

# SCIENTIFIC REPORTS



OPEN

## An integrative omics approach to unravel toxicity mechanisms of environmental chemicals: effects of a formulated herbicide

Tiago Simões<sup>1,2,3</sup>, Sara C. Novais<sup>1,3</sup>, Tiago Natal-da-Luz<sup>2</sup>, Bart Devreese<sup>4</sup>, Tjalf de Boer<sup>3</sup>, Dick Roelofs<sup>3</sup>, José P. Sousa<sup>2</sup>, Nico M. van Straalen<sup>3</sup> & Marco F. L. Lemos<sup>1</sup>

The use of integrative molecular approaches can aid in a comprehensive understanding of the effects of toxicants at different levels of biological organization, also supporting risk assessment. The present study aims to unravel the toxicity mechanisms of a widely used herbicide to the arthropod *Folsomia candida* exposed in a natural soil, by linking effects on reproduction, proteomics and genome-wide gene expression. The EC<sub>50</sub> effects on reproduction over 4 weeks was 4.63 mg glyphosate/kg of soil. The formulation included a polyethoxylated tallowamine as an adjuvant, which at 50% effect on reproduction had an estimated concentration of 0.87–1.49 mg/kg of soil. No effects were observed on survival and reproduction when using the isolated active substance, pointing the toxicity of the formulated product to the co-formulant instead of the active ingredient, glyphosate. RNA sequencing and shotgun proteomics were applied to assess differential transcript and protein expressions between exposed and control organisms in time, respectively. Specific functional categories at protein and transcriptome levels were concordant with each other, despite overall limited correlations between datasets. The exposure to this formulation affected normal cellular respiration and lipid metabolism, inducing oxidative stress and leading to impairment in biological life cycle mechanisms such as molting and reproduction.

Due to the high complexity of ecosystems, environmental risk assessment can be a very challenging task. There is a need to develop and validate innovative and reliable tools and integrated approaches for fast detection of changes in population and community structures that can be applied by regulatory agencies<sup>1</sup>. The application of omics tools (genomics, transcriptomics, proteomics, and metabolomics) to evaluate environmental health status is becoming increasingly important in environmental management and is helpful for risk assessment. Omics may allow connecting molecular events and adverse outcomes with biological levels of organization relevant for risk assessment schemes<sup>2,3</sup>. The analysis of the abundance of transcripts of RNA<sup>4–7</sup> or more recently proteomics<sup>8–11</sup> have already gained particular interest by helping to better understand the mechanistic interactions between the stressor and the organism at different stress conditions, but also to unravel the modes of action (MoA) of several toxicants.

Due to translational regulatory events, the difference in the average half-life of molecules, or even the physical properties of transcripts<sup>12</sup>, abundance of mRNA does not always correlate with protein abundance (e.g. Gygi *et al.*)<sup>13</sup>, particularly in complex biological samples<sup>14</sup>. However, the analysis of both datasets from the same experiment may provide useful insights that may not be deciphered from individual analysis of gene expression or protein levels.

Collembola (springtails) are a group with high abundance in almost any environment, and *Folsomia candida* has emerged as a standard species and genomic model organism for soil-toxicology<sup>15</sup>. Several studies have addressed the effects of different stressors, such as metals<sup>16,17</sup>, pesticides<sup>18,19</sup> or combined pollutants<sup>20</sup>, in gene

<sup>1</sup>MARE – Marine and Environmental Sciences Centre, ESTM, Polytechnic Institute of Leiria, Peniche, Portugal.

<sup>2</sup>Centre for Functional Ecology, Department of Life Sciences, University of Coimbra, Coimbra, Portugal. <sup>3</sup>Institute of Ecological Sciences, Vrije University, Amsterdam, Netherlands. <sup>4</sup>Laboratory for Microbiology (LM-Ugent), Ghent University, Ghent, Belgium. Tiago Simões and Sara C. Novais contributed equally to this work. Correspondence and requests for materials should be addressed to T.S. (email: [tiago.simoes@ipleiria.pt](mailto:tiago.simoes@ipleiria.pt))

expression of this collembolan species over the last decade. Since then, considerable advances have been made to assemble a more complete transcriptome of *F. candida*<sup>21</sup> and the complete genome for this species was also very recently made available<sup>22</sup>. However, proteomic information is still lacking. To our knowledge, there are no proteomic studies available for *F. candida* and the present study represents the first insight on its proteome.

In agricultural fields, pesticides are applied as complex commercial formulations, whose active ingredients have commonly relatively well characterized effects on target, as well and some non-target organisms. However, for some herbicide formulations, the added adjuvants confer higher or different routes of toxicity. It is therefore of major importance to also understand the general effects of the applied formulations.

Glyphosate-based herbicides are between the most used and studied formulations around the globe, with their future usage being currently under great debate by the European Union (EU). In a recent study<sup>23</sup>, 317 EU agricultural topsoils from eleven European countries were analyzed and glyphosate and/or its primary metabolite were detected in 45% of the tested samples, with a maximum concentration of 2 mg kg<sup>-1</sup>.

This work primarily intended to better understand the mechanisms of toxicity of a glyphosate-based formulation to the soil ecotoxicological model organism *F. candida*, by linking effects at different levels of biological organization: reproduction, protein levels, and gene expression. In addition, the potential of integrative gene and protein expression analysis as potential early warning indicators to be applied in environmental assessment studies was evaluated.

## Methods

**Test soil, organisms and contaminant.** The test soil was collected from an agricultural area from the lower Mondego valley in Coimbra, Portugal (40°12'44.0"N; 8°27'02.4"W). The area was kept in fallow for over 10 years without use of pesticides. After collection, soil was sieved (5 mm), and defaunated using freezing-thawing cycles, being then kept at 4 °C for a maximum of 1 month. The physicochemical and mineralogical characterization of the natural soil is provided in Supplementary Table S1.

The collembolan *F. candida*, obtained from cultures of the Laboratory of Soil Ecology and Ecotoxicology of the Centre for Functional Ecology, University of Coimbra (Portugal), was used as test species. Cultures were maintained while kept on a wet mixture of plaster of Paris and activated charcoal (11:1, w/w), at 20 ± 2 °C and with a 16:8 h light:dark photoperiod, and fed baker's yeast (Vahiné, McCormick, Italy).

The herbicide formulation Montana<sup>®</sup> was acquired from SAPEC, Portugal, with a declared component composition including glyphosate (CAS 1071-83-6) as the active ingredient (30.8%, a.i.) in the form of isopropyl ammonium salt (C<sub>6</sub>H<sub>17</sub>N<sub>2</sub>O<sub>3</sub>P) and a polyethoxylated tallow amine adjuvant (POEA; CAS 61791-26-2) with concentrations ranging from 7 to 12% (w/w). After spiking the herbicide formulation into soil, glyphosate concentrations in soil samples were determined using LC-MS/MS by an external accredited company (Eurofins Scientific, Portugal), following methanol extraction. The storage time before analysis was approximately 1 month. The POEA adjuvants are very complex mixtures and their quantification in formulations is currently very challenging<sup>24</sup>. Therefore, the spiking calculations were based on the active ingredient (glyphosate) present in the formulation. We used the ratio between glyphosate and POEA in the product label to estimate POEA concentrations mixed into soil at specific effect levels.

