in the application varies based on available data (Figure 1B, Supplementary File S2).

Data workflow
We first reanalyzed publicly available raw sequencing data, applying a Kallisto::limma-voom preprocessing pipeline to each species in parallel (Figure 1C). Key pipeline elements include: a shell script for quality control assessments and Kallisto-based alignment-free read mapping; and an R Markdown (.rmd) file for conversion of abundance data to $\log_2$CPM, followed by filtering, normalization, gene annotation, and finally, saving as static *Strongyloides* Browser input files (Supplementary Figure S1).

Next, we developed an interactive web application for browsing and on-demand analysis of the data files generated by our preprocessing pipeline (Figure 1C). Key browser features include

extracting gene expression values for user-specified genes of interest as well as known homologs in other *Strongyloides* species, and performing genome-wide limma-voom-based pairwise DGE analyses followed by gene set enrichment analyses (Figure 2, Supplementary Figure S2). Interactive app functionality was implemented using the Shiny R package v1.5.0.

Below, we described methodological details related to the preprocessing pipeline, app features, and benchmarking the results of pipeline processing and in-app data analyses.

Preprocessing pipeline
Data source
Raw reads and study design files (Supplementary File S2) were downloaded from the European Nucleotide Archive using the following study accession numbers: PRJEB3116 (*S. stercoralis*); PRJEB1376 and PRJEB3187 (*S. ratti*); PRJEB14543 (*S. papillosus*); and PRJDB3457 (*S. venezuelensis*). The following reference transcriptomes were downloaded from WormBase ParaSite:
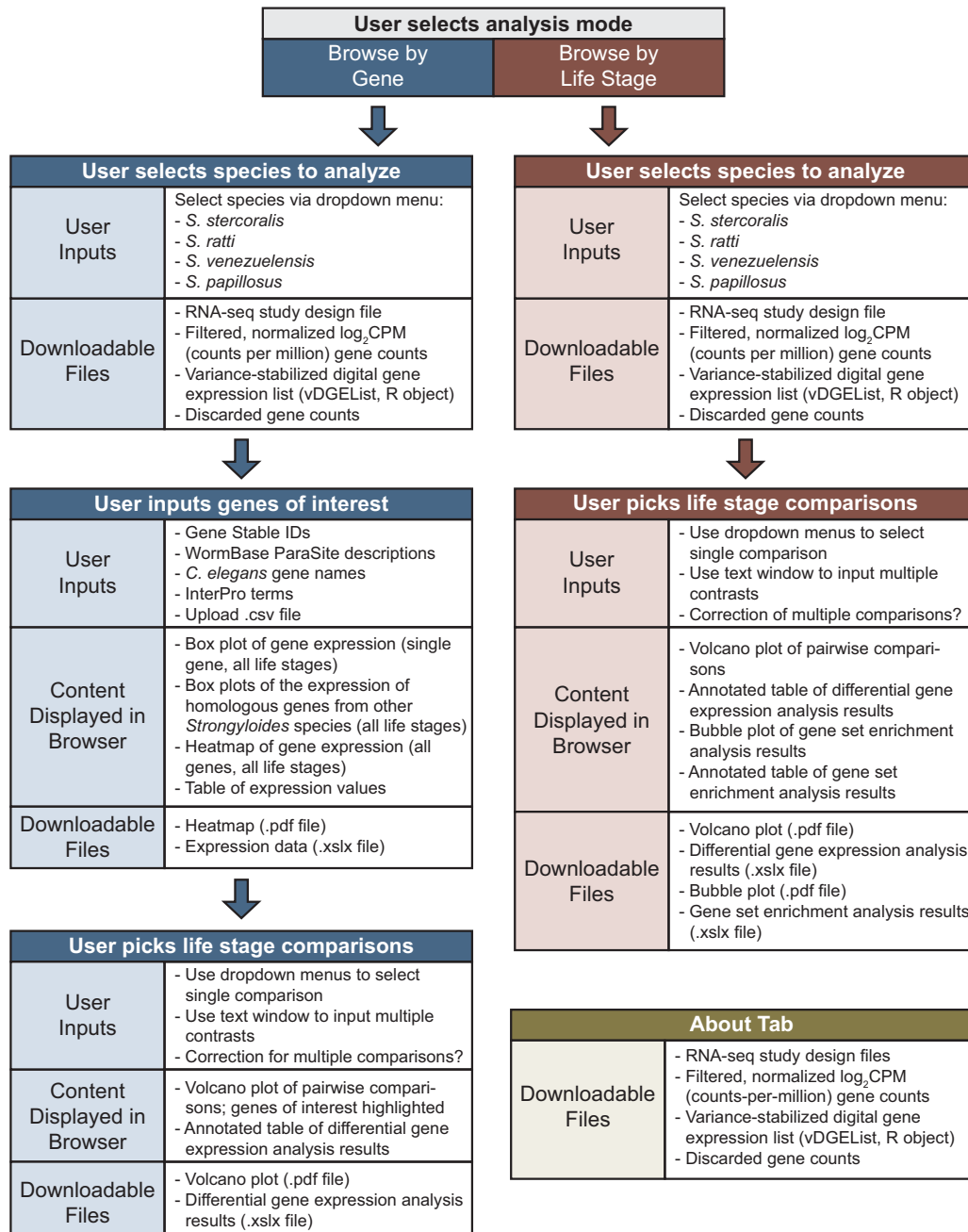

