ADDITIONAL DATA

Chronic toxicity testing including transcriptomics-based molecular profiling in *Cloeon dipterum*

Kirsten Germing[a]*, Steve U. Ayobahan[b], Hannes Reinwald[b], Maria Vogt[a], Benedikt Ringbeck[c], Bernd Göckener[c], Elke Eilebrecht[a], Lena Kosak[a] and Sebastian Eilebrecht[b]

[a]*Department of Ecotoxicology, Fraunhofer Institute for Molecular Biology and Applied Ecology, Schmallenberg, Germany*

[b]*Fraunhofer Attract Eco'n'OMICs, Fraunhofer Institute for Molecular Biology and Applied Ecology, Schmallenberg, Germany*

[c]*Department of Trace Analysis and Environmental Monitoring, Fraunhofer Institute for Molecular Biology and Applied Ecology, Schmallenberg, Germany*

**Address correspondence to:*
kirsten.germing@ime.fraunhofer.de
Fraunhofer Institute for Molecular Biology and Applied Ecology
Auf dem Aberg 1
57392 Schmallenberg
Germany
+49 2972 302177

1 ADDITIONAL MATERIALS & METHODS

Chemical analysis

a) Materials and methods

Test medium, prepared at Fraunhofer IME, was used as the aqueous portion in all samples, including calibration and quality control samples (matrix-matched calibration). Acetonitrile and methanol (CHEMSOLUTE® LC-MS grade, Th. Geyer GmbH & Co. KG, Renningen, Germany) were used for sample stabilization and as organic eluents, respectively. Ultrapure water (UHQ water, <18.2 MΩ · cm; PURELAB Ultra, ELGA LabWater, Veolia Water Technologies Deutschland GmbH, Celle, Germany) was used for the aqueous eluent. Ammonium acetate (LC-MS grade) was obtained from Honeywell (Charlotte, NC, USA).

b) Sample preparation

Samples were prepared by adding 5 mL of the test samples to 1 mL acetonitrile each in 8 mL glass tubes. The ratio of acetonitrile to test medium was thus 1+5 (v+v) in all samples. Samples were injected directly without any further preparation into the UHPLC-MS/MS system. Samples that were not immediately analyzed were stored frozen at ≤−18 °C after addition of acetonitrile until analysis.

c) UHPLC-MS/MS conditions

Sample analysis was conducted using a WATERS Acquity UPLC System with Xevo TQ-S Mass analyzer (Waters Corporation, Milford, MA, USA). Chromatographic separation was performed on an Acquity UPLC BEH C18 column (50 x 2.1 mm; 1.7 µm) (Waters) using H_2O/acetonitrile (95/5, v/v) and methanol, both with addition of 2 mM ammonium acetate, as aqueous (eluent A) and organic eluent (eluent B), respectively. Column and autosampler temperatures were 30 °C and 8 °C, respectively. The injection volume was 5 µL. The flow was 0.3 mL/min. The eluent gradient started as a composition of 30% B, held for 0.1 min, and increased to 100% B in 2.4 min. This composition was held for 2.9 min and was then decreased to initial composition of 30% B in 2.6 min (total run time 8 min).

For the short-term exposure test, a contamination of Fipronil from the eluents was observed. The contamination peak was chromatographically resolved by implementing a second chromatographic column between the eluent mixing chamber and the autosampler (Luna C18 (2) (150 x 2 mm; 3 µm) (Phenomenex Ltd., Torrance, CA, USA)). For these measurements, the chromatographic method was adapted by reducing the flow from 0.3 mL/min to 0.2 mL/min to deal with the higher back-pressure. To ensure equilibration of the column to initial eluent composition with regards to the reduced flow, the last chromatographic step was elongated from 2.6 min to 4.6 min. Due to the higher gradient delay volume and lower flow, the retention time was shifted from 3.55 min to 5.11 min. The contamination peak was baseline separated with a retention time of approx. 5.4 min. As the validity of both methods was given, the method adaption has no impact on the results of this study. Exemplary chromatograms

of both tests for a control sample and a sample at the lowest test concentration are shown in Figures S13 to S16.

Negative electrospray ionization (ESI(−)) was used and the two mass transitions of the pseudo-molecular ion ([M−H]−) at *m/z* 434.6 → 329.9 (collision energy: 31 eV) and m/z 434.6 → 249.9 (collision energy: 16 eV) were analyzed for quantification and quality control. Argon was used as collision gas. Ion source and mass analyzer settings were: Capillary 3.0 kV, Cone voltage 15.0 V; Source offset 50 V; Source Temperature 150 °C; Desolvation Temperature 350 °C; Cone Gas Flow 50 L/h; Desolvation Gas Flow 1000 L/h; Collision Gas Flow 0.15 mL/min, Nebuliser Gas Flow 7.00 bar; LM 1 Resolution 15.00; HM 1 Resolution 15.00; Ion Energy 1 0.50; MS Mode Collision Energy 0.00; MSMS Mode Collision Energy 31.00; MS Mode Entrance 1.00; MS Mode Exit 1.00; Gas On MS Mode Entrance 30.00; Gas On MS Mode Exit 30.00; Gas On MSMS Mode Entrance 1.00; Gas On MSMS Mode Exit 1.00; Gas Off MS Mode Entrance 30.00; Gas Off MS Mode Exit 30.00; Gas Off MSMS Mode Entrance 1.00; Gas Off MSMS Mode Exit 1.00; ScanWave MS Mode Entrance 1.00; ScanWave MS Mode Exit 1.00; ScanWave MSMS Mode Entrance 1.00; ScanWave MSMS Mode Exit 1.00; LM 2 Resolution 15.00; HM 2 Resolution 15.00; Ion Energy 2 0.50; Gain 1.00.

d) Calibration and quality control

Calibration and quality control samples were prepared from two individual weights of approx. 5 mg fipronil (analytical standard, identical with test substance), which were dissolved in each 10 mL acetonitrile. Six calibration samples were prepared in the range of 10 ng/L to 1000 ng/L. The ratio of test medium to acetonitrile was 5+1 (v+v) in all samples. LOD and LOQ were 10 ng/L and 30 ng/L, respectively. All concentrations referred to the aqueous portion of the samples. Validity of the method was confirmed by preparing quality control (QC) samples from the second stock solution in analogy to calibration samples. QC levels were 50 ng/L and 1000 ng/L. QC samples were measured before and after the test samples to ensure validity over the whole sample batch. Solvent blanks (acetonitrile) and matrix blanks (test medium + acetonitrile) were analyzed to ensure that no carry-over between injections or contamination took place.

Quadratic calibration (1/x weighted) was used for regression. The selection of the regression model was based on residual analysis. The coefficients of determination (r^2) were >0.999 for all calibrations. Recoveries of QC samples were within a range of 80 – 130%.

e) Results of the test samples

Results of the chemical analysis of fipronil in test samples can be found in Tables S3 and S4 for the short-term and long-term exposure tests, respectively. Results of the test samples were in the range of 80-120% of nominal values, except for the lowest test concentration, where a slightly higher variation (78-126%) was found. This variation can be explained as the concentrations are in the range of the LOQ. In this low concentration range, slight contamination during sample preparation or analysis (even at concentrations <LOD) may result in higher recoveries. On the other hand, lower recoveries might be

explained by adsorption or degradation, which often has a more pronounced effect at lower concentrations. However, values of 78% and 126%, respectively, are still close to nominal values so these slight imprecisions are deemed to have no influence on the results of this study.

Total RNA isolation and transcriptome sequencing

Total RNA was isolated from imagines (long-term exposure test) and larvae (short-term exposure test) using the NucleoSpin RNA/Protein kit (Macherey & Nagel, Düren, Germany). The organisms were homogenized in lysis buffer at 5 m/s for 40 seconds at room temperature with Lysing Matrix D ceramic beads using a FastPrep-24 homogenizer (MP Biomedicals, Irvine, USA) before proceeding with the NucleoSpin RNA/Protein kit according to the manufacturer's instructions. The yields of total RNA isolated from each sample were quantified using a Nanodrop system (Thermo Fisher Scientific, Waltham, USA), and the RNA quality was assessed using the RNA 6000 Nano Kit on an Agilent 2100 Bioanalyzer System (Agilent, Santa Clara, USA). RNA-Seq libraries were prepared using samples with RNA integrity numbers (RIN) greater than 9.0, ensuring high-quality RNA analysis. Poly(A)+ RNA was isolated from total RNA and subjected to library preparation using the TruSeq RNA library Prep Kit v2 following the manufacturer’s instructions. Libraries were sequenced in 2 x 150-bp paired-end mode on a NovaSeq 6000 system, yielding approximately 30 million raw reads per sample. Fastp was used to remove adapter sequences from demultiplexed fastq files (Chen et al. 2018) and the sequences were aligned to the *C. dipterum* reference genome (GCA 902829235.1) with STAR aligner v.2.5.2a (Dobin et al. 2013) and feature-mapped reads were counted using STAR's '—quantMode GeneCounts'. Raw and normalized read counts are shown in Figure S8.

Differential gene expression analysis (DGEA)

Mapped read tables from replicates and conditions were merged into a single count matrix and subjected to differential gene expression analysis (DGEA) in R v4.2.2 using DESeq2 v1.38.3 (Love et al. 2014). Prior to DESeq2's median of ratios normalization and statistical testing, low abundant genes with less total counts than the number of samples were removed. Unsupervised multivariate sample clustering methods were used to test the robustness of general underlying data patterns. The 75 % quantile of absolute unshrunk log2-fold change (lfc) values was used as the biological effect size cut-off for each treatment comparison. This scalable effect size cut-off adjusts with the global lfc distribution under the assumption that lfc values are equally distributed around 0. Following the determination of LFcut, lfc values were shrunk using the apeglm method (Zhu et al. 2019). Wald's t-test was applied to compare exposure treatments to the control group in a pairwise fashion. After Benjamini-Hochberg correction, p-values were corrected for multiple testing using independent hypothesis weighting (IHW) (Ignatiadis et al. 2016) (BH). Subsequently, a gene was considered as differentially expressed when its mean expression differed significantly from the control group (BH adjusted p (p.adj) ≤ 0.05) and the absolute apeglm shrunk lfc was greater LFcut (abs[apeglm(lfc)] $\geq$ LFcut). The core DEG set was defined as the common set of identified differentially expressed genes (DEGs) of low and high exposure conditions

for organisms exposed to fipronil and was evaluated using correlation analysis. ENSEMBL gene identifiers were annotated in R using biomaRt (Durinck et al. 2009). Lfc and adjusted p-value distributions are shown in Figures S10 and S11. PCA and t-SNE plots are shown in Figure S12 and the sample distance matrix is shown in Figure S13. Raw and processed data were deposited in ArrayExpress database at EMBL-EBI (www.ebi.uk/arrayexpress) (Athar et al. 2019) under accession numbers E-MTAB-12814 (Username: Reviewer_E-MTAB-12814, Password: pxwworux).

Functional genome annotation, overrepresentation analysis (ORA) and gene set enrichment analysis (GSEA)

To enable functional interpretation of *C. dipterum* RNA-Seq datasets in R via clusterProfiler (Cantalapiedra et al. 2021), a custom AnnotationDbi package was created following the workflow previously described by Loll et al. (2022) for Lemna minor (Loll et al. 2022). Similar to this approach, the org.Cdipterum.eg.db functional gene annotation package was created based on the homology of protein sequences translated from the *C. dipterum* reference genome to a reference database. Sequence homology was assessed using BLAST+ (Sayers et al. 2021) and eggnog (Cantalapiedra et al. 2021). The combined results were used for functional annotation of *C. dipterum* gene identifiers from the reference genome (GenBank assembly GCA_949628265.1). Details of the annotation workflow can be found in the supplements and in Loll et al. (2022) (Loll et al. 2022). The org.Cdipterum.eg.db package and the corresponding data files used to create this annotation are publicly available on Zenodo (www.zenodo.org) under accession number 8003805 (Reinwald 2023).

To elucidate the biological mechanisms affected by the test substance, clusterProfiler was used to perform enrichment analysis for GO terms for biological processes (BP) and molecular functions (MF). To identify BP and MF directly associated with the DEG sets of each sample group, an overrepresentation analysis (ORA) was performed. To better understand the direction of regulation in the affected BP and MF GO terms, gene set enrichment analysis (GSEA) was conducted for each sample group based on the apeglm shrunk lfc values. Details on ORA and GSEA are described in the supplements.

Functional annotation of the *C. dipterum* genome

To enable functional interpretation of RNA-Seq datasets of *C. dipterum* in R via clusterProfiler (Wu et al., 2021), a custom AnnotationDbi (Pagès et al., 2023) package was built following the workflow previously described by Loll and colleagues for Lemna minor (Loll et al., 2022). Similar to this approach, the org.Cdipterum.eg.db functional gene annotation package was constructed based on protein sequence homology of *C. dipterum*s proteome to a reference database.

At first, the reference proteome for *C. dipterum* was predicted from the gene coding sequences (CDS) obtained from the reference genome (GenBank assembly GCA_949628265.1) and respective genome annotation file created through the Darwin Tree of Life project (The Darwin Tree of Life Project

Consortium, 2022). For this, a multi fasta file listing all CDS was created using the reference genome's fasta and GTF file via cufflink's gffread function (Trapnell et al., 2010) before being translated into the respective amino acid sequences using transeq from the EMBOSS package (Rice et al., 2000).