**Reproduction tests and exposure for the omics analyses.** To find the concentration causing 50% inhibitory effect on reproduction (EC50), several dilutions of the formulation were prepared in distilled water and tested after mixing them into soil portions, according to nominal concentrations of the active ingredient, to achieve the test concentrations of 0.135; 0.27; 0.54; 1.08; 2.16 and 4.32 mg a.i./kg of dry soil. For toxicity comparison purposes, a parallel reproduction test was conducted with pure glyphosate (Sigma-Aldrich, USA), testing the same concentrations as in the formulation (plus two additional concentrations: 8.64 and 17.28 mg a.i./kg). The reproduction tests were conducted according to ISO guideline 11267<sup>25</sup> recommendations. Briefly, synchronized organisms (10–12 days old) were exposed to the above-mentioned concentrations, plus control, maintaining temperature and photoperiod at 20 ± 2 °C and 16:8 h light:dark. Five replicates per treatment were used, each with 10 organisms in 30 g of soil in a 100 mL glass vial. Organisms were fed after 14 days (2 mg of dry yeast spread in crumbs on the surface) and humidity was verified weekly by adding distilled water to the original weight. After 28 days of exposure, adults and juveniles present in each replicate were floated to the surface, a digital picture was taken and the animals were counted using image analysis software provided by ImageJ (version 1.49). The inhibitory effect on reproduction (EC50) was calculated by fitting a four-parameter logistic dose response curve ( $y = a + (b - a) / (1 + 10^{(\log(c) - x) * d})$ ) through the concentration response data, with “a” being the average juveniles at the highest concentration tested (bottom plateau), “b” the average juveniles at the control treatment (top plateau), “c” being the EC50 and “d” the hill's slope. Fitting was done with Statistica software (version 7, Statsoft, 2004). Data homoscedasticity and normality was observed a posteriori via a residue analysis and a Q-Q plot. For evaluating survival and reproduction statistical effects, and depending on the fulfillment of the assumptions of normality and homoscedasticity, either a one-way analysis of variance (ANOVA) or a Kruskal-Wallis was used, respectively.

For the mechanistic assessment of the effects caused by the formulation, and to better understand the correlations between transcriptomics and proteomics information through time, organisms were then exposed to the EC50 (calculated through the reproduction test), using the formulation (plus control) and sampled at different time-points (4, 7, and 10 days). The exposure was performed with slight modifications from the reproduction tests: four replicates per treatment and time-point were used (24 in total), consisting of vials with 50 g of soil and 75 organisms. The organisms were fed only at the beginning of the exposure period with 2 mg of dry yeast. In parallel, to follow and verify the actual reproduction and mortality effects of the EC50 exposure after 28 days, 5 extra replicates from each treatment were used with 30 g of soil and 10 organisms, following the same ISO guideline recommendations<sup>25</sup>.

**Biological material extraction and isolation.** The total number of organisms recovered from each test vial were pooled and considered a replicate for the omics analyses. After 4, 7 and 10 days of exposure, average organisms recovered per replicate were, respectively,  $70.3 \pm 7.6$ ,  $70.0 \pm 7.5$ ,  $75.0 \pm 0.0$  for control samples and  $72.0 \pm 7.7$ ,  $73.3 \pm 3.7$ ,  $70.2 \pm 7.2$  for contaminated samples. Both RNA and proteins were extracted from the same pools of animals, using the TRIZOL<sup>®</sup> extraction method (Invitrogen, Belgium). After the first RNA precipitation, wash and resuspension steps, total RNA was submitted to a DNase treatment for RNA purification and further resuspended in DEPC water (20  $\mu$ L). RNA integrity was checked for every sample before proceeding with the preparation for RNA-sequencing. For this purpose, 500 ng from each sample were analysed with an Agilent RNA 6000 Nano Kit in a 2100 Bioanalyzer equipment (Agilent Technologies, USA). Total RNA was quantified using the Nanodrop 2000 (Thermo Scientific, USA). DNA contamination was also assessed with Qubit<sup>®</sup> 2.0 Fluorometer (ThermoFisher Scientific, USA), and concentrations of this nucleic acid in the RNA samples were generally lower than 10 ng/ $\mu$ L (less than 1% contamination). RNA samples were kept at  $-80^{\circ}\text{C}$  until further analysis.

Total proteins were precipitated and washed according to TRIZOL<sup>®</sup> protocol and resuspended in 200  $\mu$ L of the following buffer: 40 mM Tris-HCl (pH 8.0) with 7 M urea, 2 M thiourea, 1% ASB-14, 0.5% triton-X-100, 0.1% SDS, 50 mM MgCl<sub>2</sub>, 0.5 mg/ml bovine pancreas DNase I (Roche, Germany), 0.25 mg/ml bovine pancreas RNase A (Roche, Germany), and a protease inhibitor cocktail mixture (Roche, Germany), according to the manufacturer recommendations. Total protein was quantified using Pierce<sup>™</sup> Coomassie (Bradford) Protein Assay Kit (ThermoFisher Scientific, USA). After quantification, about 5  $\mu$ g were used to run an electrophoretic SDS-Page gel for protein integrity and extraction efficiency determinations. Protein samples were kept at  $-80^{\circ}\text{C}$  until further analysis.

**Transcriptomics.** About 1  $\mu$ g of total RNA from each sample was used for mRNA selection and followed the TruSeq<sup>™</sup> RNA Library Preparation Kit v2 (Illumina<sup>®</sup>) manufacturer's protocol to be converted into a library of DNA fragments ready for sequencing. Briefly, mRNA was fragmented and converted to cDNA. After second strand synthesis, the steps included end repair, ligation of the sequencing adapters and the enrichment of 200–500 base pairs (bp) fragments. Quality and concentration of the fragmented and enriched library was verified with Agilent DNA 7500 Kit in the Agilent 2100 Bioanalyzer equipment (Agilent Technologies, USA). DNA libraries were normalized to 10 nM and pooled for sequencing. Next-generation sequencing (NGS) of the *F. candida* fragments was performed on the Illumina HiSeq2000 platform (Illumina, USA), following a Paired-End RNA Sequencing with  $2 \times 100$  bp read length.

The resulting sequencing reads were mapped against the annotated transcriptome of *F. candida*<sup>21</sup> using Bowtie software<sup>26</sup>, generating the total number of reads per transcript. Differentially expressed transcripts were obtained using Limma and edgeR packages in R (version 3.1.3), following a GLM (Generalized Linear Model) regression, with false discovery rate (FDR) correction for multiple testing. Transcripts with adjusted *p*-value below 0.05, after FDR correction, were considered to be differentially expressed. Gene Ontology (GO) enrichment analysis of the differentially expressed transcripts was performed using topGO package on R, using Fisher exact test with the weight function ( $p < 0.05$ ). Clustering analysis of gene expression data was performed using MeV 4.9.0 software<sup>27</sup>, following the Self Organizing Tree Algorithm (SOTA), Pearson Uncentered metric distance, and cell diversity as cell division criteria. Gene interaction analysis was performed using GeneMANIA plugin from Cytoscape software (version 3.4.0)<sup>28</sup>.