Figure 2 *Strongyloides* RNA-seq Browser UI overview. Overview of the *Strongyloides* RNA-seq Browser, including user inputs, content displayed in browser, and downloadable files. The application permits browsing RNA-seq data in two modes: "Browse by Life Stage" mode (red panels) and "Browse by Gene" mode (blue panels). Separate panels linked by arrows indicate the dependencies in the user experience. For example, in "Browse by Gene" mode, users first select a species to analyze, then input genes of interest, then pick life stage comparisons. Subpanels labeled "Content Displayed in Browser" and "Downloadable Files" indicate software elements that are available after users have submitted the indicated user inputs. The user interface also includes an "About" tab that displays methods information and a menu for downloading preprocessed datafiles. For downloadable files, file formats are .csv files unless otherwise indicated.

PRJEB528.WBPS14.mRNA_transcripts (*S. stercoralis*)
PRJEB125.WBPS14.mRNA_transcripts (*S. ratti*)
PRJEB525.WBPS14.mRNA_transcripts (*S. papillosus*)
PRJEB530.WBPS14.mRNA_transcripts (*S. venezuelensis*)

Kallisto alignment and gene annotation

For each species, Kallisto was used to perform ultra-fast read mapping of raw reads to a reference transcriptome (Bray *et al.* 2016). Quality controls for raw data and Kallisto alignments were assessed and summarized using FastQC v0.11.9 and MultiQC v1.9 (Andrews 2010; Ewels *et al.* 2016). Kallisto-generated transcript abundance data were imported into R v3.6.3 using the tximport package v1.14.2; transcript read counts were generated using the lengthScaledTPM option (Soneson *et al.* 2015).

Gene annotations were imported from WormBase ParaSite via the biomaRt package v2.42.1 (Durinck *et al.* 2005, 2009). Annotation information includes: *C. elegans* homologs/percent homology, UniProKB number, InterPro terms, gene ontology (GO) terms, and gene homologs/percent homology for the other three *Strongyloides* species.

Data filtering and normalization

Transcript read counts were converted to $\log_2$ counts per million ($\log_2$CPM) using the edgeR package v3.28.1 (Robinson *et al.* 2010), then filtered to remove transcripts with low counts (Supplementary File S3). The filtering criteria varied across species as follows, depending on the number of biological replicates (Figure 1B): *S. stercoralis* and *S. venezuelensis*, 1 $\log_2$CPM in at least 3 samples; *S. ratti* and *S. papillosus*, 1 $\log_2$CPM in at least 1 sample.

Filtered read counts were normalized using the trimmed mean of M-values (TMM) method (Robinson and Oshlack 2010) to permit between-samples comparisons. The mean-variance relationship was modeled using a precision weights approach via the voom function in the limma package v3.42.2 (Law *et al.* 2014). For *S. venezuelensis* and *S. ratti*, the RNA-seq study design included technical replicates (Figure 1B); in those cases, voom-modeled data were condensed by replacing technical replicates with their weighted average using the limma::avearrays function.

Application function and benchmarking

In-app gene lookup

Users may search for genes of interest using the following identifiers: gene IDs with prefixes "SSTP," "SRAE," "SPAL," and "SVE"; keywords matched against WormBase ParaSite gene descriptions; InterPro term; or parasite Ensembl Compara protein family names (Hunt *et al.* 2016). For *S. ratti*, users may also search for gene IDs with the prefix "WB." Users may provide *C. elegans* gene names, which will retrieve known *Strongyloides* homologs based on WormBase ParaSite classifications.