In the next steps these *C. dipterum* proteome sequences were queried for homologous sequences using a) eggNOG (Huerta-Cepas et al., 2019) and b) the basic local alignment search tool BLAST+ (Camacho et al., 2009) against a custom-built query database. The query database for the blastp search was constructed from the Uniprot reference genomes of Apis mellifera (UP000005203) and Drosophila melanogaster (UP000000803) as functionally well annotated representatives of the insect class as well as all available protein sequences found for odonatan, ephemeroptera and acrythosiphon. Translated *C. dipterum* CDS sequences were subjected to a blastp search against this database and each gene was matched with the Uniprot ID of the best hit for each reference species, scored by %-alignment. The search results were cleaned from alignment lengths < 20 amino acids and alignment similarities < 35%. Each *C. dipterum* gene ID from these filtered results was then merged with the combined set of unique GO terms associated with the matched Uniprot IDs across the different queried taxa. For this step, the GO terms for the filtered blastp hits (Uniprot-ID) were obtained from Uniprot before.

For the eggNOG-based annotation, translated *C. dipterum* sequences were subjected to the eggNOG annotation pipeline with default settings. All non-insect related matches were removed from the eggNOG annotation, before results were merged with the filtered blastp search results. Hereby, GO terms from both searches were combined and the resulting gene2GO table (one gene mapped to multiple GO terms) was converted to a GO2gene format (one GO term mapped to multiple genes) using the topGO package function inverseList(). Deprecated GO terms were removed from the GO2gene table and GO ids were annotated with GO descriptions and ontology features (BP, MF or CC) using the clusterProfiler package functions go2term() and go2ont().

Ultimately, all information on the *C. dipterum* gene IDs (GID) were compiled and integrated in the AnnotationDbi organism package org.Cdipterum.eg.db. The package was built using the AnnotationForge package function makeOrgPackage(). This package maps the gene identifiers listed in the reference genome (GenBank assembly GCA_949628265.1) to a variety of functional information such as general gene descriptors, PFAM (Protein Family) domains and most importantly GO terms and KEGG pathways. The later is especially relevant as it greatly facilitates the usage of the GO term annotation for any type of functional enrichment analysis via the powerful clusterProfiler package. This package is suited to be directly called as the OrgDb variable in the enrichGO() and gseGO() function for overrepresentation analysis (ORA) and gene set enrichment analysis (GSEA) respectively. The org.Cdipterum.eg.db package as well as respective data files used to build this annotation (e.g. the reference genomes CDS fasta, BLAST query database, GID annotation table, etc.) is made publicly available on Zenodo (www.zenodo.org) under accession 8003805 (Reinwald, 2023).

Functional analysis of toxicogenomic signatures

To elucidate the biological mechanisms affected by the test compound, enrichment analysis for biological processes (BP) and molecular functions (MF) GO terms was performed using clusterProfiler v4.4.4 .in R. To identify BP and MF directly associated with the DEG set (p.adj ≤ 0.05 & abs(lfc) > LFcut) of each sample group, overrepresentation analysis (ORA) was performed via the enrichGO() function call within compareCluster(). The universe background was composed of the total set of genes analyzed with DESeq2 after low read count genes were removed. To better understand the direction of regulation of affected BP and MF, gene set enrichment analysis (GSEA) was performed additionally using the gseGO() function for each treatment group. The ranked input gene list for GSEA was created from the DESeq2 differential gene expression analysis (DGEA) result tables (for which low count genes were already removed prior DESeq2 run). Genes were ranked by their apeglm shrank log2-fold change values. In both cases (ORA & GSEA) the orgDb parameter was specified as "org.Cdipterum.eg.db" with keyType="GID". Minimal size of genes annotated by Ontology term for testing was set to 10 and maximal size of genes annotated for testing was set to 5. p value adjustment method for multiple testing was performed after BH. Terms with p.adjust ≤ 0.05 were considered significantly enriched. To allow for simplification of enriched GO terms, semantic similarities among the different GO terms were computed after the "Wang" method via the godata() function from the GOSemSim package v2.22. Subsequently redundant GO term results were simplified using the simplify() function with a similarity cutoff of 0.8 from the same package. The main function calls for ORA and GSEA are provided on Zenodo (www.zenodo.org) under accession 8003805. Enrichment data visualizations were created in R using clusterProfiler and ggplot2.

2 ADDITIONAL TABLES

Table S1: Composition of nutrient solution for *Navicula pelliculosa* culture.

Nutritional salt	Concentration in nutrient solution [g/L]
$NaNO_3$	5.25
KH_2PO_4	1.15
Na_2SiO_3	5.29

Table S2: Water parameters of the long- and short-term exposure tests.

	Treatment [ng/L]	pH	Oxygen concentration [mg/L]	Oxygen saturation [%]	Light intensity [lux]	Tempe-rature [°C]
		Mean ± Standard abbreviation				
Long-term exposure test	Control	7.89 ± 0.47	7.44 ± 0.75	86.8 ± 8.7	49.7 ± 12.5	20.7 ± 0.3
	38.0	7.88 ± 0.50	7.57 ± 0.80	88.4 ± 9.5		
	75.0	7.96 ± 0.55	7.48 ± 0.98	87.2 ± 11.5		
	150	8.01 ± 0.57	7.59 ± 0.84	88.4 ± 9.8		
	300	8.01 ± 0.60	7.61 ± 0.79	88.7 ± 9.3		
	600	7.99 ± 0.60	7.64 ± 0.65	89.4 ± 7.9		
Short-term exposure test	Control	7.64 ± 0.38	7.64 ± 0.62	96.2 ± 9.2	34.3 ± 3.6	20.5 ± 0.2
	38.0	7.76 ± 0.35	7.44 ± 1.01	94.4 ± 15.2		
	75.0	7.67 ± 0.40	7.28 ± 0.85	92.4 ± 12.9		
	150	7.72 ± 0.36	7.30 ± 0.69	92.7 ± 10.7		
	300	7.77 ± 0.36	7.35 ± 0.54	93.1 ± 9.0		

Table S3: Time-weighted mean concentrations of fipronil of the long-term exposure test.

Nominal conc. [ng/L]	Mean conc. of fresh test solutions [µg/L]	Recovery of nominal [%]	Mean conc. of aged test solutions [µg/L]	Recovery of nominal [%]	Recovery of initial [%]	Time-weighted mean conc. [µg/L]	Recovery of nominal [%]
Control	< LOD	-	-	-	-	-	-
38.0	0.047	122.36	0.045	122.60	96.15	0.047	125.92
75.0	0.081	108.74	0.082	98.93	101.92	0.080	106.41
150	0.157	105.48	0.146	97.33	93.19	0.148	98.54
300	0.299	100.70	0.281	93.63	93.92	0.277	92.34
600	0.596	100.85	0.573	95.56	96.22	0.568	94.73

Table S4: Time-weighted mean concentrations of fipronil of the short-term exposure test.

Nominal conc. [ng/L]	Mean conc. of fresh test solutions [µg/L]	Recovery of nominal [%]	Mean conc. of aged test solutions [µg/L]	Recovery of nominal [%]	Recovery of initial [%]	Time-weighted mean conc. [µg/L]	Recovery of nominal [%]
Control	< LOD	-	-	-	-	-	-
38.0	0.024*	64.0	0.034	89.3	139.6	0.029	78.0
75.0	0.070	92.7	0.065	86.7	93.5	0.067	89.4
150	0.145	96.7	0.120	79.7	82.4	0.132	87.8
300	0.303	101.0	0.193	64.3	63.7	0.244	81.3

*value <LOQ but still within the calibration range (>LOD)

Table S5: Percentage of larvae ≥ L4 per replicate and treatment after three days exposure in the long-term exposure toxicity test with *C. dipterum*.

	Nominal concentration [ng/L]					
	Control	**38.0**	**75.0**	**150**	**300**	**600**
	Percentage of larvae ≥ L4 [%]					
/1	40	25	0	50	0	60
/2	60	60	60	60	20	20
/3	60	40	60	40	40	60
/4	60	60	40	40	80	0
Mean	**55**	**46**	**40**	**48**	**35**	**35**
Standard abbreviation	**10.0**	**17.0**	**28.3**	**9.6**	**34.2**	**30.0**

Table S6: Percentage of larvae ≥ L6 per replicate and treatment after seven days exposure in the long-term exposure toxicity test with *C. dipterum*.

	Nominal concentration [ng/L]					
	Control	**38.0**	**75.0**	**150**	**300**	**600**
	Percentage of larvae ≥ L6 [%]					
/1	100.0	100.0	40.0	50.0	50.0	50.0
/2	80.0	60.0	50.0	60.0	20.0	0.0
/3	100.0	80.0	60.0	25.0	40.0	0.0
/4	60.0	100.0	40.0	50.0	20.0	0.0
Mean	**85.0**	**85.0**	**47.5**	**46.3**	**32.5**	**12.5**
Standard abbreviation	**19.1**	**19.1**	**9.6**	**14.9**	**15.0**	**25.0**

Table S7: Percentage of larvae ≥ L7 per replicate and treatment after ten days exposure in the long-term exposure toxicity test with *C. dipterum*.

	Nominal concentration [ng/L]					
	Control	**38.0**	**75.0**	**150**	**300**	**600**
	Percentage of larvae ≥ L7 [%]					
/1	100.0	50.0	20.0	50.0	0.0	33.3
/2	80.0	50.0	75.0	40.0	25.0	0.0
/3	80.0	60.0	40.0	25.0	25.0	33.3
/4	60.0	60.0	50.0	25.0	20.0	0.0
Mean	**80.0**	**55.0**	**46.3**	**35.0**	**17.5**	**16.7**
Standard abbreviation	**16.3**	**5.8**	**22.9**	**12.2**	**11.9**	**19.2**

Table S8: Percentage of larvae ≥ L4 per replicate and treatment after three days exposure in the short-term exposure toxicity test with *C. dipterum*.

	Nominal concentration [ng/L]				
	Control	**38.0**	**75.0**	**150**	**300**
	Percentage of larvae ≥ L4 [%]				
/1	40	0	40	60	20
/2	60	40	40	0	0
/3	60	80	20	75	40
/4	40	40	60	40	40
/5	80	-	-	-	-
/6	40	-	-	-	-
Mean	**53**	**40**	**40**	**44**	**25**
Standard abbreviation	**16.3**	**32.7**	**16.3**	**32.5**	**19.1**

Table S9: Percentage of larvae ≥ L6 per replicate and treatment after seven days exposure in the short-term exposure toxicity test with *C. dipterum*.

	Nominal concentration [ng/L]				
	Control	**38.0**	**75.0**	**150**	**300**
	Percentage of larvae ≥ L6 [%]				
/1	80.0	33.3	60.0	75.0	50.0
/2	80	100.0	80.0	50.0	40.0
/3	100	100	60	50	40.0
/4	100	80.0	66.7	50.0	50.0
/5	60	-	-	-	-
/6	60	-	-	-	-
Mean	**80.0**	**78.3**	**66.7**	**56.3**	**45.0**
Standard abbreviation	**17.9**	**31.4**	**9.4**	**12.5**	**5.8**

3 ADDITIONAL FIGURES

Figure S1: Left: *C. dipterum* larva in larval stage L3; Right: *C. dipterum* larva in larval stage N1.

Figure S2: Larval development stages based on wing pad development of *C. dipterum* (adopted from Cianciara 1976).

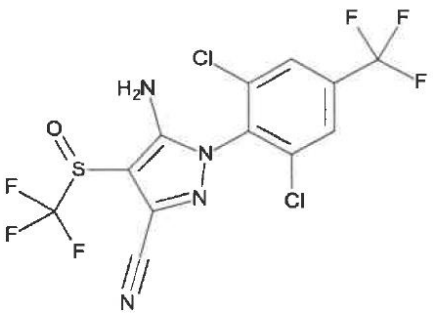


Figure S3: Molecular sturcture of the test substance fipronil.