**Proteomics.** Protein denaturation, reduction, digestion with trypsin and labelling of the digested samples with the isobaric tags, were performed according to the iTRAQ<sup>®</sup> Reagents - 8 plex protocol (Applied Biosystems, USA). Total amount from each sample was normalized to exactly 20  $\mu$ g and each sample to be directly compared was subsequently labeled with different isobaric tags. After being randomly labeled, samples to be directly compared were pooled together (8 samples for iTRAQ-8plex: 4 control and 4 pesticide EC50 treatment samples from the same time-point). Samples were dried in a SC110 Savant<sup>™</sup> Speedvac (ThermoFisher Scientific, USA) and stored at  $-20^{\circ}\text{C}$  until analysis. Afterwards, samples underwent a 2 dimension reversed phase (RP) high-pressure liquid chromatographic (HPLC) separation of the peptides: an alkaline fractionation (pH = 10), followed by acidic fractionation (pH = 3). Conditions for both fractionations can be consulted in Supplementary Text. The fractions of interest were collected, pooled into six different fractions and conditioned for the next separation by drying in the Speedvac, followed by resuspension in 20  $\mu$ L of 50% ACN with 0.1% trifluoroacetic acid (TFA) and stored at  $-20^{\circ}\text{C}$ . Of these fractions, 10  $\mu$ L was loaded for the acidic fractionation. The effluent was directly spotted by the Probot device (Dionex-LC Packings) on an OptiTOF<sup>™</sup> LC/MALDI target plate (512 spots), with 30 s intervals. The plates were previously spotted with the supporting matrix, consisting of 4 mg/mL  $\alpha$ -cyano-4-hydroxycinnamic acid, 70% ACN, 0.1% TFA, 10 mM dibasic ammonium citrate and 0.01 pmol/ $\mu$ L Glu-fibrinopeptide B. A 4800 Plus MALDI-TOF/TOF analyzer mass spectrometer (AB SCIEX, USA) was used for the analysis. Identification of the peptides was based on both MS and MS/MS information using the combined peaks for all iTRAQ<sup>®</sup> labelled peptides, whereas the relative quantitation of the peptides was accomplished in the MS/MS mode using the peak intensity of the eight different iTRAQ<sup>®</sup> reporter ions. All spectra were analyzed and compiled with 4000 Series Explorer<sup>™</sup> Software. ProteinPilot<sup>™</sup> software (version 4.0), involving the Paragon Pro-Group algorithms from ABSciex<sup>29</sup>, was used for identification and relative quantitation of peptides and correspondent predicted proteins, using *F. candida* genome as database<sup>22</sup>. The selected parameters were: iTRAQ-label modification at the peptide level, fixed MMTS modification of cysteine, variable biological modifications and amino acid substitutions. Background correction was applied. A false discovery rate (FDR) analysis was performed automatically by the software. After, the proteins with a local FDR  $< 5\%$  were exclusively retained. Of these, only proteins with an “unused score” greater than 1.0 (identification with 90% confidence) and quantified in all replicates, were considered for further analysis. Furthermore, to address differences between control

GO ID	Domain	Term	Annotated	4 Days		7 Days		10 Days	
				Signif.	Adj. p-value	Signif.	Adj. p-value	Signif.	Adj. p-value
<b>Up-regulated</b>									
GO:0055114	BP	Oxidation-Reduction Process	1100	1	0.316	2	<b>0.0391</b>	0	1
GO:0016491	MF	Oxidoreductase Activity	1271	1	1	2	<b>0.0337</b>	0	1
<b>Down-regulated</b>									
GO:0006030	BP	Chitin Metabolic Process	155	0	1	0	1	4	<b>6.90E-06</b>
GO:0006508	BP	Proteolysis	827	1	0.245	0	1	11	<b>2.10E-05</b>
GO:0030704	BP	Vitellogenesis	19	0	1	0	1	2	<b>0.0016</b>
GO:0008654	BP	Phospholipid Biosynthetic Process	73	0	1	0	1	2	<b>0.00217</b>
GO:0008061	MF	Chitin binding	135	0	1	0	1	5	<b>2.70E-05</b>
GO:0003841	MF	1-acylglycerol-3-phosphate O-acyltransferase	6	0	1	0	1	2	<b>0.0001</b>
GO:0016787	MF	Hydrolase	2751	2	1	1	1	24	<b>0.00018</b>
GO:0004568	MF	Chitinase	27	0	1	0	1	2	<b>0.0024</b>
GO:0004181	MF	Metalloprotease	30	0	1	0	1	2	<b>0.0029</b>

**Table 1.** GO enrichment analysis of the differentially expressed transcripts after 4, 7 and 10 days of *F. candida* exposure to the herbicide formulation at the EC50 concentration. GOs presenting singletons were not considered. Significantly enriched GO terms at each time-point (Adjusted p-value < 0.05) are highlighted in bold. BP (Biological Process); MF (Molecular Functions); Signif. (Significant).

and treatment samples, a two-tailed Student's *t*-test was performed on the logarithmic transformed ratios and the significance level was set at  $p < 0.05$ .

**Data availability.** The datasets generated and analyzed during the current study were made available as Supplementary Datasets S4 (differentially expressed transcripts list), S5 (Functional annotation), S6 (detected protein levels). Raw Illumina sequence reads were deposited to the NCBI SRA under accession SRP115951 (Bioproject PRJNA399149).

## Results

**Effects on survival and reproduction.** The measured concentrations of glyphosate in the soil samples showed an average deviation of 17% (<10% in the upper concentrations) compared to the nominal concentrations (see Supplementary Table S2). The estimated effect concentrations were based on the analytical concentrations. The reproduction tests with formulation and pure substance fulfilled the validity criteria described by the standard guideline<sup>25</sup>. No effects on survival were observed in the treatments ( $F_{6,28} = 0.916$ ,  $p = 0.498$ ). The estimated EC50 of the formulation on reproduction was 4.63 mg a.i./kg (95% CI: 0.58–8.68 mg/kg). As for the results with the active ingredient alone, no effects on *F. candida* survival nor on reproduction were observed ( $H = 6.257$ ,  $df = 7$ ,  $p = 0.510$  and  $F_{7,32} = 1.561$ ,  $p = 0.183$ , respectively).

For the mechanistic assessment of the effects caused by the formulation, extra replicates were used to assess if the exposure to EC50 reduced reproduction significantly in this particular experiment (analytical concentration: 4.71 mg a.i./kg; estimated POEA concentration: between 0.87 and 1.49 mg/kg; Supplementary Table S2). After 28 days, on average  $9.2 \pm 0.45$  adult organisms and  $522.6 \pm 117.13$  juveniles were present in control replicates versus  $9.0 \pm 1.22$  adults and  $390.6 \pm 51.77$  juveniles present in EC50 concentration samples, producing no effects on survival ( $t = 0.34$ ,  $df = 4$ ,  $p = 0.37$ , one-tailed), but a significant reduction of approximately 25% in the reproductive output ( $t = 2.3$ ,  $df = 4$ ,  $p = 0.02$ , one-tailed).