In-app DGE

The limma package is used to conduct pairwise DGE analyses between life stages (Ritchie *et al.* 2015; Phipson *et al.* 2016). Specific contrasts are defined via user inputs. For all species, variance-stabilized, filtered, normalized $\log_2$CPM values are fit to a linear model using the limma::lmFit function. The design matrix used to fit the linear model specifies no intercept/blocking, with comparisons across life stages. To provide increased statistical power, we use empirical Bayes smoothing of gene-wise standard deviations via the limma::eBayes function (Smyth 2004). Differentially expressed genes are identified using the limma::decideTests function: *p*-values are adjusted for multiple gene-wise

comparisons using the Benjamini-Hochberg false discovery rate (FDR) method. When optional correction for multiple pairwise comparisons is specified, the "global" method for multiple testing is applied; otherwise the "separate" option is used. Significantly expressed genes are classified based on an FDR of ≤0.05 and an absolute log fold change ($\log_2$FC) of ≥1. In "Browse by Gene" mode, DGE analyses are conducted first on the entire genome; then results for user-specified genes of interest are extracted.

Functional enrichment analysis

Gene set enrichment analysis (GSEA) is performed via the GSEA function in the clusterProfiler package v3.14.3 (Yu *et al.* 2012) using gene sets extracted from a database of parasite Ensembl Compara protein families (Hunt *et al.* 2016). On-demand GSEA is performed on gene lists rank-ordered by $\log_2$FC. Offline GSEA of genes contributing to PC1/PC2 identity was performed on a gene list rank-ordered by PC1 or PC2 gene scores; the two ranked lists were analyzed independently, then merged for plotting.

In-app data visualization

For heatmaps of $\log_2$CPM gene expression, life stages are ordered using Spearman clustering and genes are ordered using Pearson clustering; clustering is performed on user-defined genes using the stats package. Heatmaps are plotted using a local copy of the ggheatmap function in the heatmaply package v1.1.1 (Galili *et al.* 2018); this copy alters the default plot margins, as well as how tick labels and plot titles are displayed. Clustering dendrograms are calculated and plotted using the dendextend package v1.4.0. Plots of individual gene expression across life stages, expression of *Strongyloides* gene homologs across life stages, DGE volcano plots, and GSEA bubble plots are generated using the ggplot2 package v3.3.2 (Wickham 2009). Identification of *Strongyloides* homologs is based on gene homology annotations in WormBase ParaSite. Interactive tables are generated using the DT package v0.14 and saved using the openxlsx package v4.2.3. Users may download any plot or table as a PDF or Excel file, respectively. When downloading DGE analysis results as an Excel file, users are offered several filtering options, including: download genes with a specific direction of differential expression (*e.g.*, only upregulated genes); download a specific percentage of differentially expressed genes based on $\log_2$FC value; and only download genes with selected expression types in all searched pairwise comparisons (*e.g.*, only genes upregulated in iL3s relative to all other life stages).

PCA and benchmarking analyses

For each species, principal component analysis (PCA) was run on filtered, normalized $\log_2$CPM data using the stats::prcomp function. For statistical comparisons of expression levels in genes exclusively found in the *Strongyloides* RNA-seq Browser relative to genes also included in benchmarking datasets, the median $\log_2$CPM expression values across biological replicates were calculated for each life stage. Next, 2-way ANOVAs (Type III) and Tukey HSD *post-hoc* tests were run using the stats package and the car package version 3.0-8.

Web resources

A web-hosted version of the app is available at: https://hallemlab.shinyapps.io/strongyloides_rnaseq_browser/