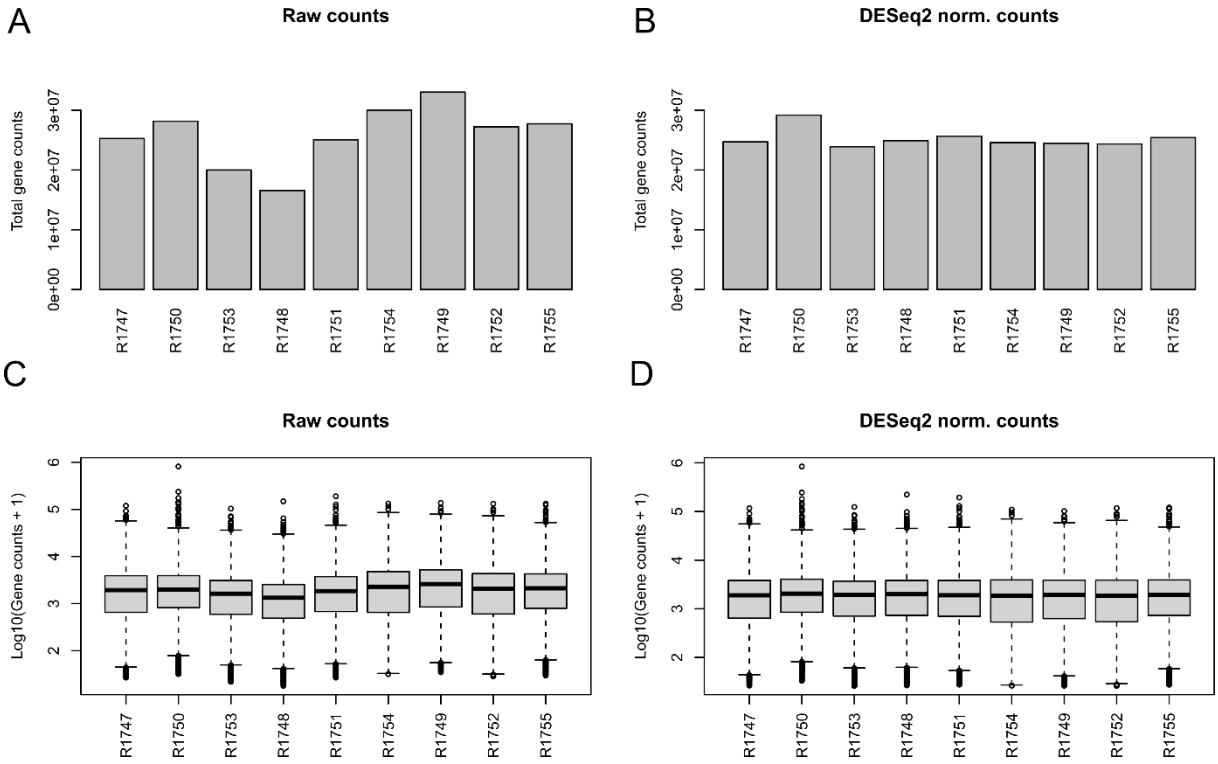
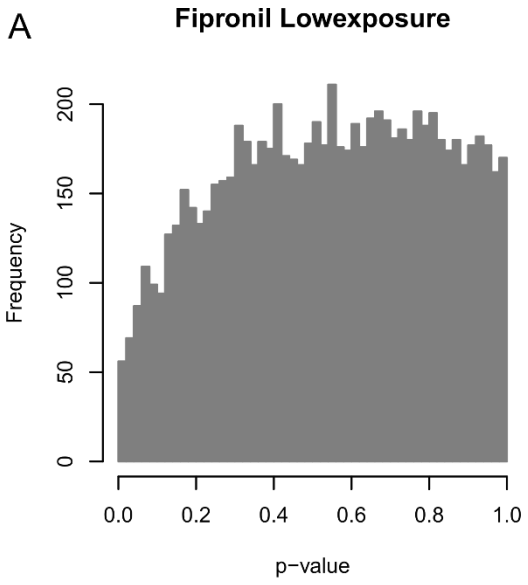
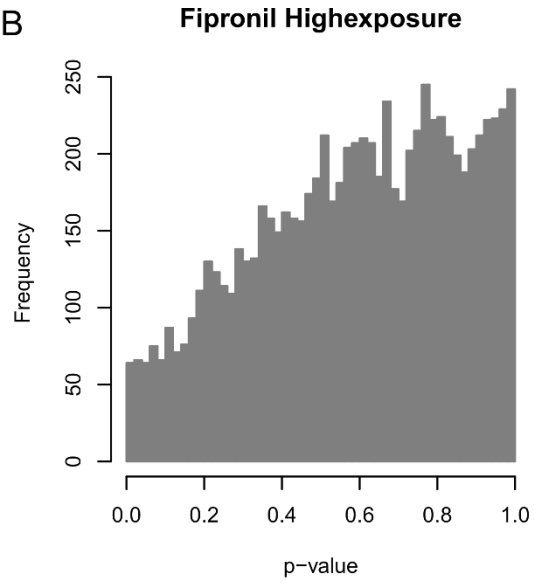


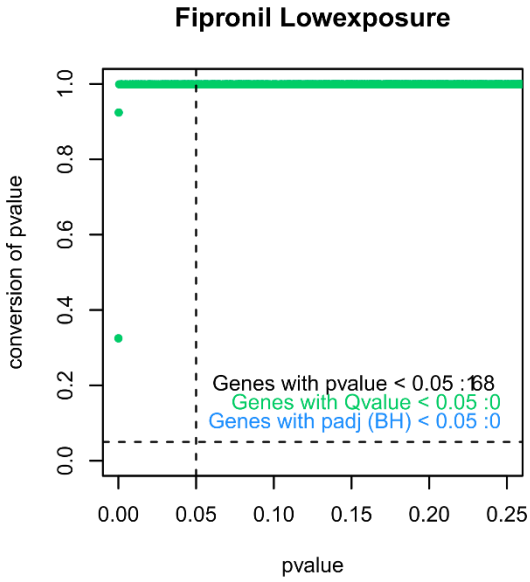
Figure S4: RNA-Seq read count normalisation using DESeq2. (A) Raw read counts (left) and relative log expression (RLE) normalised read counts (right) of adult *C. dipterum* exposed to low (75.0 ng/L) and high (300 ng/L) concentrations of fipronil and corresponding controls. Biological replicates are numbered as shown.

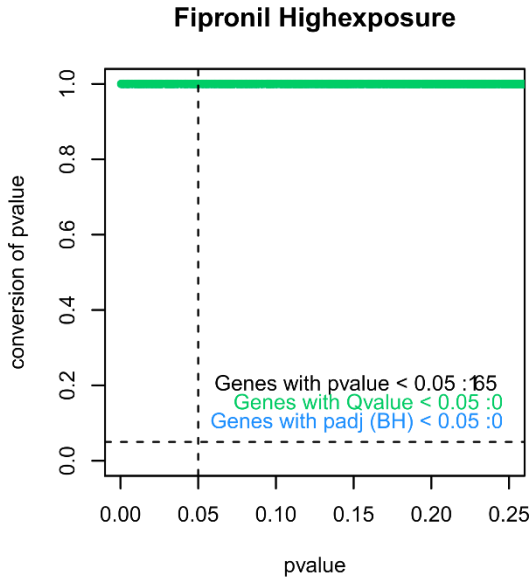
A

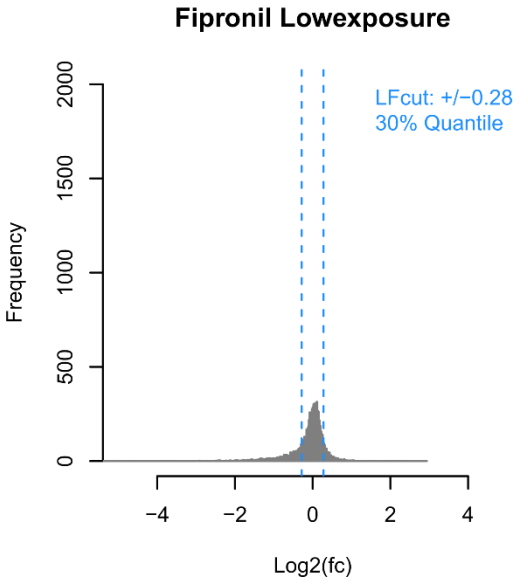

Fipronil Lowexposure
Frequency
0
50
100
150
200
0.0
0.2
0.4
0.6
0.8
1.0
p−value

B


Fipronil Highexposure
Frequency
0
50
100
150
200
250
0.0
0.2
0.4
0.6
0.8
1.0
p−value


Fipronil Lowexposure
conversion of pvalue
0.0
0.2
0.4
0.6
0.8
1.0
Genes with pvalue < 0.05 :168
Genes with Qvalue < 0.05 :0
Genes with padj (BH) < 0.05 :0
0.00
0.05
0.10
0.15
0.20
0.25
pvalue


Fipronil Highexposure
conversion of pvalue
0.0
0.2
0.4
0.6
0.8
1.0
Genes with pvalue < 0.05 :165
Genes with Qvalue < 0.05 :0
Genes with padj (BH) < 0.05 :0
0.00
0.05
0.10
0.15
0.20
0.25
pvalue


Fipronil Lowexposure
Frequency
0
500
1000
1500
2000
LFcut: +/−0.28
30% Quantile
−4
−2
0
2
4
Log2(fc)

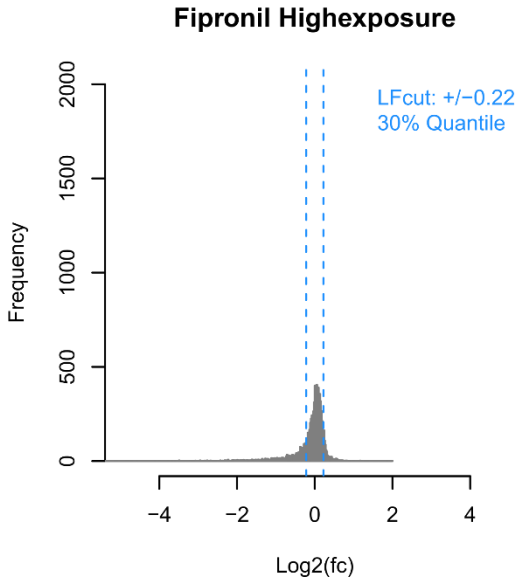

Fipronil Highexposure
Frequency
0
500
1000
1500
2000
LFcut: +/−0.22
30% Quantile
−4
−2
0
2
4
Log2(fc)

Figure S5: Distributions of p-values, p-value conversion, and distributions of log(2)-fold changes upon low and high levels of fipronil exposure in adult *C. dipterum* observed by gene expression data in contrast to the control. Distribution of p-values for all genes after exposure to (A) low (75.0 ng/L) and (B) high (300 ng/L) fipronil concentrations (Wald's t-test) (top). Wald's p-values were obtained and compared to converted p-values for multiple testing using Benjamini-Hochberg, as well as the corresponding p-values for exposure to low (75.0 ng/L) and high (300 ng/L) concentrations of fipronil (Middle). The distribution of log(2)-fold change values for all genes following fipronil exposure at low (75.0 ng/L) and high (300 ng/L) concentrations (bottom). For each case, the dotted line denotes the log2-fold change cut-off.

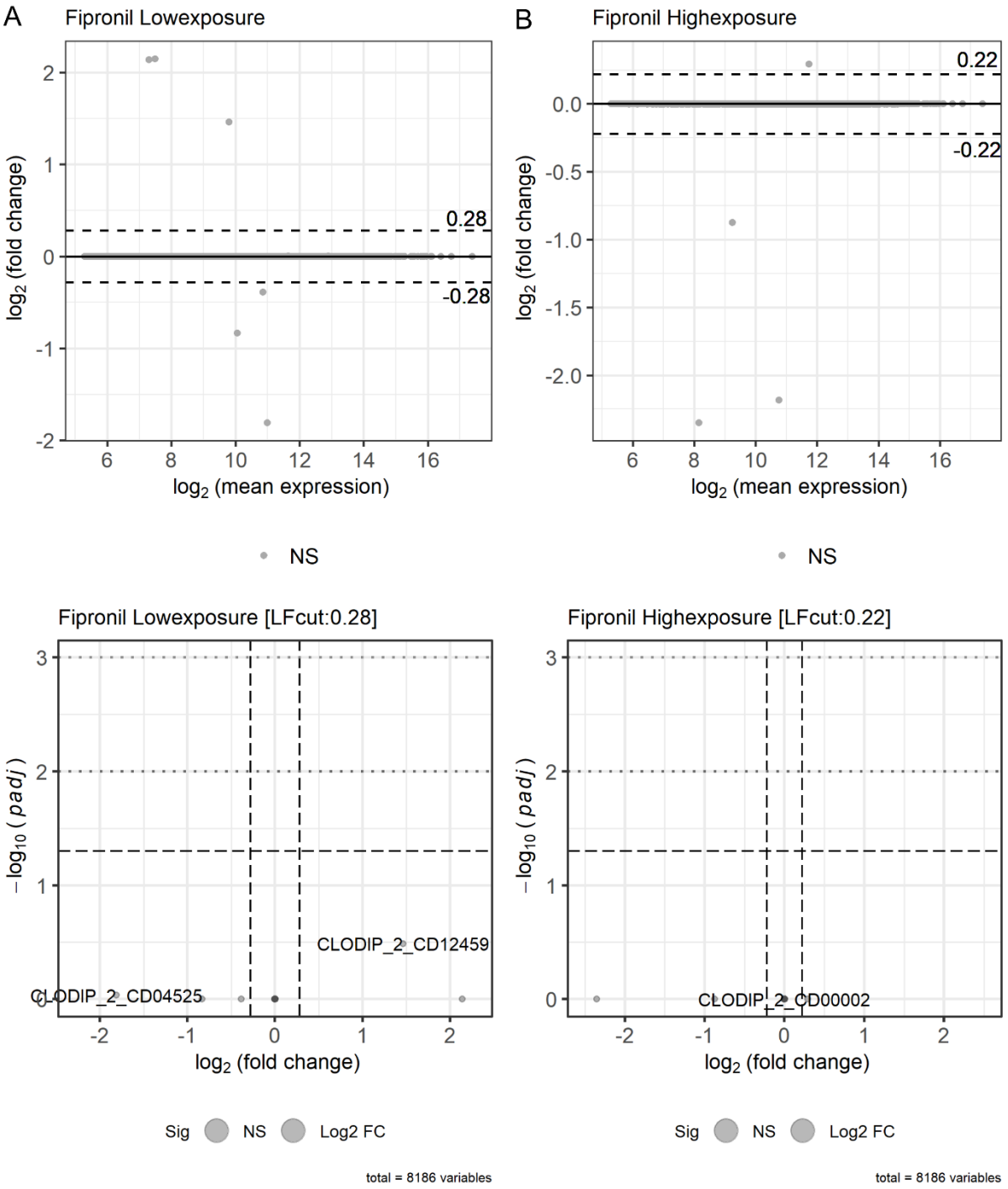


Figure S6: MA-plot (top) showing log(2) transformed mean expression (gene count) to apeglm shrunk log(2)-fold change for every expressed gene in (A) Low and (B) High exposure condition in comparison to the control and Volcano plot (bottom) depicts the log(2)-fold change (lfc) values

against the corresponding –log(10)(padj) values of genes that were differentially expressed after (A) Low and (B) High fipronil exposure to adult *C. dipterum*. The lfc value cut-off as well as the padj cut-off are indicated as dotted lines. Genes applying to both of them are coloured red. Black dots represent genes with no significant difference.