**Transcriptomics.** Illumina sequencing yielded 24,124,447 paired-end reads, after trimming. An average of  $83.26 \pm 3.46\%$  of reads per sample was successfully mapped to the transcriptome (38,015 transcripts). Sequencing related statistics per sample can be consulted in Supplementary Table S3. Statistical analysis of differential expression resulted in a total of 117 transcripts differentially expressed over the three different exposure time-points to the formulation, with 8 transcripts differentially regulated after 4 days (8 up, 0 down), 7 differentially regulated after 7 days (7 up, 0 down) and 102 differentially regulated after 10 days (6 up and 96 down), evidencing no overlaps between gene lists from the different time-points. A gene cluster analysis on the complete list of differentially expressed transcripts can be consulted in Supplementary Fig. S1.

The 4-days exposure resulted in six affected annotated transcripts, three of which encoded for a calcium-activated chloride channel regulator. The 7 and 10 days exposures resulted in 4 and 59 affected transcripts with known homologies to functional proteins, most of which are related to lipid metabolism, body development, chitin metabolism, and reproduction. The list of all differentially expressed transcripts with significant blast homologies and correspondent fold change values can be found in Supplementary Table S4.

A GO enrichment analysis was performed individually for each time-point to statistically evaluate the likelihood of certain biological processes (BP) and molecular functions (MF) to be significantly enriched along time (Table 1).

In general, the BP and MF categories significantly associated with up-regulated transcripts were related to oxidation-reduction processes (Table 1). Four BP and five MF were associated with down-regulated transcripts,



Transcript Acc. Number	Protein Homolog	Unused Score	Log <sub>2</sub> Fold Change	p-value
<b>7 Days</b>				
GAMN01010555	50S ribosomal protein L9	2.12	3.76	0.005
GAMN01032080	Cytochrome C	2	3.33	1.55E-06
GAMN01009208	Vitellogenin	8.07	2.71	0.001
GAMN01000325	Ribosomal Prot. S19 40S	2	1.84	0.008
GAMN01003635	Hsp 70 kDa cognate 2	1.62	1.79	0.012
GAMN01002103	ADP/ATP carrier protein	2	1.55	0.008
GAMN01006818	Vitellogenin-2	9.09	1.47	0.006
GAMN01009199	Cuticle protein 6	2	-0.64	0.029
<b>10 Days</b>				
GAMN01000323	Superoxide dismutase [Cu-Zn]	2	6.333	2.02E-06
GAMN01009208	Vitellogenin	14.68	3.97	0.017
GAMN01006818	Vitellogenin-2	14.63	3.03	0.022
GAMN01001328	Alpha-actinin, sarcomeric	2.56	1.46	0.002
GAMN01000430	Fatty acid-binding protein	3.77	0.591	0.003
GAMN01003161	Tropomyosin	14.35	-0.99	0.028
GAMN01001827	Catalase	2	-0.704	0.025
GAMN01004453	NDUFV2, mitochondrial	2	-0.64	0.014
GAMN01006823	Glutathione S-transferase 5	2	-0.591	0.049

**Table 2.** List of proteins with significant altered levels ( $p < 0.05$ ) in *F. candida* after 4, 7 and 10 days of exposure to the herbicide formulation at the EC<sub>50</sub> concentration.

involved in lipid and chitin metabolism, body development and reproduction (Table 1). All significant differential transcripts within each GO term are given in Supplementary Table S5.

**Proteomics.** Protein profiling yielded a total of 101 potential proteins for which 89 (88.1%) known functional homologies were retrieved. The complete protein list can be consulted in Supplementary Table S6. From these proteins, 17 (16.8%) were found to be significantly different between control and contaminated samples, with twelve proteins being more abundant (seven after 7 days and five after 10 days) and five proteins less abundant (one after 7 days and four after 10 days) (Table 2).

The major biological functions of significantly altered proteins are involved in oxidative stress (e.g. superoxide dismutase, catalase, cytochrome c), response to xenobiotic compounds (glutathione S-transferase), and reproduction (e.g. vitellogenin, vitellogenin 2). Less representative processes affected are mitochondrial transport (ADP/ATP carrier protein and NDUFV2) and body development (cuticle protein 6).

After 4 days of exposure, there were no significant differences in protein abundances (see Supplementary Table S6) and a cuticle-related protein was the only significantly low abundant peptide after 7 days. In contrast, proteins such as cytochrome C, heat shock 70 kDa protein cognate 4, ADP-ATP carrier protein, and ribosomal proteins were significantly more abundant at exposure day 7. Egg yolk precursors such as vitellogenin and vitellogenin 2 were observed to become more abundant with time, with increasing values from 7 to 10 days in the contaminated samples.

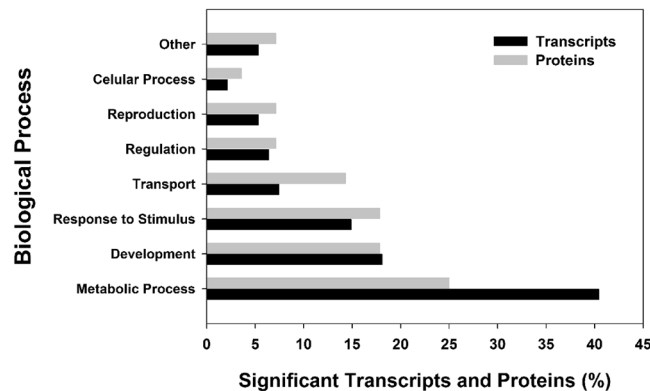
The levels of detoxification and antioxidant related enzymes decreased after 10 days of exposure (e.g. glutathione S-transferase, catalase), whereas the antioxidant superoxide dismutase [Cu-Zn] was significantly higher, being up to 80 times more abundant comparatively to control samples.

**Gene and protein correlations.** The whole set of identified proteins, and corresponding coding transcripts, were used for correlation analyses to increase the power of analysis (Pearson correlation coefficient,  $r_p$ ;  $p < 0.05$ ). Several correlations were tested between datasets belonging to the same time-point, but also between time-points. The correlation coefficients and  $p$ -value results for all analyses are presented in Supplementary Table S7. Transcription levels of stress-related proteins at day 4 were positively correlated with their protein levels at day 10 ( $r_p = 0.60$ ,  $p = 0.007$ ), and this correlation increased when transcript levels at 10 days were compared to protein levels from the same time-point ( $r_p = 0.84$ ,  $p = 0.002$ ). Moreover, proteins related to motor activity were also significantly correlated between transcript and protein level at 4 days ( $r_p = 0.998$ ,  $p = 0.011$ ), but also between transcript levels at 4 days and protein levels at 7 days ( $r_p = 0.997$ ,  $p = 0.047$ ).

Despite the low significance of correlations between transcript and protein levels, there was substantial similarity in the biological processes drawn from associated GO terms (Fig. 1).

## Discussion

The exposure of *F. candida* to Montana<sup>®</sup>, a widely used herbicide formulation in Portugal resulted in a reproduction EC<sub>50</sub> of approximately 5 mg a.i./kg of soil [95% CI: 0.58–8.68], when expressed in terms of the active ingredient glyphosate. At this concentration of active ingredient, the adjuvant concentration in soil is estimated between 0.87–1.49 mg POEA/kg soil.