Due to Shinyapps.io memory constraints, computationally intensive analyses (*e.g.*, exploring gene expression for >2000 genes) may require users to run apps locally.
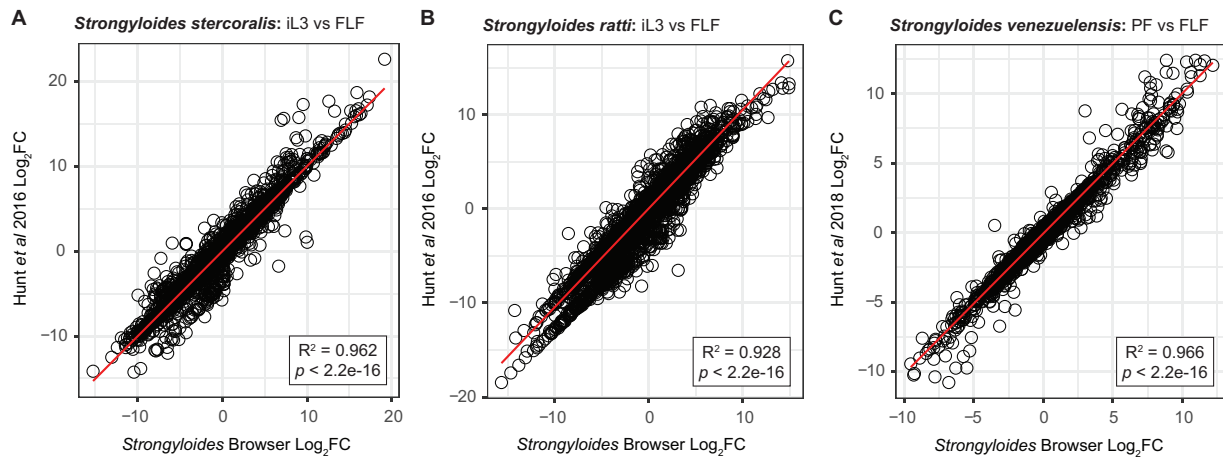
Figure 3 Benchmarking of selected differential expression analyses. Comparison of results generated by the *Strongyloides* RNA-seq Browser to previously published analyses in which the same datasets were analyzed using different pipelines. (A-B) For *S. stercoralis* and *S. ratti*, iL3 versus free-living female (FLF) DGE results are benchmarked against data from Hunt *et al.* (2016). (C) For *S. venezuelensis*, parasitic female (PF) versus free-living female (FLF) DGE results are benchmarked against data published in Hunt *et al.* (2018). For *S. papillosus*, $\log_2$ fold change ($\log_2$FC) values from previously published analyses are not available. For all plots, black circles represent individual genes. *X*-axis values are $\log_2$FC values as calculated by the *Strongyloides* RNA-seq Browser; *y*-axis values are previously published $\log_2$FC values. Red lines indicate a linear regression comparing previously published values to those calculated by the *Strongyloides* RNA-seq Browser. Plot inlays report R^2 and significance values for goodness of fit. These values indicate that for individual genes, there is a high correspondence between DGE as calculated by the *Strongyloides* RNA-seq Browser and previously published values. Only genes included in both the *Strongyloides* RNA-seq Browser and previously published datasets were included in this analysis.

App source code and deployment instructions are available at: https://github.com/HallemLab/Strongyloides_RNAseq_Browser

Data availability

The source code for: species-specific preprocessing pipelines, including MultiQC reports; example analyses, including additional multivariate quantification and plots; and other Supplementary materials are available at: https://github.com/HallemLab/Bryant-DeMarco-Hallem-2021.

Supplementary figures and the following Supplementary files have been uploaded to figshare. Supplementary File S1 contains a code freeze of the *Strongyloides* RNA-seq Browser. Supplementary File S2 contains metadata tables for the four *Strongyloides* species included in the *Strongyloides* RNA-seq Browser. For all species, metadata information was retrieved from the European Nucleotide Archive. For *S. stercoralis*, additional protocol information was retrieved from ArrayExpress. Supplementary File S3 contains low-count genes and associated $\log_2$CPM values discarded during preprocessing. Supplementary File S4 contains a list of the top 10% of genes contributing to PC1 and PC2 identity for each species. Genes were extracted based on PCA variable scores using tidyverse package version 1.3.0. Supplementary File S5 contains a list of chemoreceptor gene IDs for all *Strongyloides* species (Langeland *et al.* 2020; Wheeler *et al.* 2020). Supplementary File S6 is an example .xlsx results file generated by the *Strongyloides* RNA-seq Browser containing a DGE table listing 64 *S. stercoralis* chemoreceptors that are consistently upregulated in iL3s versus all other life stages. Supplementary material is available at figshare: https://doi.org/10.25387/g3.14273627.

Results and discussion
Software functionality

The *Strongyloides* RNA-seq Browser has two usage modes: "Browse by Life Stage" mode and "Browse by Gene" mode (Figure 2, Supplementary Figure S2). In "Browse by Gene" mode, users

search for gene(s) of interest by providing gene IDs or keywords that are matched against a gene annotation database. The program next offers four plotting options: a heatmap that plots the clustered expression profiles for all genes of interest across life stages, a box and whisker plot of $\log_2$CPM expression across life stages for each gene of interest, box and whisker plots of $\log_2$CPM expression across life stages for the known homologs of each gene of interest in the other three *Strongyloides* species, and a table of expression values. For both modes, users next input pairwise contrasts between life stages for limma-voom-based DGE analysis. Analysis results are displayed as downloadable volcano plots and data tables.

In "Browse by Life Stage" mode, the results of DGE analyses are used to perform GSEA, using an established parasite Ensembl Compara protein family database (Hunt *et al.* 2016). The GSEA analysis returns a downloadable bubble plot of enriched gene families as well as a downloadable table containing normalized gene enrichment scores that represent the degree to which the elements of the gene set are over-represented at the edges of the ranked gene list.

Benchmarking and example usage
Benchmarking

To assess the relative accuracy of the Kallisto::limma-voom analysis pipeline, we compared *Strongyloides* RNA-seq Browser data and results against previously published analyses. For *S. stercoralis* and *S. ratti*, we conducted benchmarking relative to a dataset generated via a TopHat::edgeR pipeline (Hunt *et al.* 2016). For *S. venezuelensis*, we compared results to a dataset generated via an HT-seq::cufflinks::edgeR pipeline (Hunt *et al.* 2018).