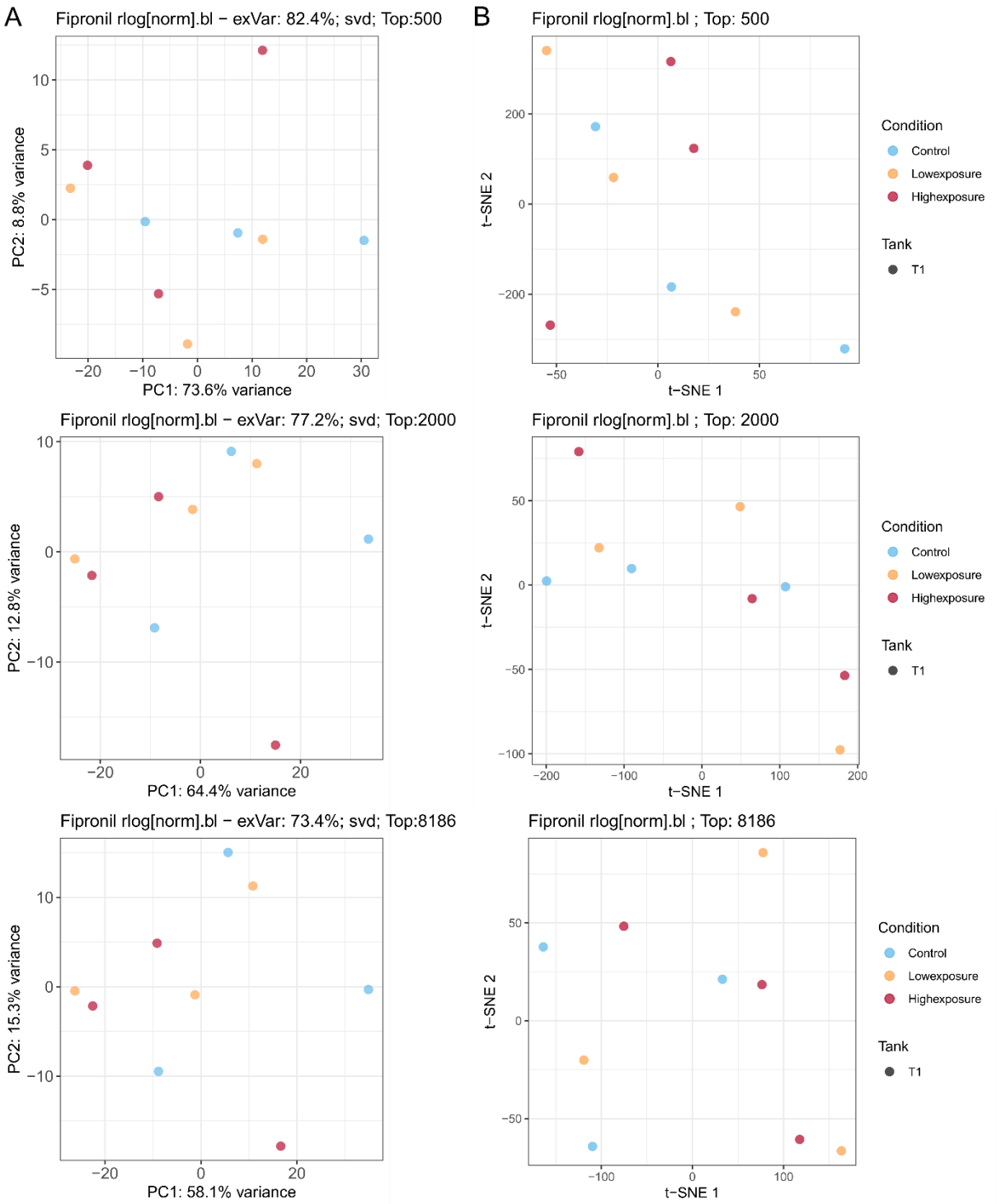


Figure S7: PCA (A) and t-distributed stochastic neighbour embedding (t-SNE) (B) of adult *C. dipterum* exposure to low (75.0 ng/L) and high (300 ng/L) concentrations of Fipronil, as showed by RNA-Seq. On top of each panel, the total number of normalized genes used for clustering is displayed.

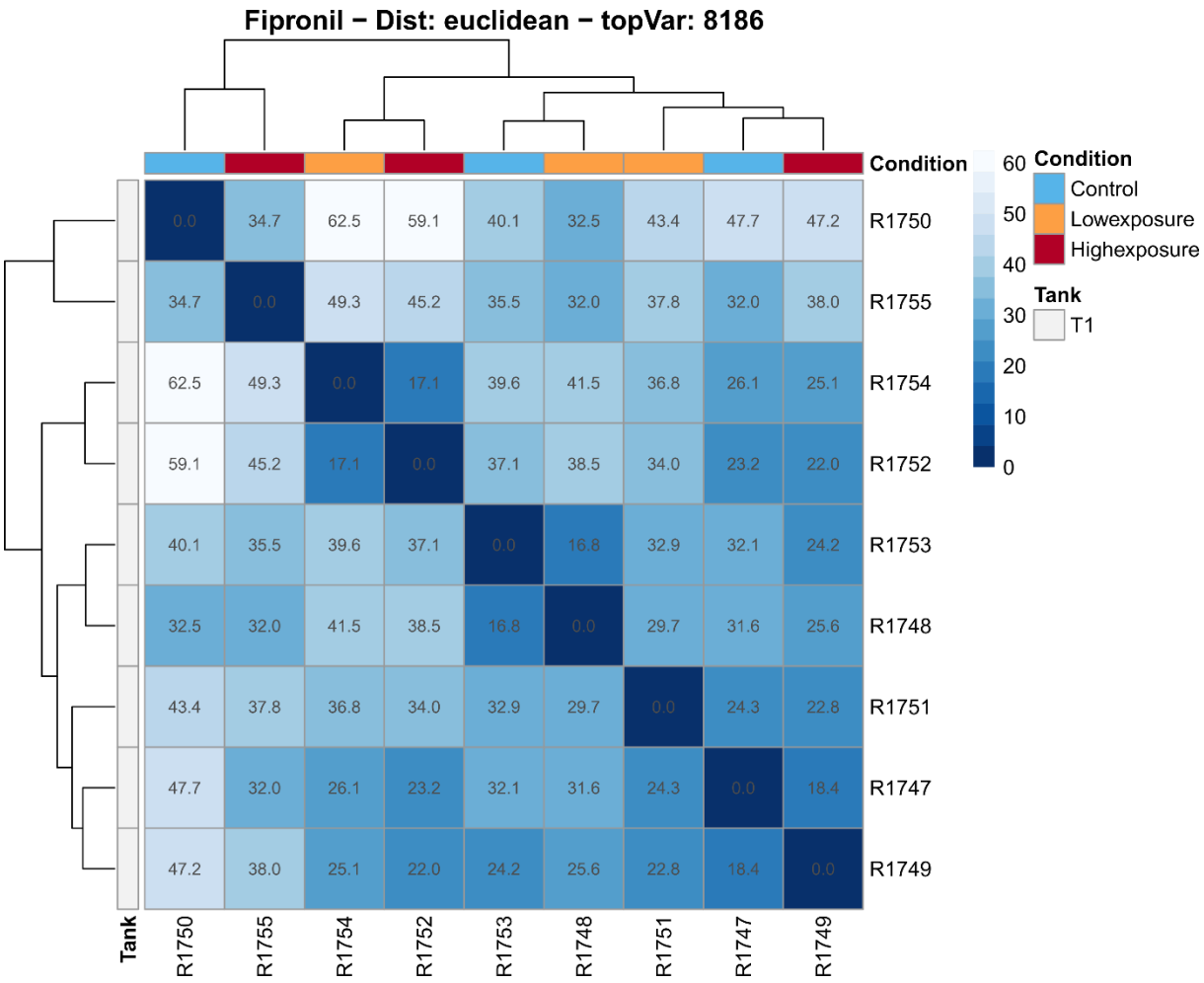


Figure S8: Sample distance matrix analysis of adult *C. dipterum* exposed to Fipronil at low (75.0 ng/L) and high (300 ng/L) concentrations, as well as the corresponding controls. The conditions are denoted by a color code.

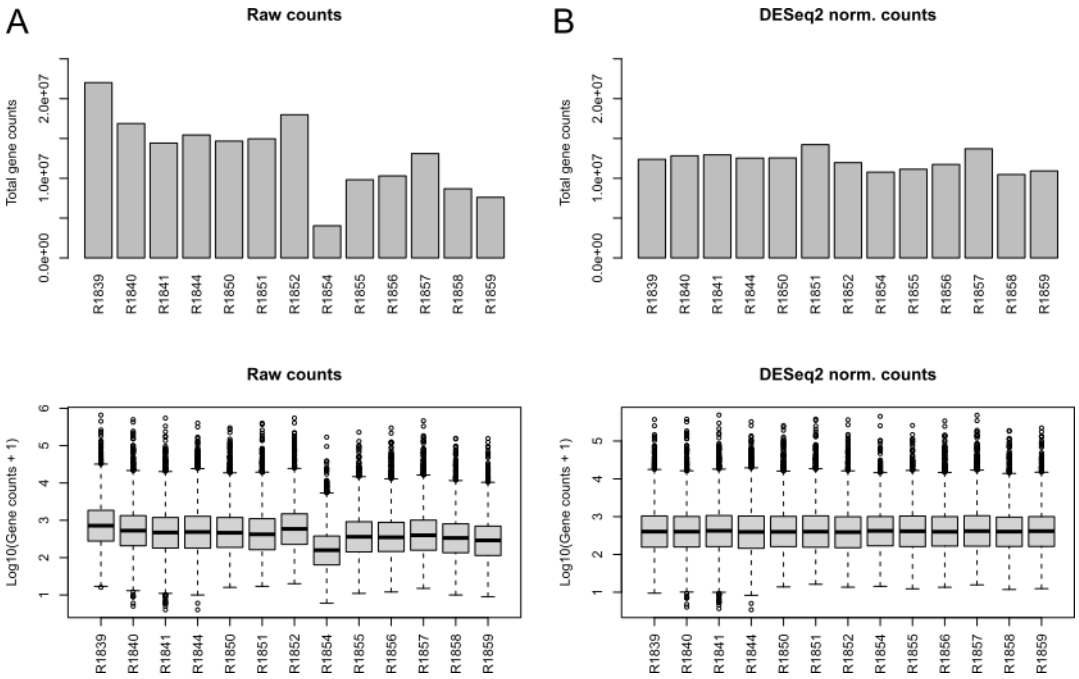


Figure S9: RNA-Seq read count normalization using DESeq2. (A) Raw reads (left) and Relative log Expression (RLE) normalized read counts (right) of *C. dipterum* larvae exposed to low (75.0 ng/L), mid (150 ng/L), and high (300 ng/L) concentrations of fipronil, as well as the corresponding controls. Biological replicates are numbered.

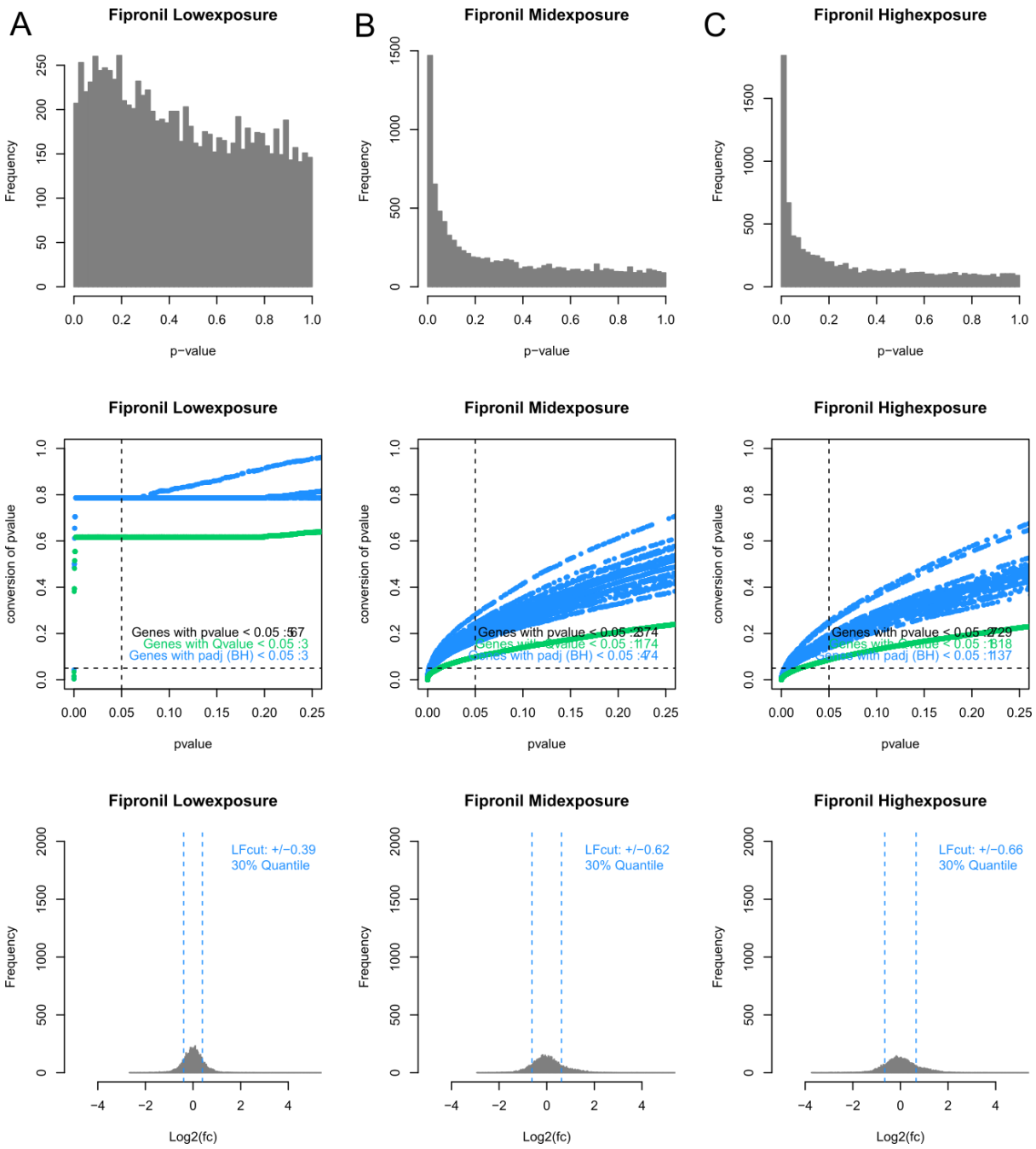


Figure S10: Distributions of p-values, p-value conversion, and distributions of log(2)-fold changes following low, mid, and high exposure of *C. dipterum* larvae to fipronil as observed by gene expression data in comparison to the control. Distribution of p-values for all genes following exposure to (A) low (75.0 ng/L), (B) mid (150 ng/L), and (C) high (300 ng/L) concentrations of fipronil (Wald's t-test) (top). Obtained Wald's p-values in comparison to converted p-values for multiple testing following Benjamini-Hochberg, as well as the corresponding p-values for exposure to low (75.0 ng/L), mid (150 ng/L), and high (300 ng/L) concentrations of fipronil (Middle). Distribution of log(2)-fold change values of all genes after exposure to fipronil at low (75.0 ng/L), mid (150 ng/L), and high (300 ng/L) concentrations (bottom). The dotted line represents the log(2)-fold change cut-off for each condition.