**Figure 1.** Categorization of all differential *F. candida* gene expression and protein levels classified by GO-term (Biological Process) after exposure to the herbicide formulation during 4, 7 and 10 days. Results are expressed in percentage (%). Blast2GO software (version 4.0) and UniProt database ([www.uniprot.org](http://www.uniprot.org)) were used to group both differential transcript expression and protein levels into the GO categories, according to the major biological processes involved.

It is doubtful whether the effects of Montana<sup>®</sup> on *F. candida* are due to glyphosate. No effects were observed on either reproduction or survival when the pure active ingredient was tested, even at concentrations higher than in the tested formulation. According to Madsen *et al.*<sup>30</sup>, glyphosate was initially patented as a chelator agent (U.S. Patent No. 3.160.632, 1964) due to its propensity to bind metal cations, making it a strong protonophore<sup>31</sup>. It is therefore possible that when mixed into soil, the active ingredient becomes inactivated by binding to clay minerals, causing no effects in organisms at ecologically relevant concentrations, as was verified by Von Mérey *et al.*<sup>32</sup> for earthworms, springtails, and predatory soil mites, and also in the present study for springtails. The differences between observations with the active substance and the formulation tested in our study must therefore consider the adjuvants used in the formulation to enhance stability and penetration of the active compounds into plant cells<sup>33</sup>. Adjuvants are declared as inert diluents, not being directly responsible for the intended pesticide activity<sup>24</sup>, although effects on other non-target organisms are known. According to the safety data sheet of the formulation tested in the present study, POEA was used as adjuvant, though the exact concentration was not mentioned (range from 7 to 12% p/p). Ethoxylated adjuvants are substances with a surfactant action, having alkyl chains and/or ethoxy groups in the ethoxylated chains, which can penetrate cell membranes disrupting their structure and functions<sup>34</sup>. Contrary to what is expected from a pesticide adjuvant, POEA has been reported in the last few years to present even higher toxicity to non-plant organisms than glyphosate alone<sup>33,35</sup>. A statement released by EFSA<sup>36</sup> on this subject concluded that “compared to glyphosate, a higher toxicity of the POE-tallowamine was observed on all endpoints investigated”. This was the basis for the recommendation made by the EC in June 2016 (MEMO-16-2012) to ban POEA from glyphosate-based products, which has already been followed by some member states, including Portugal (the formulation used in this study was purchased prior to this market withdrawal). However, this adjuvant is still largely used worldwide not only in glyphosate-based formulations<sup>24</sup>, which increases the need to understand the real effects and the toxicity mechanisms of such formulations on non-target soil organisms of agricultural fields. Acute effects of POEA on the aquatic environment have already been largely addressed by Tush *et al.*<sup>37</sup>, where a list of acute toxic levels was presented for 24 aquatic species, or more recently in the bivalve mollusk *Crassostrea gigas*<sup>38</sup>. There is, however, limited information available for the terrestrial environment.

In this study, different responses to the formulation could be observed over time, at both gene and protein levels, highlighting the importance of following different time-points for a more accurate assessment of toxicity mechanisms. This pattern of temporal effects in gene expression profiles after exposure to chemical contaminants has been observed in previous ecotoxicogenomics studies with *F. candida* and other soil invertebrates<sup>20,39–41</sup>. Novais *et al.*<sup>40</sup> exposed a potworm to the reproduction EC10 and EC50 of another herbicide and found a higher alteration in gene expression at day 2, with a subsequent decrease in the number of affected genes at day 4 and a late new increase in response after 21 days. This lead us to hypothesize that the lack of differential gene expression responses at 4 and 7 days in the present work, along with the observed increase in affected protein levels already at 7 days of exposure, might be a result of an earlier and undetected gene expression response (before 4 days of exposure), as it is often observed for early stress response enzymes such as catalase or glutathione S-transferase.

Nevertheless, the obtained results give valuable information about the mechanisms of toxicity of an herbicide formulation in this collembolan species, pointing towards induction of cellular oxidative stress, alterations in lipid metabolism, and impairment of development and reproduction. These mechanisms will be hereafter discussed in different sections, focusing on the integrative analysis of effects on both gene expression and protein levels.

**Cellular stress effects.** Although major effects/responses to the formulation were only observed after 10 days, one of the first measured responses of *F. candida* to this exposure was the down-regulation of three transcripts (GAMN01012505, GAMN01007081, GAMN01014780) homologous to the calcium-activated chloride channel regulator 1 gene (CLCA1) at 4 days (Supplementary Fig. S1 and Supplementary Table S4). This gene encodes for a member of the calcium sensitive chloride conductance protein family, which is activated by calcium cations (Ca<sup>2+</sup>)<sup>42</sup>. There are studies with similar formulations to the one used here reporting cellular

calcium disturbance upon exposure. In a work by De Liz Oliveira Cavalli *et al.*<sup>43</sup> with Wistar rat testis or Sertoli cells exposed to another glyphosate-based formulation (Roundup®), the authors reported a Ca<sup>2+</sup> imbalance and proposed that the formulation promotes an intracellular Ca<sup>2+</sup> overload and subsequent cell signaling misregulation, potentially causing cellular stress. More recently, these Ca<sup>2+</sup> imbalance effects were also discussed in a study addressing effects of glyphosate in marine invertebrate *Mytilus galloprovincialis*<sup>44</sup>. After 7 days of exposure, the response of the organisms to a stressful condition was evidenced by over-representation of transcripts involved in oxidation-reduction processes (GO: GO:0055114; Table 1). Two transcripts coding for a Tbh-like monooxygenase protein (Tbh, GAMN01003982, GAMN01004068; Supplementary Fig. S1), were significantly overexpressed, along with a glucosylceramidase homolog (GAMN01005039; CG33090) (Supplementary Fig. S1 and Supplementary Table S4). Arthropods synthesize tyramine beta-hydroxylase (TBH) instead of dopamine-β-hydroxylase (DBH), which converts tyramine into octopamine, the analogous neurotransmitter to noradrenaline of arthropods<sup>45,46</sup>. A stress response in insects is elicited by release of octopamine to cope with stressful energy demanding processes (e.g. flight response, locomotion)<sup>45,47,48</sup>. It is therefore expected that the overexpression of the gene encoding octopamine synthesis after 7 days, represents a response to stress induced by the contaminant. Glucosylceramidase (EC 3.2.1.45) is an enzyme involved in formation of D-glucose and sphingosines, which are important membrane signaling molecules. They are also commonly considered to protect cell surface against harmful environmental factors by forming a mechanically stable and chemically resistant outer leaflet of the plasma membrane lipid bilayer<sup>49</sup>. Thus, the up-regulation of these glucosylceramidase transcripts could indicate a strategy for springtails to protect their cells and avoid penetration of the xenobiotic metabolites.