We first assessed the ability of our preprocessing pipeline to generate filtered and normalized expression data similar to previously published datasets. For all species, the number of genes included in the *Strongyloides* RNA-seq Browser was greater than the number of genes listed in the published datasets (Supplementary Figure S3A). This discrepancy may be due to differences in the preprocessing pipelines, as well as differences in threshold
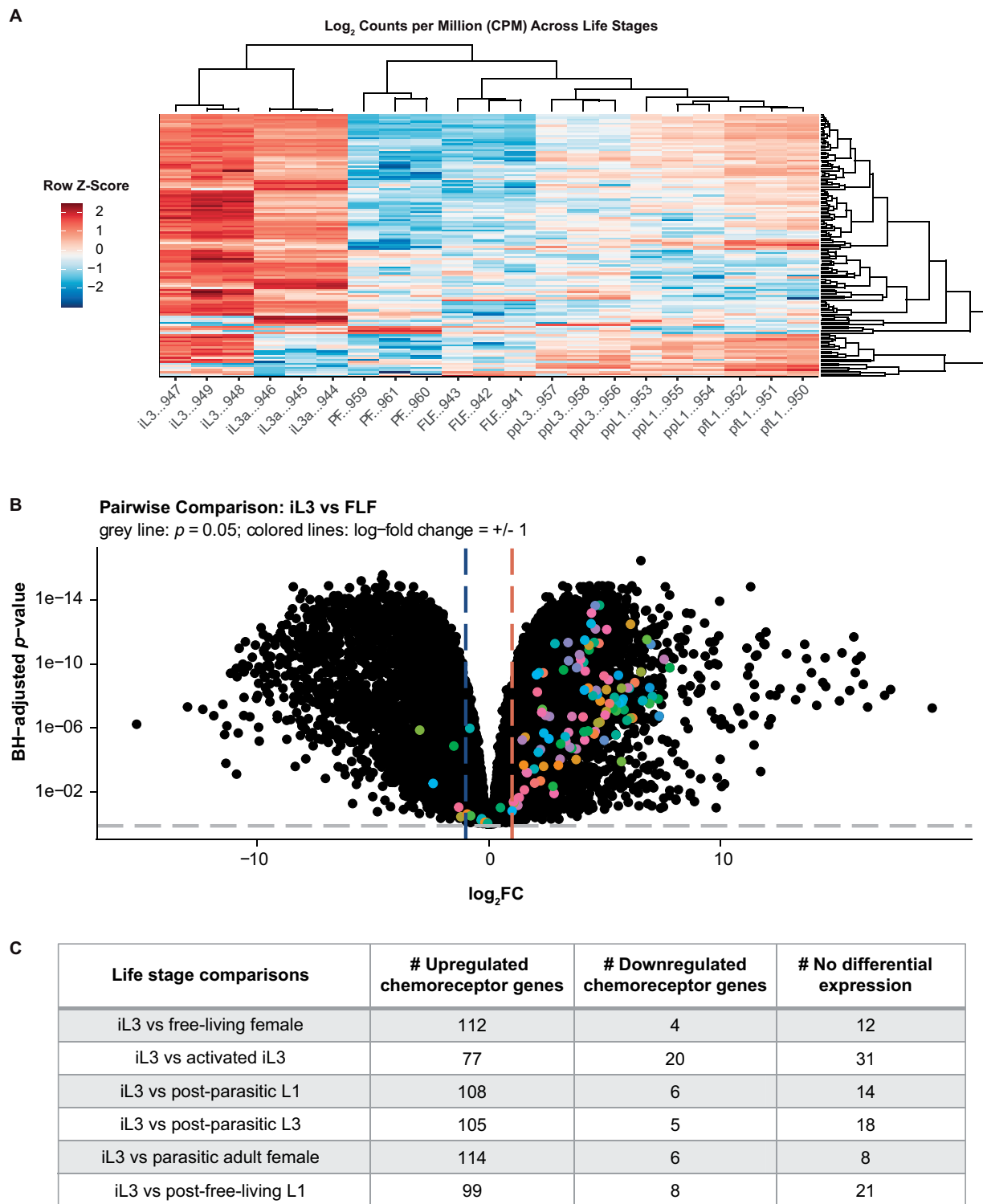
A

Log$_2$ Counts per Million (CPM) Across Life Stages

B

Pairwise Comparison: iL3 vs FLF
grey line: p = 0.05; colored lines: log–fold change = +/- 1