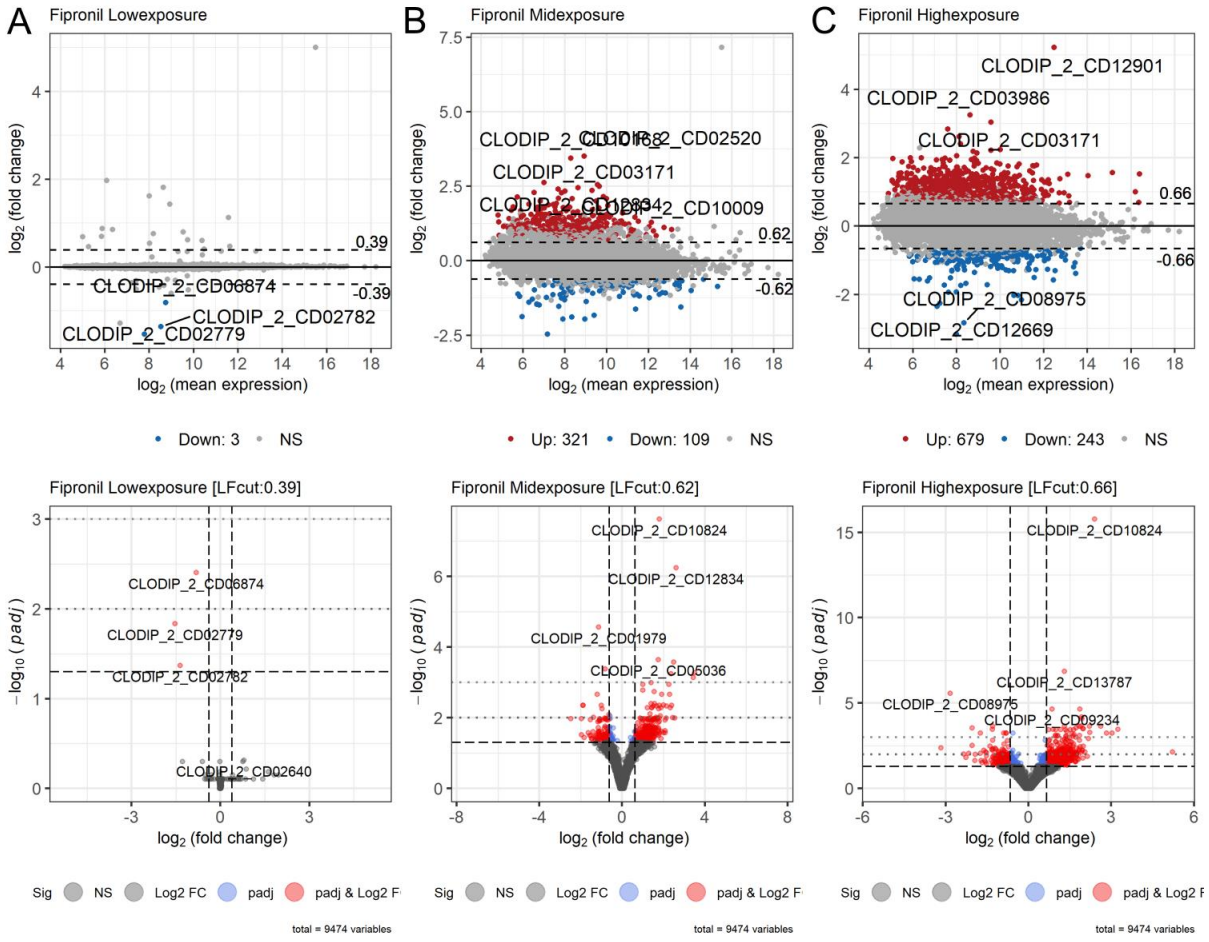


Figure S11: MA-plot (top) showing log(2) transformed mean expression (gene count) to apeglm shrunk log(2)-fold change for every expressed gene in (A) Low, (B) Mid and (C) High exposure condition in comparison to the control and Volcano plot (bottom) depicts the log(2)-fold change (lfc) values against the corresponding –log(10)(padj) values of genes that were differentially expressed after (A) Low, (B) Mid and (C) High fipronil exposure to *C. dipterum* larvae. The lfc value cut-off as well as the padj cut-off are indicated as dotted lines. Genes applying to both of them are coloured red. Black dots represent genes with no significant difference.

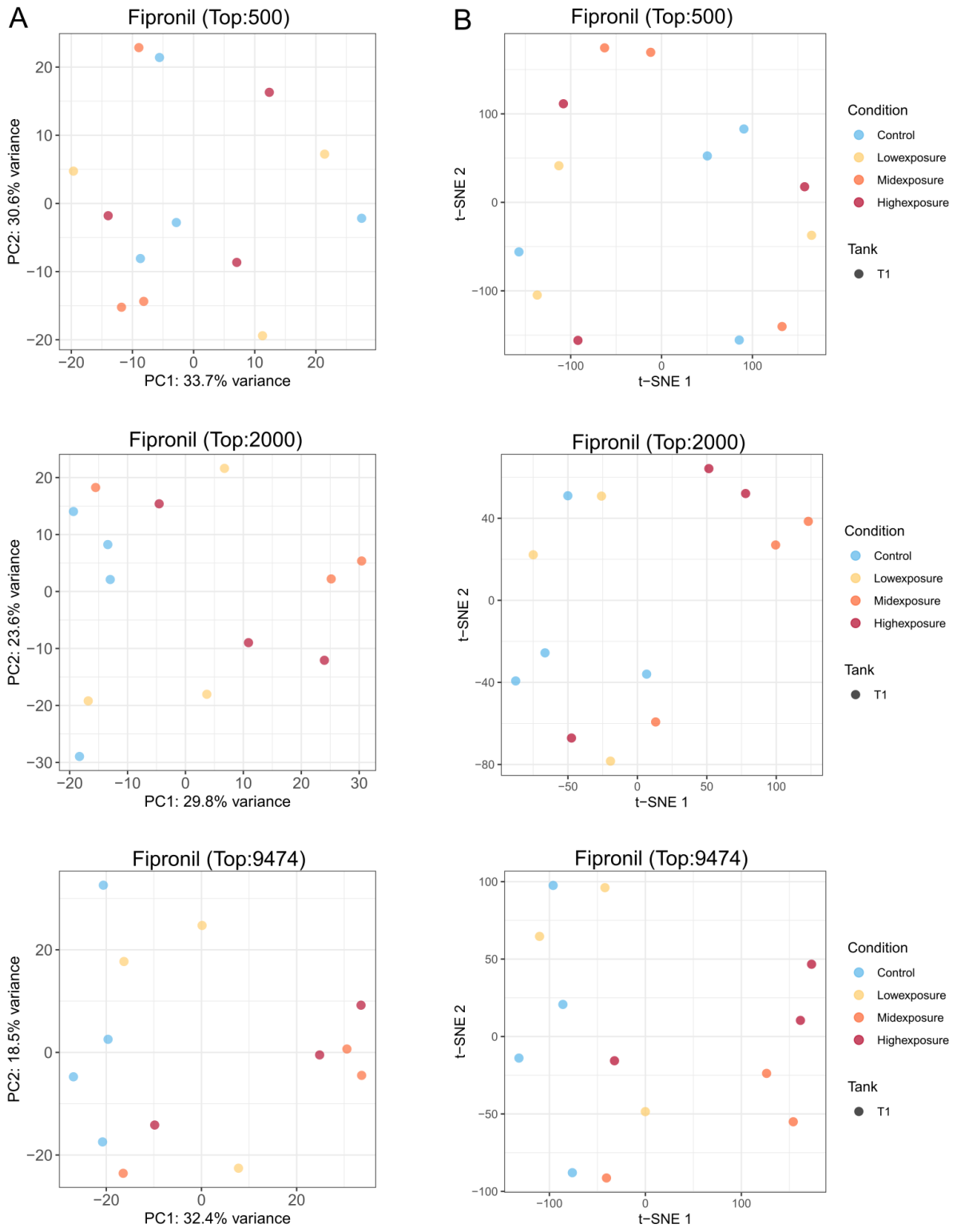


Figure S12: PCA (A) and t-distributed stochastic neighbour embedding (t-SNE) (B) of samples upon exposure to low (75.0 ng/L), mid (150 ng/L), and high (300 ng/L) concentrations of fipronil to *C. dipterum* larvae as showed by RNA-Seq. On top of each panel, the total number of normalized genes used for clustering is displayed.

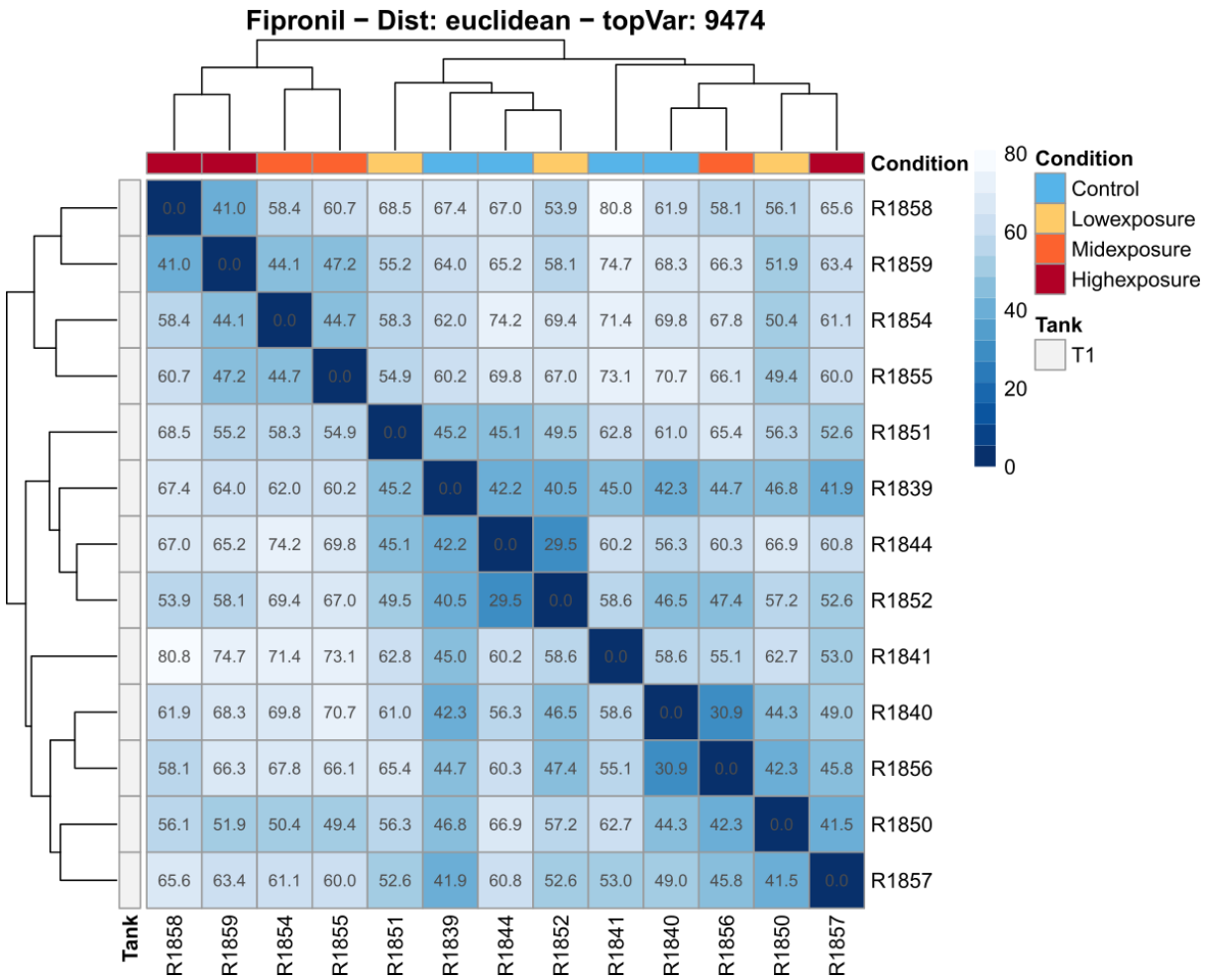


Figure S13: Sample distance matrix analysis of *C. dipterum* larvae samples exposed to fipronil at low (75.0 ng/L), mid (150 ng/L), and high (300 ng/L) concentrations, as well as the corresponding controls. The conditions are denoted by a color code.