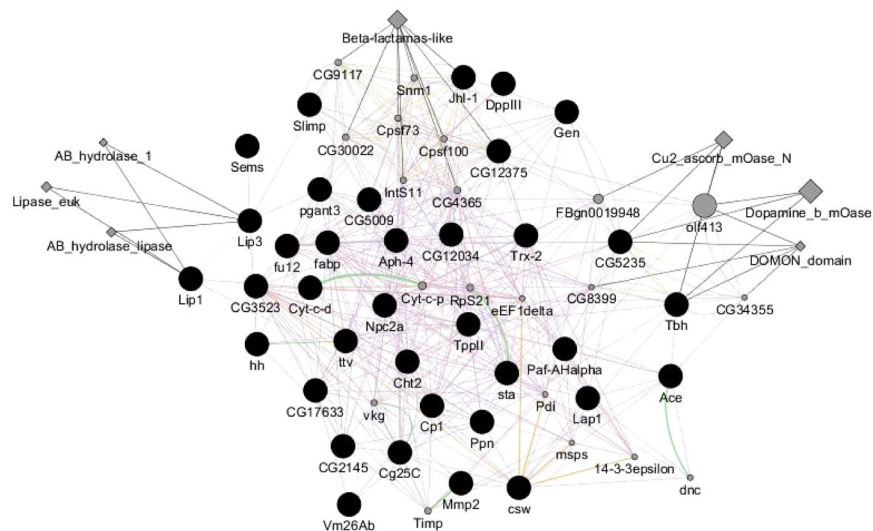
Down-regulation of two collagen (Col14a1) homologs (GAMN01019740; GAMN01005839) was also observed after 10 days, which may be attributed to an increase in oxidative stress, as demonstrated by Webster *et al.*<sup>50</sup> when exposing brown trout to formulated glyphosate (Roundup®). In addition to gene expression responses, stress related proteins such as Cyt C and heat shock 70 kDa protein cognate 4 increased. After 10 days, however, this tendency seemed to be reverted given that some stress response mechanisms were shown to be down-regulated (GO:0004181; catalase, NDUFV2; Tables 2 and S5). Nevertheless, an important antioxidant enzyme, a SOD homolog protein, was found to be highly abundant at 10 days. This is indicative of ROS accumulation, and specially superoxide anions. The high abundance of SOD suggests an extra need for antioxidants to cope with such radicals and cellular stress. Contardo-Jara *et al.*<sup>51</sup> found similar effects in SOD activity when exposing the worm *Lumbriculus variegatus* to environmental concentrations in a Roundup® formulation during four days.

**Developmental and reproductive effects.** The exposure to the herbicide formulation triggered different genes related to disturbances in lipid metabolism, namely by inhibiting the expression of transcripts involved in fatty acid elongation, sphingolipid metabolism, and steroid biosynthesis (for a KEGG<sup>52</sup> pathway analysis see Supplementary Fig. S2). As lipids are key structural components of cellular membranes and participate in signaling and intracellular transport processes, they are expected to be among the first metabolic response agents to a stressor. There is evidence showing effects of glyphosate formulations on lipid metabolism in other non-target species through ROS production, steroid biosynthesis and ultimately, developmental and reproductive effects<sup>38,50,53,54</sup>. To note that after this exposure period the organisms were already 20–22 days old. Since *F. candida* reaches complete sexual maturity after 21 to 24 days of age<sup>55</sup>, this is indicative that at this stage the organisms most likely had started to reproduce. Several metabolic functions implied in body development (GO:0006030; GO:0006508; GO:0004568) and reproduction (GO:0030704) were down-regulated or inhibited after 10 days (Table 1). Molecular events related to reproduction are of particular interest, since they allow a direct link to the observed effect at a higher level of organization (25% inhibition on reproduction). The particular set of down-regulated genes after 10 days of exposure are presented in Fig. 2, illustrating physical and genetic interactions, co-expressions, and protein domain attributes.

A study in *Drosophila melanogaster* showed that 1-acylglycerol-3-phosphate-O-acyltransferase 2 (GAMN01037272, GAMN01010825; fu12) is mostly present in adipocytes and predicted to be involved in lipid storage and lipid droplet turnover during developmental stages and oogenesis<sup>56</sup>. Down-regulation of fu12 causes impairment of lipid mobilization, ultimately leading to developmental arrest and eventual embryo mortality, eventually unraveling the decreased reproductive rates verified in this work. The human homolog (AGPAT2) also plays a critical role in the early stages of growth and development of adipocytes, and in the biosynthesis of major components of cell membrane (glycerophospholipids)<sup>57</sup>, a process where the fatty acid synthase (GAMN01010449; CG3523) also participates. Fatty acid synthase is a multi-enzyme protein, catalyzing the formation of long-chain fatty acids from acetyl-CoA and malonyl-CoA, being expressed by lipogenic tissues, hormone-sensitive cells and proliferating fetal cells. Both genes have been previously shown to be implicated in adipocyte disturbance when down-regulated under xenobiotic contamination<sup>58</sup>. According to Fig. 2, these two genes are co-expressed under xenobiotic exposure<sup>59</sup>, which also seems to be applicable to the formulation tested here.

These mechanisms might also involve the observed expression inhibition of other co-expressed lipid related genes, such as the homolog to the human PFAH1B1 (Paf-AHalpha; GAMN01013576), a putative sphingomyelinase (CG12034; GAMN1008482) and epididymal secretory protein homolog (Npc2a; GAMN01006832) (Fig. 2). The first two are involved in development and embryogenesis<sup>60,61</sup>. Npc2a could also be related to the developmental and reproduction effects since it plays an important role in cholesterol homeostasis and ultimate transport to plasma membrane, controlling the availability of sterol substrate and affecting steroidogenesis such as ecdysteroids<sup>62</sup>. Webster *et al.*<sup>50</sup> also reported disturbances in cholesterol metabolism in fish after exposure to an herbicide formulation containing POEA.

The proteomics analysis also showed that vitellogenin (VTG) and vitellogenin-2 (VTG2) were significantly increased after 7 and 10 days of exposure (Table 2). In oviparous vertebrates and invertebrates, the egg yolk-precursor proteins are very important phospholipoglycoproteins involved in oocyte



**Figure 2.** Gene interaction analysis [co-expression (purple), co-localization (light blue), physical interactions (orange), genetic interactions (green) and consolidated pathways (grey)] performed with *F. candida*'s differentially expressed transcripts after exposure to the herbicide formulation for 10 days. Protein domain attributes and genes related to query genes are presented in grey nodes and query genes presented in black nodes. Interaction analysis was performed with the list of *Drosophila melanogaster* genome homologous genes to *F. candida*, using GeneMANIA plugin from Cytoscape software (version 3.4.0).

maturation<sup>63</sup>. Vitellogenin is produced in fat bodies and taken up by the developing oocytes through a specific receptor-mediated endocytosis in the oocyte outer membrane. After uptake into oocytes, the VTGs are stored in a crystalline form as vitellin, a reserve food-source for the future embryo development<sup>64</sup>. Although no up-regulation of vitellogenin genes (vtgs) was observed at any time-point, the accumulation of these proteins in exposed *F. candida* may be indicative that either their production was stimulated in fat cells/adipocytes or their conversion inside the oocyte was obstructed by some other mechanisms, revealing a clear disturbance in the reproductive cycle and embryogenesis that started to be noticed after 7 days of exposure. Provost-Javier and Rasgon<sup>65</sup> reported a post-transcriptional regulation of vtgs, although using a different arthropod (mosquito). The authors demonstrated that the hormone 20-hydroxyecdysone (ecdysteroid hormone, 20E) was responsible for regulating post-transcriptional splicing of vtg through RUST (Regulated Unproductive Splicing and Translation) mechanism.