C

Life stage comparisons	# Upregulated chemoreceptor genes	# Downregulated chemoreceptor genes	# No differential expression
iL3 vs free-living female	112	4	12
iL3 vs activated iL3	77	20	31
iL3 vs post-parasitic L1	108	6	14
iL3 vs post-parasitic L3	105	5	18
iL3 vs parasitic adult female	114	6	8
iL3 vs post-free-living L1	99	8	21

Figure 4 Chemoreceptor gene expression is upregulated in *S. stercoralis* infective larvae. (A) Heatmap of log$_2$ counts per million (CPM) expression of 128 *S. stercoralis* chemoreceptor genes across developmental life stages. Columns represent RNA-seq samples and are labeled with the appropriate life stage abbreviation and the last 3 digits of the sample ID number. Column ordering and dendrogram reflect results of Spearman clustering; row order and dendrogram reflect results of Pearson clustering. Heatmap color scale is based on a row Z-score calculated by centering and row-scaling the log$_2$CPM values for each gene. Life stage abbreviations are as follows: infective third-stage larvae (iL3), activated iL3s (iL3a), parasitic adult females (PF), free-living females (FLF), post-parasitic first-stage larvae (ppL1), post-parasitic third-stage larvae (ppL3), and post-free-living first-stage larvae (pfL1). (B) Volcano plot of DGE between iL3s and free-living females (FLF). Positive log$_2$FC values indicate enrichment in iL3s relative to free-living females; negative log$_2$FC values indicate enrichment in free-living females relative to iL3s. Black dots are values for all *S. stercoralis* genes; colored dots indicate chemoreceptor genes. Gray line indicates Benjamini-Hochberg-adjusted p-value of 0.05, colored lines indicate log$_2$FC = 1. (C) Summary of differentially expressed chemoreceptor gene counts in iL3s versus other life stages. Genes were categorized as significantly upregulated or downregulated if they displayed an absolute log$_2$FC of $\geq$1 and a false-discovery-rate-corrected p-value of $\leq$0.05.

cutoffs used to filter genes for inclusion in DGE datasets. Indeed, genes found exclusively in the *Strongyloides* RNA-seq Browser dataset displayed lower expression levels than genes found in both datasets, for all species across life stages (Supplementary Figure S3, B–D). This suggests that more permissive inclusion criteria contribute to the greater size of the *Strongyloides* RNA-seq Browser dataset.

Next, we assessed whether our preprocessing pipeline preserves known associations between RNA-seq samples. For each species, we assessed the variability in expression across RNA-seq samples by performing PCA using the filtered, normalized gene expression data included in the *Strongyloides* RNA-seq Browser. In general, samples representing the same life stage clustered together on plots of the first two principal components for each species; biological and technical replicates of individual life stages were similar (Supplementary Figures S4 and S5). Consistent with previous findings (Lu *et al.* 2020), we observed that for *S. stercoralis*, the first two principal components distinguish between adult and larval life stages (PC1; Supplementary Figure S5) and between non-infective and parasitic life stages (PC2; Supplementary Figure S5).

To test whether genes influencing PC1 and PC2 identity in *S. stercoralis* are enriched for specific parasite gene families, we performed GSEA on a ranked list of PC1/PC2 scores, using previously established parasite gene families (Hunt *et al.* 2016). Gene families significantly contributing to *S. stercoralis* PC1 and PC2 include multiple families encoding hypothetical proteins as well as families that were previously identified as having putative roles in parasitism: astacin-like proteins, SCP/TAP proteins, trypsin inhibitors, and acetylcholinesterases (Supplementary Figure S6) (Hunt *et al.* 2016, 2018). For all species, we extracted and saved the top 10% of genes contributing to PC1 and PC2 identity (Supplementary File S4).

To benchmark DGE results, we used "Browse by Life Stage" mode to perform the following previously published pairwise comparisons: *S. stercoralis* iL3s *vs* free-living females, *S. ratti* iL3s *vs* free-living females, and *S. venezuelensis* parasitic females *vs* free-living females (Hunt *et al.* 2016, 2018). For all species, the magnitude of differential expression in individual genes was highly similar across methods (Figure 3). Thus, the *Strongyloides* RNA-seq Browser implements a DGE analysis pipeline that replicates known differential expression patterns in *Strongyloides* species, while providing flexible, on-demand analysis options.