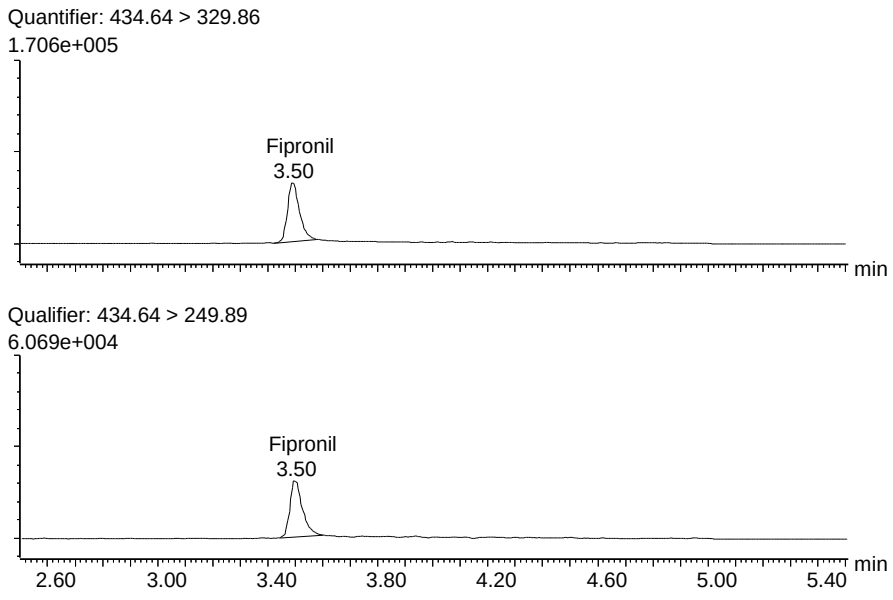


Figure S14: Chromatogram of fipronil (quantifier and qualifier mass transition) of the control sample at day 21 (fresh) of the chronic test. The measured concentration in this sample was 0.018 µg/L (<LOQ). The y-axis was scaled to the conc. 1 sample at the same day (see Figure S 4).

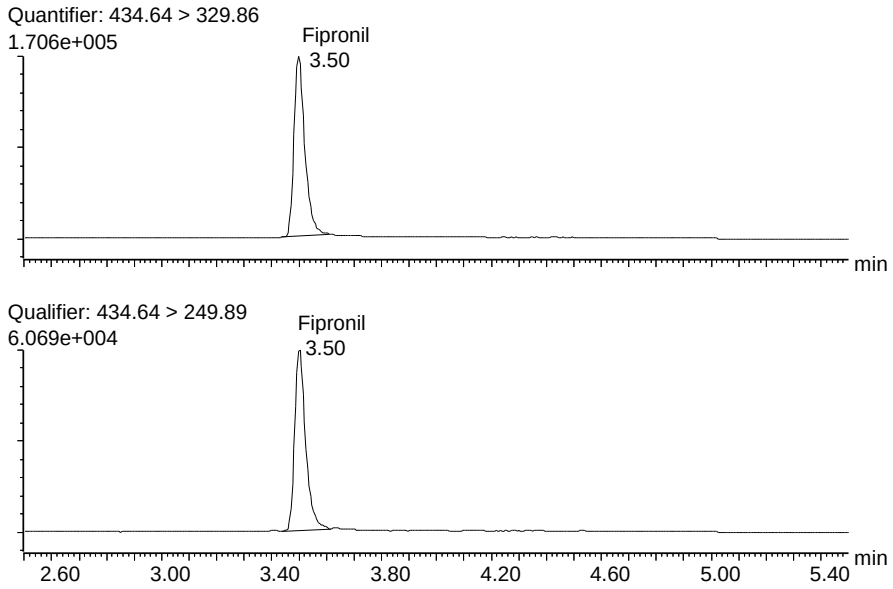


Figure S15: Chromatogram of fipronil (quantifier and qualifier mass transition) of the concentration level 1 sample at day 21 (fresh) of the chronic test. The measured concentration in this sample was 0.052 µg/L.

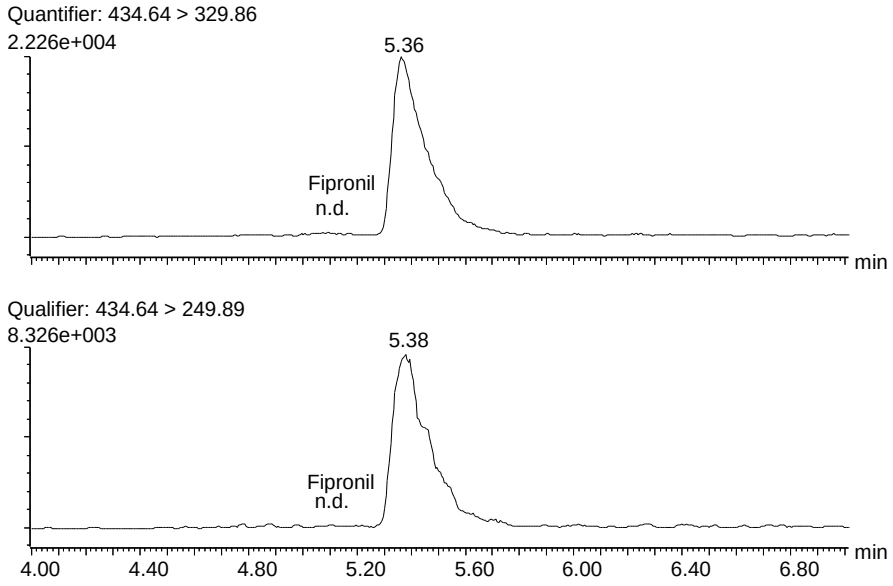


Figure S16: Chromatogram of fipronil (quantifier and qualifier mass transition) of the control sample at day 21 (fresh) of the short term exposure test. No peak was detected (n.d.). The y-axis was scaled to the conc. 1 sample at the same day (see Figure S 6).

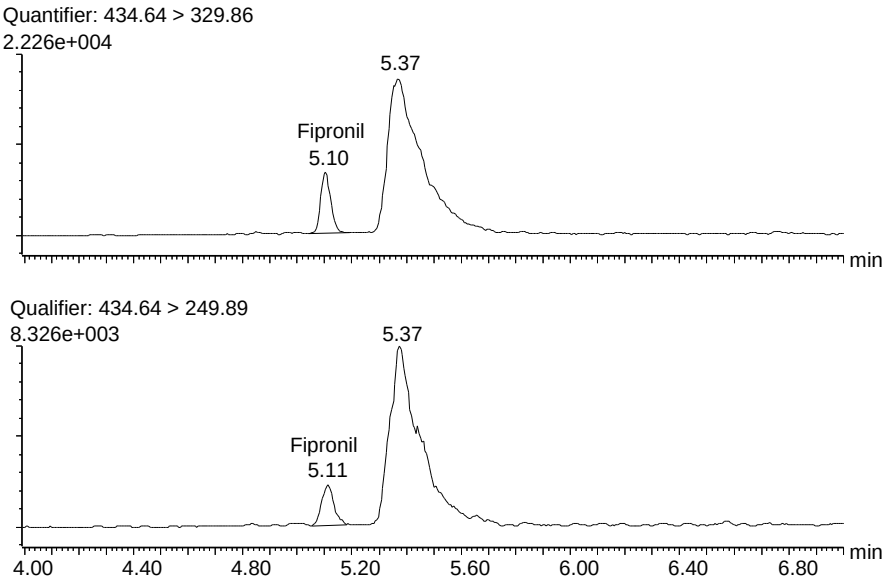


Figure S17: Chromatogram of fipronil (quantifier and qualifier mass transition) of the concentration level 1 sample at day 21 (fresh) of the short term exposure test. The measured concentration in this sample was 0.027 µg/L.

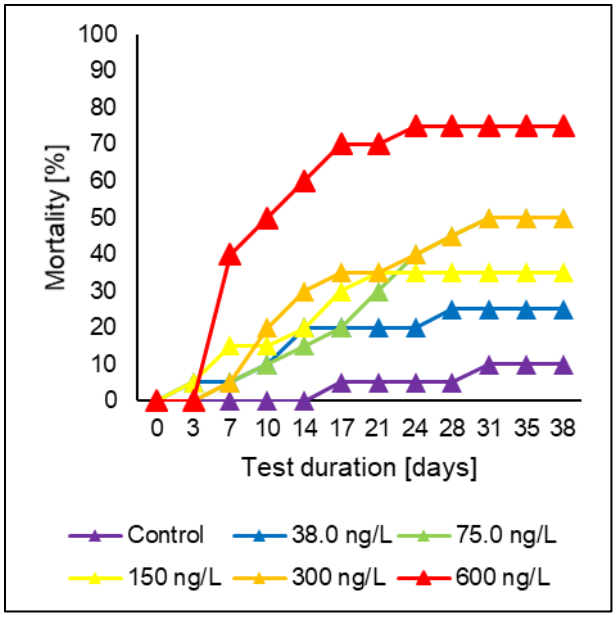


Figure S18: Mortality induced by Fipronil in *C. dipterum* per treatment during a long-term exposure test with a total duration of 38 days.

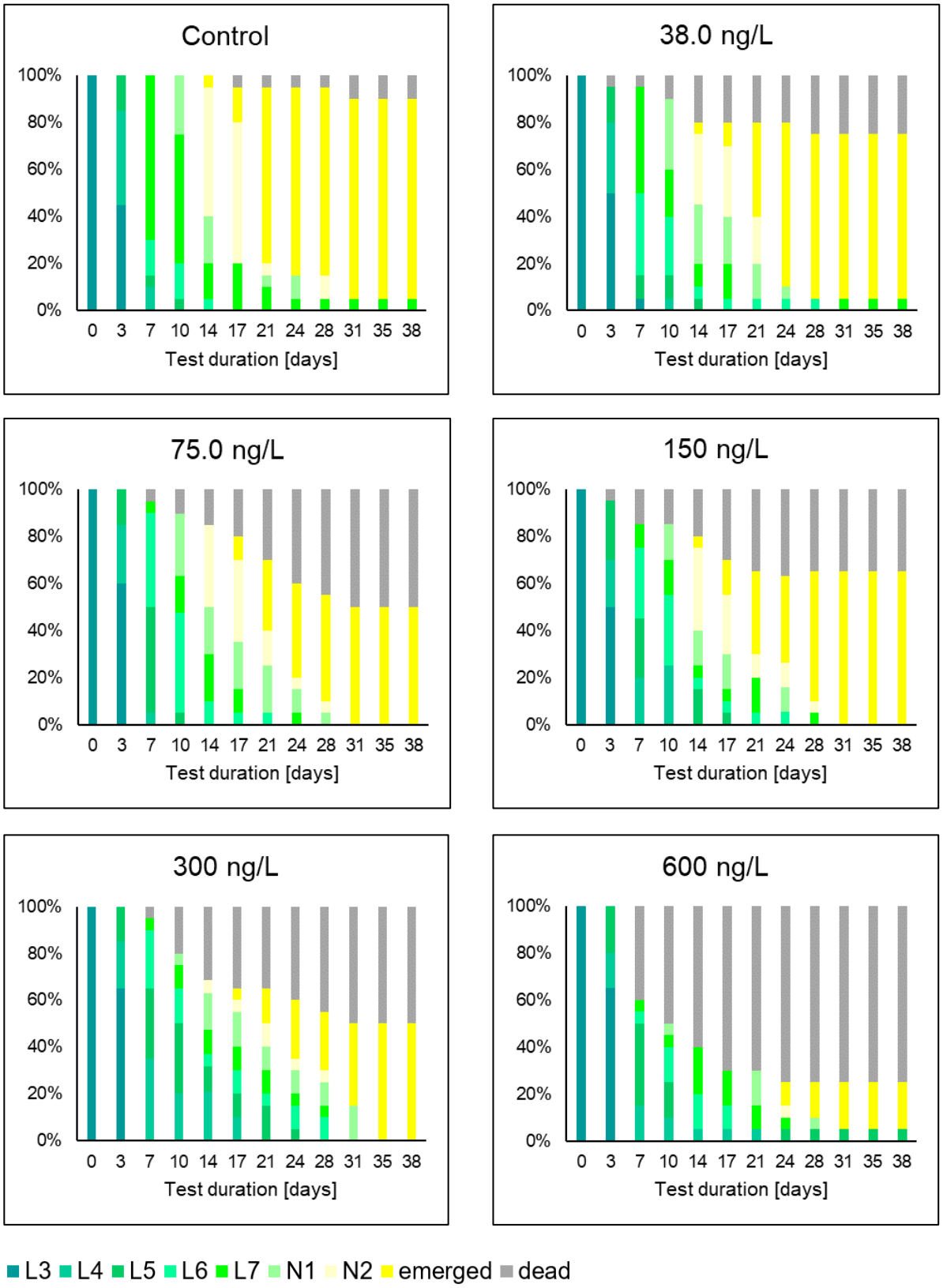


Figure S19: Larval development in each treatment for the test duration of 38 days in the long-term exposure test.