Oocyte maturation, fertilization, and embryo development are also general calcium mediated processes<sup>63</sup>. Webster *et al.*<sup>50</sup> showed that low concentrations of Roundup<sup>®</sup> significantly affected calcium signaling pathways and calcium homeostasis in brown trout, disturbing reproduction. It cannot be excluded here that the reproductive impairment could also be a secondary effect caused by calcium imbalance induced by the formulation. Considering that calcium was described to be a good indicator of vitellogenin status in crustaceans and fish species<sup>66</sup> and implied in vitellogenin synthesis<sup>67</sup>, it is therefore plausible the influence of calcium imbalance in the VTG accumulation observed for exposed organisms. In future research, it would be interesting to analyze the vitellin accumulation in the embryos of *F. candida* to further clarify the reproduction impairment mechanism involving these proteins. In a study with *Daphnia magna*, glyphosate did not induce alterations in vtg gene expression<sup>68</sup>, whereas when other authors tested Roundup<sup>®</sup> formulation, a significant effect in fecundity (embryogenesis) was observed<sup>69</sup>. The same reproductive effects were reported by Druart *et al.*<sup>70</sup>, exposing the terrestrial snail *Cantareus asperses* to an environmental concentration of a glyphosate-based herbicide containing 10 to 20% POEA (Bypass<sup>®</sup>). These reports raise some evidences that the reproductive effects observed in *F. candida* might be attributed to the surfactants of the formulation or possible synergistic effects and not to glyphosate itself.

There are also some indications that the impairment of reproduction could be related with a deficiency in octopamine hormone. The two transcripts involved in octopamine synthesis (tbh, mentioned in section 4.1), were found up-regulated at 7 days as a response to stress but were subsequently down-regulated at 10 days (Fig. 2, Supplementary Fig. S1 and Supplementary Table S4). Some studies with *D. melanogaster* and other insects have reported that the absence of TBH protein and consequent impairment of octopamine biosynthesis results in decreased egg deposition by a blockage of mature oocytes within the ovaries, given that octopamine is believed to modulate ovarian contractions and oviductal muscles for oviposition<sup>47,71,72</sup>. This Tbh gene appears to be co-expressed with papilin homolog gene in *D. melanogaster* (Ppn, GAMN01037019; Fig. 2), which is an essential extracellular matrix glycoprotein required for morphogenesis and hypodermal enclosure in the embryo<sup>73,74</sup>.

Downregulation of chitinase (Cht2, GAMN01037556; GAMN01007170) and transferase pgant 3 (GAMN01001593) are examples of developmental inhibition (Figs 2, S1 and Supplementary Tables S4 and S5). Cuticle proteins 6 and 7 were also detected in lower amounts in contaminated samples, although only cuticle protein 6 was found to be statistically decreased (Tables 2 and S6). In arthropods, cuticle proteins are constituents of the organism's exoskeleton and together with chitinases they are key molecules in consecutive ecdysis and



growth stages. The lower abundance of such proteins observed here suggests a slowdown in the development of *F. candida* when compared to control organisms. Similar results were obtained when Qiao *et al.*<sup>75</sup> exposed the collembolan *F. candida* to the pesticide pentachlorophenol, where down-regulation of transcripts involved in molting processes was observed. The delay in development after contaminant exposure is a very well reported behavioral feature of molting organisms, including *F. candida*<sup>76,77</sup>.

**Gene and protein correlations.** Generally for eukaryotes, the cellular concentrations of proteins and their corresponding RNAs have correlation coefficients of approximately 40%. The remaining 60% have been suggested to be mostly controlled at the translation level<sup>78–80</sup>. However, reports of higher rates can be found, especially in simpler biological systems such as bacteria and cell lines<sup>81,82</sup>, or when focusing on specific tissues<sup>83</sup> and differentially expressed genes and proteins<sup>78</sup>. There are several biological aspects to consider when facing these poor correlations between datasets, such as transcriptional and translational machineries and regulatory events, which can also be tissue specific<sup>84,85</sup> and can therefore result in different correlations for the same set of genes and proteins in different tissues<sup>81</sup>. Considering this and the fact that in the present study, the entire body composition of a pool of approximately 75 organisms was used in each replicate, it is possible that different organs and tissues could strongly contribute to the low correlations observed (see Supplementary Table S7).

Contrary to expectations, correlations between datasets from different time-points proved to be lower than those generated at identical time-points. The higher and most significant correlations between gene expression and protein levels were observed at 10 days for a specific subset of stress responses, which can also result from a higher number of stress regulation related genes and enzymes affected by the treatment at this time-point. Also important to mention is the fact that within the same time-point, RNA and proteins to be compared were extracted from exactly the same organisms and these higher correlations could be reflecting the lower inherent biological variation in these cases.

Additionally, only moderate protein diversity was identified throughout this study. There are several environmental proteomics reports indicating low amounts of proteins, possibly related to the biological complexity of the organisms or the proteome range methodologies available<sup>83</sup>, resulting in low correlations between RNA and protein levels. Many ecotoxicological reports refer to around 100 or less proteins identified in chitin based invertebrate species<sup>86–89</sup> but also in other invertebrates<sup>90,91</sup>. Despite the low correlation rates observed here, the information provided by these two levels of datasets revealed to be complementary and corroborative.

## Conclusions

This study was based on state of the art methodologies (omics) to perform an integrated analysis for unraveling the mechanisms of toxic action of an herbicide formulation in *F. candida*. Despite the effects of the formulation, survival and reproduction were not affected by the active ingredient in the present study, even at higher concentrations. Omics analyses allowed identification of mechanisms of toxicity of the formulation, mainly through change in normal cellular respiration and lipid metabolism, leading to increased oxidative stress and impairment of molting and reproduction. These results presented higher similarities with other studies using analogous formulations or even with POEAs, inferred by an integrative analysis of transcript expression and protein abundance. Integrated omics could thus provide useful insights that could hardly be inferred from individual analysis of gene or protein expressions, proving the relevance of integrative analyses in toxicological studies.