Analysis of chemoreceptor genes

As an example of application usage, we applied the *Strongyloides* RNA-seq Browser to a published dataset of *S. stercoralis* chemoreceptor genes (Wheeler *et al.* 2020). In parasitic nematodes, chemosensation plays an important role in driving life-stage-specific, ethologically relevant behaviors such as host seeking and host invasion (Bryant and Hallem 2018; Banerjee and Hallem 2020; Wheeler *et al.* 2020). For example, *S. stercoralis* iL3s respond robustly to a number of human-associated odorants and have distinct olfactory preferences from those of other life stages (Castelletto *et al.* 2014; Gang *et al.* 2020). Many nematode chemoreceptors are G protein-coupled receptors (GPCRs), and nematode genomes encode large numbers of predicted chemoreceptor GPCRs (Bargmann 2006; Langeland *et al.* 2020; Wheeler *et al.* 2020). Putative chemosensory GPCR genes are highly divergent across nematode species (Wheeler *et al.* 2020), leading to the functional diversity that likely underlies species-specific differences in chemosensory behaviors (Hallem *et al.* 2011; Dillman *et al.* 2012; Castelletto *et al.* 2014; Lee *et al.* 2016; Gang *et al.* 2020). Using the *Strongyloides* RNA-seq Browser, we asked whether *Strongyloides* putative chemoreceptor genes display life-stage-specific expression patterns that may underlie life-stage-specific behaviors.

We analyzed 167 *S. stercoralis* chemoreceptor genes using "Browse by Gene" mode (Supplementary File S5). Clustering analysis of $\log_2$CPM expression of *S. stercoralis* chemoreceptor genes across life stages revealed that expression patterns in iL3s and activated iL3s are highly similar, compared to all other life stages (Figure 4A). We next used the browser to perform pairwise comparisons between *S. stercoralis* chemoreceptor gene expression in iL3s versus every other life stage; these comparisons confirmed that the expression of most chemoreceptor genes is significantly higher in iL3s ($\log_2$FC $\geq$ 1, FDR $\leq$ 0.05) (Figure 4B and C). Furthermore, we identified 64 chemoreceptor genes that are selectively upregulated in iL3s relative to all other life stages (Supplementary File S6). These results agree with previous observations that key elements of the chemosensory signal transduction pathway that are located downstream of GPCRs are upregulated in iL3s and that generally, genes upregulated in iL3s are associated with sensing the environment (Stoltzfus *et al.* 2012; Hunt *et al.* 2016).

Interestingly, although most GPCR genes are downregulated in non-iL3 life stages, *S. stercoralis* free-living adults display robust chemosensory behaviors, including broad attraction to host-associated odorants (Castelletto *et al.* 2014; Gang *et al.* 2020). Furthermore, despite profound increases in GPCR gene expression levels in *S. stercoralis* iL3s, the iL3s appear to be attracted to fewer host-associated odorants than free-living adults, a difference that appears to be dominated by a lack of attraction to fecal odorants in iL3s (Gang *et al.* 2020). How the widespread upregulation of GPCR genes in iL3s relates to the more selective olfactory preferences of iL3s for host odorants is not yet clear. One intriguing possibility is that the GPCR genes upregulated in iL3s may act redundantly, reinforcing iL3 attraction to a limited number of host odorants in order to ensure successful detection of host animals. Alternatively, iL3-upregulated GPCRs may confer responses to host-specific odorants that have not yet been tested but that selectively contribute to iL3-specific behaviors such as skin penetration and host invasion. In the future, characterizing the function of these GPCRs may provide critical mechanistic insight into the parasitic behaviors of *Strongyloides* nematodes.

In summary, the *Strongyloides* RNA-seq Browser prioritizes a user experience that provides access to *Strongyloides* genomics data and bioinformatics analyses without requiring previous coding experience. The open-source code supports future expansion: additional bioinformatics analyses can be included as requested, and the bulk RNA-seq processing and analysis pipelines may be easily reused as additional data become publicly available. Ultimately, we hope that the *Strongyloides* RNA-seq Browser will facilitate future genetic studies of *Strongyloides* species by acting as a user-friendly resource for researchers seeking to understand the functional roles of specific genes and gene families in parasitic nematodes.

Acknowledgments

Funding

Literature cited

Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: 2021 April 6).

Banerjee N, Hallem EA. 2020. The role of carbon dioxide in nematode behaviour and physiology. Parasitology. 147:841–854.

Bargmann C. 2006. Chemosensation in *C. elegans*. In WormBook, http://www.wormbook.org (Accessed: 2021April 6).

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 34:525–527.

Bryant AS, and Hallem EA. 2018. Terror in the dirt: sensory determinants of host seeking in soil-transmitted mammalian-parasitic nematodes. Int J Parasitol Drugs Drug Resist. 8:496–510.

Bryant AS, Ruiz F, Gang SS, Castelletto ML, Lopez JB, *et al.* 2018. A critical role for thermosensation in host seeking by skin-penetrating nematodes. Curr Biol. 28:2338–2347.