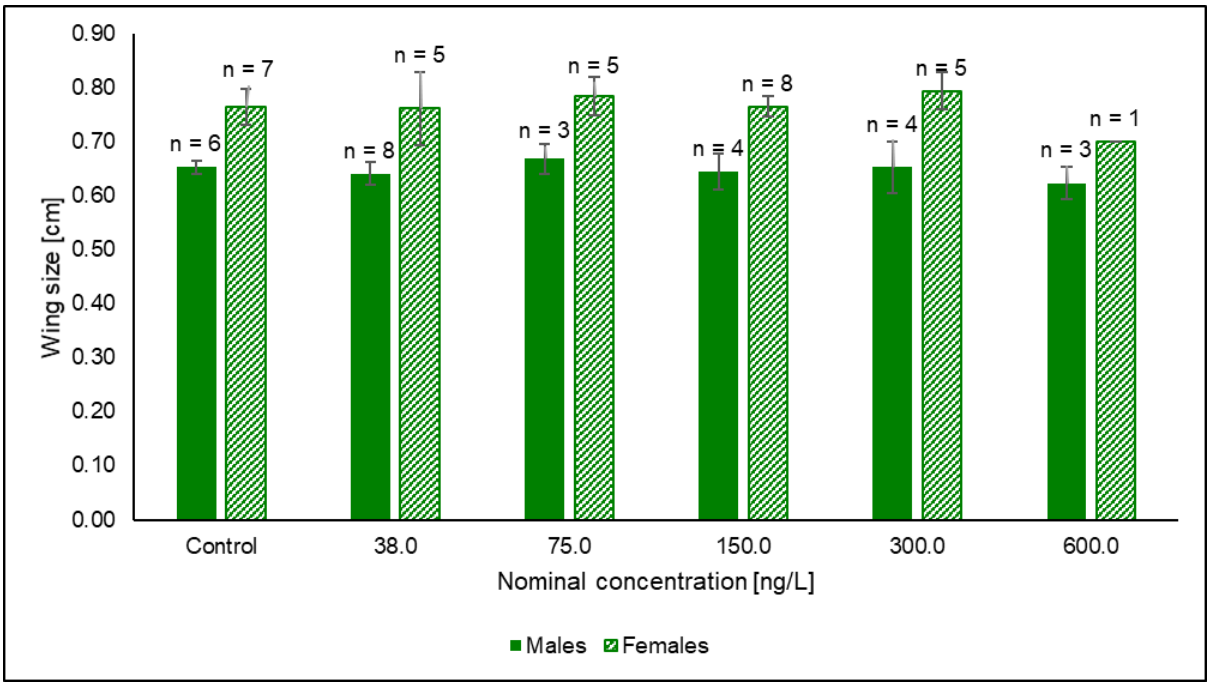


Figure S20: Wing size of *C. dipterum* imagines per treatment during in the long-term exposure test with a total duration of 38 days.

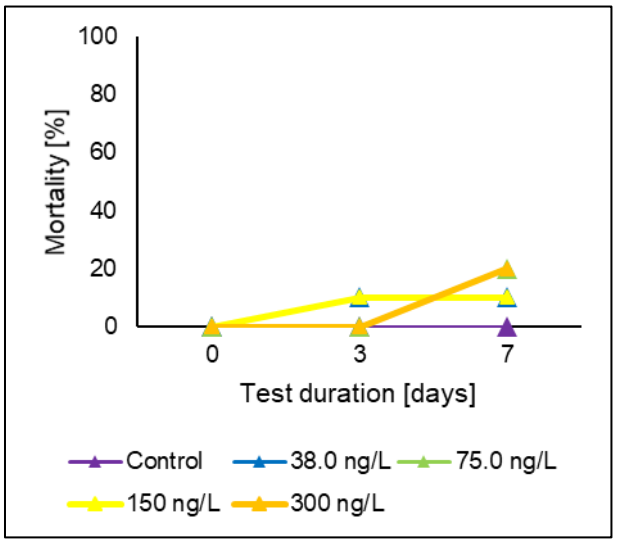


Figure S21: Mortality induced by fipronil in *C. dipterum* per treatment during a short-term exposure test with a total duration of 7 days.

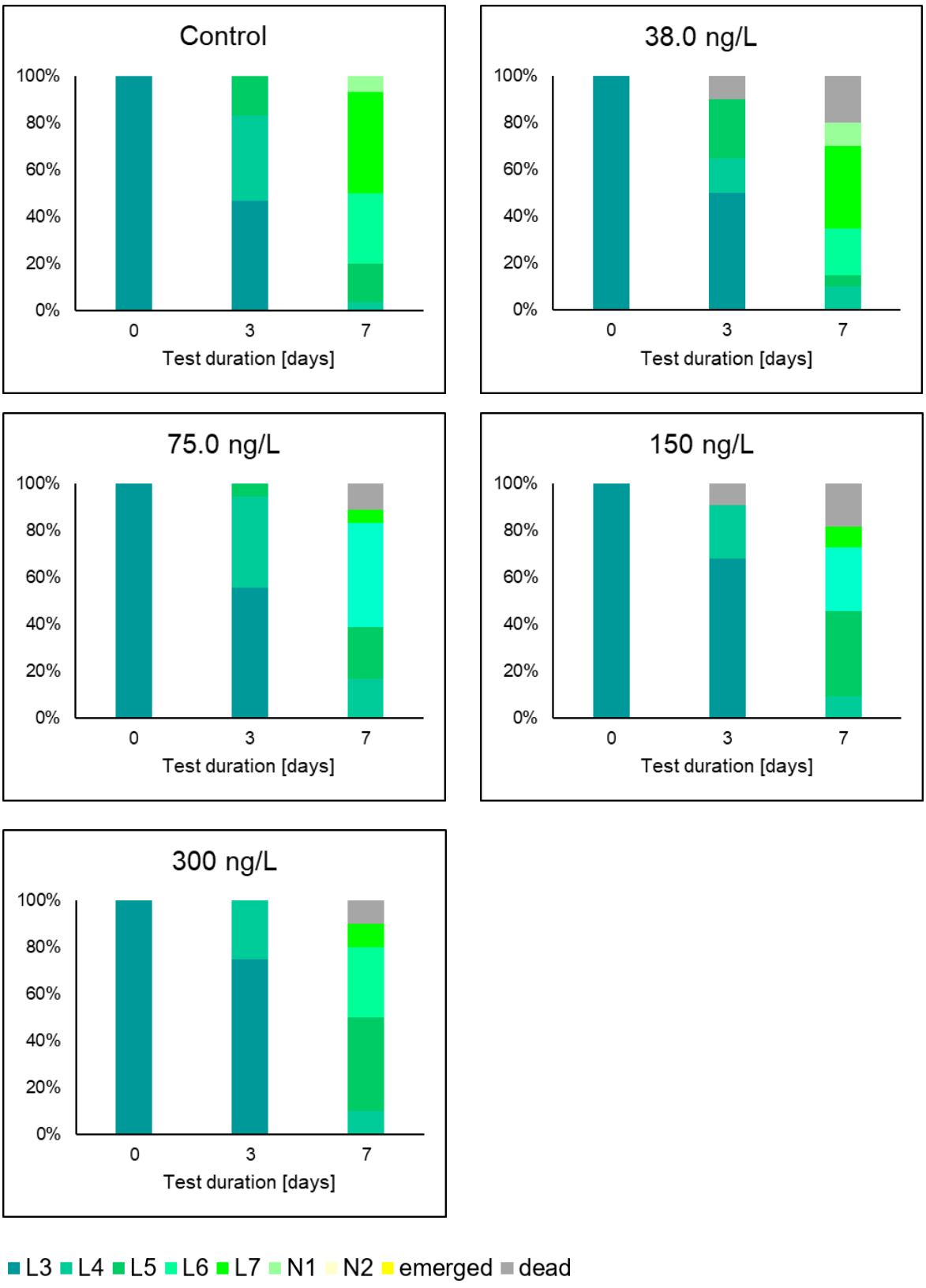


Figure S22: Larval development in each treatment for the test duration of 7 days in the short-term exposure test.

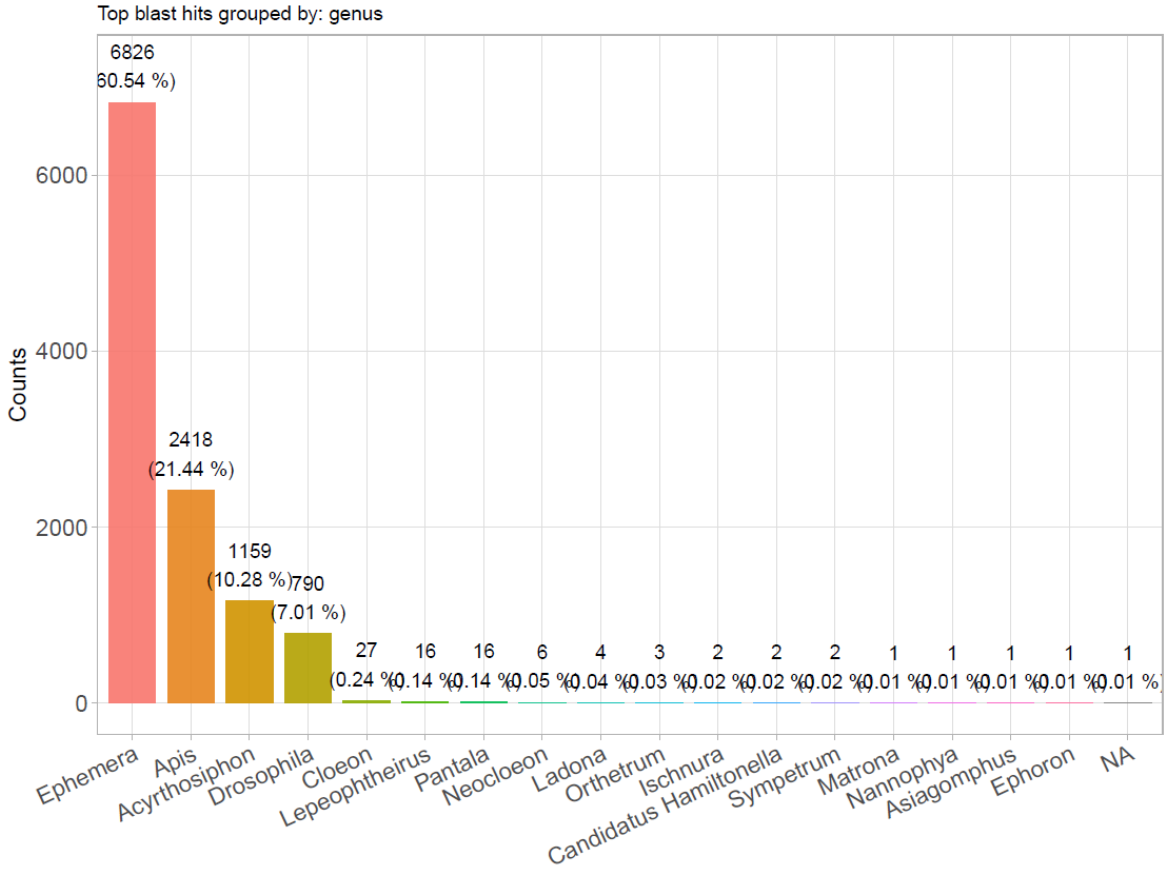


Figure S23: Total blastp top hit search results per genus illustrating the top hits distribution in relation to the *C. dipterum* gene IDs from the reference genome.

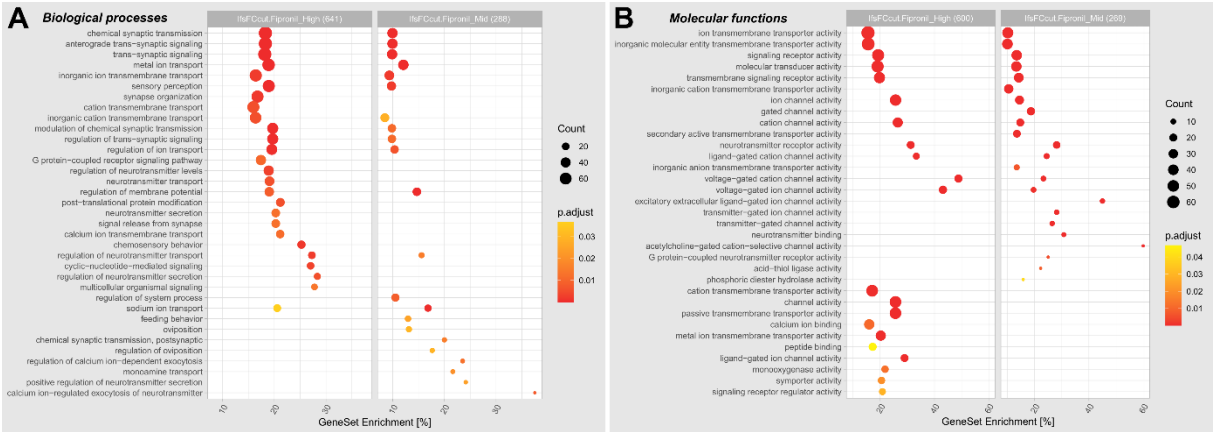


Figure S24: Overrepresentation analysis of significantlly regulated genes (A) Biological processes (B) Molecular functions following the sublethal exposure exposure of *C. dipterum* larvae to fipronil. The log(2)-converted gene count for each ontology term is represented by the size of the bubble, and the geneset enrichment (%) is plotted on the x-axis. For biological processes that are statistically significantly enriched (BH adjusted p<0.05), bubbles are drawn. Color codes are assigned to adjusted p-values.