Buonfrate D, Bisanzio D, Giorli G, Odermatt P, Fürst T, *et al.* 2020. The global prevalence of *Strongyloides stercoralis* infection. Pathogens. 9:468.

Castelletto ML, Gang SS, and, Hallem EA. 2020. Recent advances in functional genomics for parasitic nematodes of mammals. J Exp Biol. 223:jeb206482.

Castelletto ML, Gang SS, Okubo RP, Tselikova AA, Nolan TJ, *et al.* 2014. Diverse host-seeking behaviors of skin-penetrating nematodes. PLoS Pathog. 10:e1004305.

Dillman AR, Guillermin ML, Lee JH, Kim B, Sternberg PW, *et al.* 2012. Olfaction shapes host-parasite interactions in parasitic nematodes. Proc Natl Acad Sci USA. 109:E2324–E2333.

Durinck S, Moreau Y, Kasprzyk A, Davis S, Moor BD, *et al.* 2005. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 21:3439–3440.

Durinck S, Spellman PT, Birney E, Huber W. 2009. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc. 4:1184–1191.

Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 32:3047–3048.

Galili T, O'Callaghan A, Sidi J, Sievert C. 2018. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. Bioinformatics. 34:1600–1602.

Gang SS, Castelletto ML, Bryant AS, Yang E, Mancuso N, *et al.* 2017. Targeted mutagenesis in a human-parasitic nematode. PLoS Pathog. 13:e1006675.

Gang SS, Castelletto ML, Yang E, Ruiz F, Brown TM, *et al.* 2020. Chemosensory mechanisms of host seeking and infectivity in skin-penetrating nematodes. Proc Natl Acad Sci USA. 117:17913–17923.

Hallem EA, Dillman AR, Hong AV, Zhang Y, Yano JM, *et al.* 2011. A sensory code for host seeking in parasitic nematodes. Curr Biol. 21:377–383.

Howe KL, Bolt BJ, Shafie M, Kersey P, and, Berriman M. 2017. WormBase ParaSite – a comprehensive resource for helminth genomics. Mol Biochem Parasitol. 215:2–10.

Hunt VL, Hino A, Yoshida A, Kikuchi T. 2018. Comparative transcriptomics gives insights into the evolution of parasitism in *Strongyloides* nematodes at the genus, subclade and species level. Sci Rep. 8:5192.

Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, *et al.* 2016. The genomic basis of parasitism in the *Strongyloides* clade of nematodes. Nat Genet. 48:299–307.

Langeland A, Hawdon JM, O'Halloran DM. 2020. NemChR-DB: a database of parasitic nematode chemosensory G-protein coupled receptors. Int J Parasitol. 7519:30314–3.

Law C W, Chen Y, Shi W, Smyth G K. 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15:R29.

Lee JH, Dillman AR, and, Hallem EA. 2016. Temperature-dependent changes in the host-seeking behaviors of parasitic nematodes. BMC Biol. 14:36.

Lok JB, Shao H, Massey HC, Li X. 2017. Transgenesis in *Strongyloides* and related parasitic nematodes: historical perspectives, current functional genomic applications and progress towards gene disruption and editing. Parasitology. 144:327–342.

Lu MR, Lai C-K, Liao B-Y, Tsai IJ. 2020. Comparative transcriptomics across nematode life cycles reveal gene expression conservation and correlated evolution in adjacent developmental stages. Genome Biol Evol. 12:1019–1030.

Phipson B, Lee S, Majewski IJ, Alexander WS, Smyth GK. 2016. Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann Appl Stat. 10:946–963.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, *et al.* 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43:e47.

Roberts LS, Janovy J. 2005. Gerald D. Schmidt & Larry S. Roberts' Foundations of Parasitology. Boston, MA: McGraw-Hill.

Robinson MD, McCarthy DJ, and, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26:139–140.

Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11:R25.

Schafer TW, , and, Skopic A. 2006. Parasites of the small intestine. Curr Gastroenterol Rep. 8:312–320.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 3:1–25.

Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 4:1521.

Stoltzfus JD, Minot S, Berriman M, Nolan TJ, Lok JB. 2012. RNAseq analysis of the parasitic nematode *Strongyloides stercoralis* reveals divergent regulation of canonical dauer pathways. PLoS Negl Trop Dis. 6:e1854.

Wheeler NJ, Heimark ZW, Airs PM, Mann A, Bartholomay LC, *et al.* 2020. Genetic and functional diversification of chemosensory pathway receptors in mosquito-borne filarial nematodes. PLoS Biol. 18:e3000723.

Wickham H. 2009. Ggplot2: elegant Graphics for Data Analysis. New York, NY: Springer.

Yu G, Wang L-G, Han Y, He Q-Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 16:284–287.

Communicating editor: M.-A. Félix