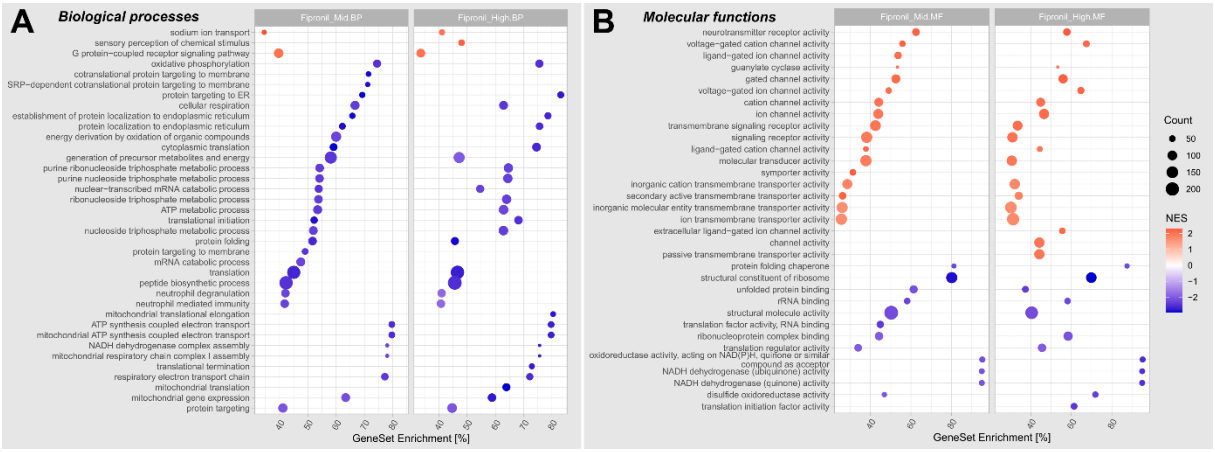


Figure S25: Gene set enrichment analysis of all identified genes (A) Biological processes (B) Molecular functions following sublethal exposure of *C. dipterum* larvae to fipronil. The size of the bubble represents the log(2)-converted gene count for each ontology term, and the x-axis plots the geneset enrichment (%). The colors of the bubbles represent a positive (red) or negative (blue) normalized enrichment score (NES). Geneset enrichment is a percentage measure of how many genes in a given gene set were enriched relative to set size (GeneSet = (Count / gene set size) * 100).

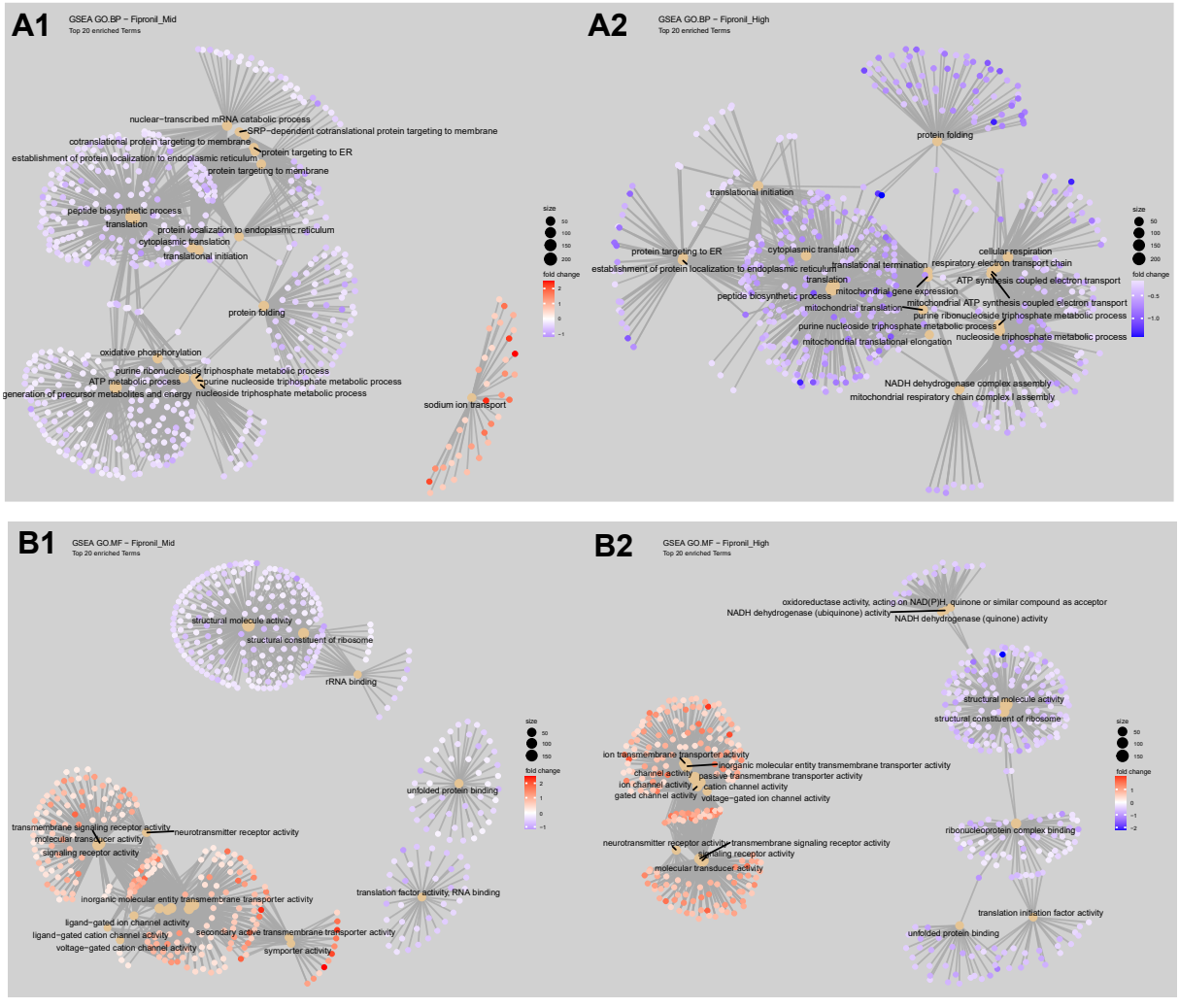


Figure S26: Gene network cluster showing the top 20, most significantly (padj BH ≤ 0.05) enriched (A) Biological processes (B) Molecular functions GO terms from gene set enrichment analysis

(GSEA). As input for the GSEA, substance specific gene sets were selected based on the shared DEGs between mid (150 ng/L) and high (300 ng/L) exposure condition. GO term bubble size reflects the number of genes from the input gene set associated with this biological process. Red indicates an enhanced and blue a suppressed expression relative to the global mean expression of a gene linked with the corresponding GO terms.

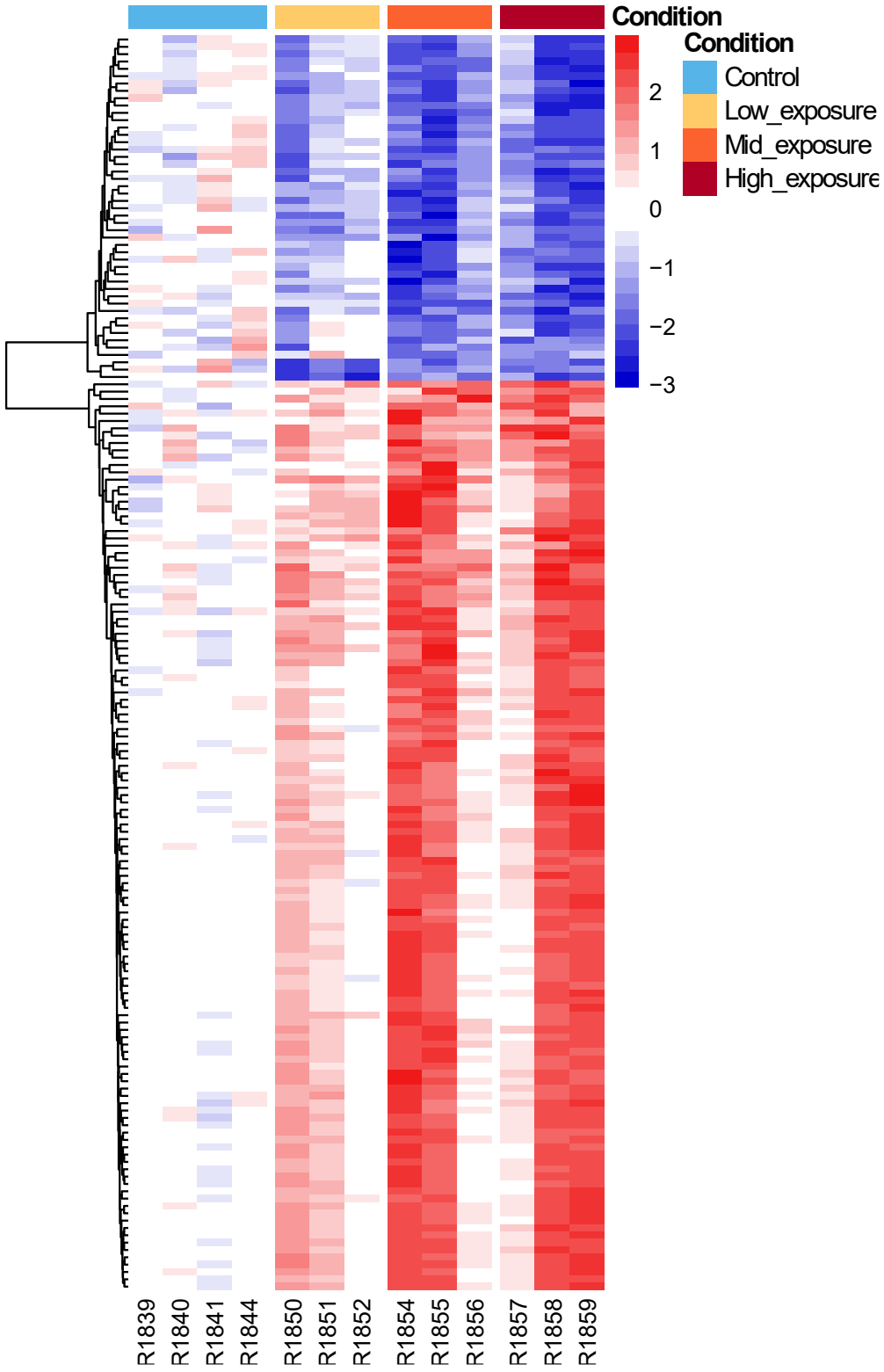


Figure S27: Heatmap illustrating the relative expression patterns of the common set of ME and HE DEGs ranked by their abs(lfc) for genes above 10 CPM (counts per million reads). Relative to the global control mean expression of a gene, red indicates up-regulation and blue indicates down-regulation. Gene clustering was performed using Euclidean distance. The colour code above each column describes the assignment of the sample to the respective exposure condition.

4 SUPPLEMENTAL REFERENCES

Athar, Awais; Füllgrabe, Anja; George, Nancy; Iqbal, Haider; Huerta, Laura; Ali, Ahmed et al. (2019): ArrayExpress update - from bulk to single-cell expression data. In Nucleic acids research 47 (D1), D711-D715. DOI: 10.1093/nar/gky964.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. https://doi.org/10.1186/1471-2105-10-421

Cantalapiedra, Carlos P.; Hernández-Plaza, Ana; Letunic, Ivica; Bork, Peer; Huerta-Cepas, Jaime (2021): eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. In Molecular biology and evolution 38 (12), pp. 5825–5829. DOI: 10.1093/molbev/msab293.

Chen, Shifu; Zhou, Yanqing; Chen, Yaru; Gu, Jia (2018): fastp: an ultra-fast all-in-one FASTQ preprocessor. In Bioinformatics (Oxford, England) 34 (17), i884-i890. DOI: 10.1093/bioinformatics/bty560.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., von Mering, C., Bork, P., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research 47, D309–D314. https://doi.org/10.1093/nar/gky1085

Ignatiadis, Nikolaos; Klaus, Bernd; Zaugg, Judith B.; Huber, Wolfgang (2016): Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. In Nature methods 13 (7), pp. 577–580. DOI: 10.1038/nmeth.3885.

Loll, A., Reinwald, H., Ayobahan, S.U., Göckener, B., Salinas, G., Schäfers, C., Schlich, K., Hamscher, G., Eilebrecht, S., 2022. Short-Term Test for Toxicogenomic Analysis of Ecotoxic Modes of Action in Lemna minor. Environ. Sci. Technol. 56, 11504–11515. https://doi.org/10.1021/acs.est.2c01777

Pagès, H., Carlson, M., Falcon, S., Li, N., 2023. AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. https://doi.org/10.18129/B9.bioc.AnnotationDbi

Reinwald, H., 2023. Cloeon dipterum annotation R package (org.Cdipterum.eg.db) and respective files for the custom built. https://doi.org/10.5281/zenodo.8003805

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16, 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Sayers, Eric W.; Beck, Jeffrey; Bolton, Evan E.; Bourexis, Devon; Brister, James R.; Canese, Kathi et al. (2021): Database resources of the National Center for Biotechnology Information. In Nucleic acids research 49 (D1), D10-D17. DOI: 10.1093/nar/gkaa892.

The Darwin Tree of Life Project Consortium, 2022. Sequence locally, think globally: The Darwin Tree of Life Project. Proceedings of the National Academy of Sciences 119, e2115642118. https://doi.org/10.1073/pnas.2115642118

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28, 511–515. https://doi.org/10.1038/nbt.1621

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., Yu, G., 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. The Innovation 2, 100141. https://doi.org/10.1016/j.xinn.2021.100141

Zhu, Anqi; Ibrahim, Joseph G.; Love, Michael I. (2019): Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. In Bioinformatics (Oxford, England) 35 (12), pp. 2084–2092. DOI: 10.1093/bioinformatics/bty895