## References

1. Tarazona, J. V. Use of new scientific developments in regulatory risk assessments: challenges and opportunities. *Integr. Environ. Assess. Manag.* **9**, 1–23 (2013).
2. Pillai, S. *et al.* Linking toxicity and adaptive responses across the transcriptome, proteome, and phenotype of *Chlamydomonas reinhardtii* exposed to silver. *Proc. Natl. Acad. Sci. USA* **111**, 3490–3495 (2014).
3. Van Aggelen, G. *et al.* Integrating omic technologies into aquatic ecological risk assessment and environmental monitoring: Hurdles, achievements, and future outlook. *Environmental Health Perspectives* **118**, 1–5 (2010).
4. Piña, B., Casado, M. & Quirós, L. Analysis of gene expression as a new tool in ecotoxicology and environmental monitoring. *TrAC - Trends Anal. Chem.* **26**, 1145–1154 (2007).
5. van Straalen, N. M. & Roelofs, D. Genomics technology for assessing soil pollution. *J. Biol.* **7**, 19 (2008).
6. Novo, M. *et al.* Different routes, same pathways: Molecular mechanisms under silver ion and nanoparticle exposures in the soil sentinel *Eisenia fetida*. *Environ. Pollut.* **205**, 385–393 (2015).
7. Chen, G., Den Braver, M. W., Van Gestel, C. A. M., Van Straalen, N. M. & Roelofs, D. Ecotoxicogenomic assessment of diclofenac toxicity in soil. *Environ. Pollut.* **199**, 253–260 (2015).
8. Nestler, H., Groh, K. J., Schönenberger, R., Eggen, R. I. L. & Suter, M. J. F. Linking proteome responses with physiological and biochemical effects in herbicide-exposed *Chlamydomonas reinhardtii*. *J. Proteomics* **75**, 5370–5385 (2012).
9. Lemos, M. F. L. *et al.* Protein differential expression induced by endocrine disrupting compounds in a terrestrial isopod. *Chemosphere* **79**, 570–576 (2010).
10. Ji, C. *et al.* Proteomic and metabolomic analysis of earthworm *Eisenia fetida* exposed to different concentrations of 2,2',4,4'-tetrabromodiphenyl ether. *J. Proteomics* **91**, 405–416 (2013).
11. Wu, S., Xu, X., Zhao, S., Shen, F. & Chen, J. Evaluation of phenanthrene toxicity on earthworm (*Eisenia fetida*): An ecotoxicoproteomics approach. *Chemosphere* **93**, 963–971 (2013).
12. Haider, S. & Pal, R. Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* **14**, 91–110 (2013).
13. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between Protein and mRNA Abundance in Yeast. *Mol. Cell. Biol.* **19**, 1720–1730 (1999).
14. Maier, T., Güell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS Lett.* **583**, 3966–3973 (2009).
15. Filser, J., Wiegmann, S. & Schröder, B. Collembola in ecotoxicology—Any news or just boring routine? *Appl. Soil Ecol.* **83**, 193–199 (2014).
16. Chen, G. *et al.* Integrating transcriptomics into triad-based soil-quality assessment. *Environ. Toxicol. Chem.* **33**, 900–909 (2014).
17. Nota, B., Vooijs, R., Van Straalen, N. M. & Roelofs, D. Expression of mtc in *Folsomia candida* indicative of metal pollution in soil. *Environ. Pollut.* **159**, 1343–1347 (2011).

18. Janssens, T. K. S. *et al.* Narcotic mechanisms of acute toxicity of chlorinated anilines in *Folsomia candida* (Collembola) revealed by gene expression analysis. *Environ. Int.* **37**, 929–939 (2011).
19. Nota, B., Bosse, M., Ylstra, B., van Straalen, N. M. & Roelofs, D. Transcriptomics reveals extensive inducible biotransformation in the soil-dwelling invertebrate *Folsomia candida* exposed to phenanthrene. *BMC Genomics* **10**, 236 (2009).
20. de Boer, T. E., Janssens, T. K. S., Legler, J., van Straalen, N. M. & Roelofs, D. Combined Transcriptomics Analysis for Classification of Adverse Effects As a Potential End Point in Effect Based Screening. *Environ. Sci. Technol.* **49**, 14274–81 (2015).
21. Faddeeva, A. *et al.* Collembolan transcriptomes highlight molecular evolution of Hexapods and provide clues on the adaptation to terrestrial life. *Plos One* **10** (2015).
22. Faddeeva-Vakhrusheva, A. *et al.* Coping with living in the soil: the genome of the parthenogenetic springtail *Folsomia candida*. *BMC Genomics* **18**, 1–14 (2017).
23. Silva, V. *et al.* Distribution of glyphosate and aminomethylphosphonic acid (AMPA) in agricultural topsoils of the European Union. *Sci. Total Environ.* **621**, 1352–1359 (2017).
24. Tush, D. & Meyer, M. T. Polyoxyethylene Tallow Amine, a Glyphosate Formulation Adjuvant: Soil Adsorption Characteristics, Degradation Profile, and Occurrence on Selected Soils from Agricultural Fields in Iowa, Illinois, Indiana, Kansas, Mississippi, and Missouri. *Environ. Sci. Technol.* **50**, 5781–5789 (2016).
25. ISO. Soil quality-inhibition of reproduction of Collembola (*Folsomia candida*) by soil pollutants. International Standards Organisation, Geneva, Switzerland (ISO11267), <https://doi.org/10.3403/01790628U> (1999).
26. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
27. Saeed, A. I. *et al.* TM4: A free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378 (2003).
28. Montojo, J. *et al.* GeneMANIA cytoscape plugin: Fast gene function predictions on the desktop. *Bioinformatics* **26**, 2927–2928 (2010).
29. Shilov, I. V. *et al.* The Paragon Algorithm, a Next Generation Search Engine That Uses Sequence Temperature Values and Feature Probabilities to Identify Peptides from Tandem Mass Spectra. *Mol. Cell. Proteomics* **6**, 1638–1655 (2007).
30. Madsen, H. E. L. *et al.* Stability Constants of Copper(II), Zinc, Manganese(II), Calcium, and Magnesium Complexes of N-(Phosphonomethyl)glycine (Glyphosate). *Acta Chem. Scand.* **32a**, 79–83 (1978).
31. Olorunsogo, O. O. Modification of the transport of protons and Ca<sup>2+</sup> ions across mitochondrial coupling membrane by N-(phosphonomethyl)glycine. *Toxicology* **61**, 205–209 (1990).
32. von Mérey, G., Manson, P. S., Mehrsheikh, A., Sutton, P. & Levine, S. L. Glyphosate and aminomethylphosphonic acid chronic risk assessment for soil biota. *Environ. Toxicol. Chem.* **35**, 2742–2752 (2016).
33. Prosser, R. S., Rodriguez-Gil, J. L., Solomon, K. R., Sibley, P. K. & Poirier, D. G. Effects of the herbicide surfactant MON 0818 on oviposition and viability of eggs of the ramshorn snail (*Planorbella pilsbryi*). *Environ. Toxicol. Chem.* **36**, 522–531 (2017).
34. Nobels, I., Spanoghe, P., Haesaert, G., Robbens, J. & Blust, R. Toxicity ranking and toxic mode of action evaluation of commonly used agricultural adjuvants on the basis of bacterial gene expression profiles. *PLoS One* **6** (2011).
35. Defarge, N. *et al.* Co-formulants in glyphosate-based herbicides disrupt aromatase activity in human cells below toxic levels. *Int. J. Environ. Res. Public Health* **13** (2016).
36. Efsa & European Food Safety Authority. Request for the evaluation of the toxicological assessment of the co-formulant POE-tallowamine. *EFSA J.* **13**, 4303 (2015).
37. Tush, D., Loftin, K. A. & Meyer, M. T. Characterization of polyoxyethylene tallow amine surfactants in technical mixtures and glyphosate formulations using ultra-high performance liquid chromatography and triple quadrupole mass spectrometry. *J. Chromatogr. A* **1319**, 80–87 (2013).
38. Séguin, A. *et al.* Sub-lethal effects of a glyphosate-based commercial formulation and adjuvants on juvenile oysters (*Crassostrea gigas*) exposed for 35 days. *Mar. Pollut. Bull.* **117**, 348–358 (2017).
39. Nota, B. *et al.* Gene expression analysis of collembola in cadmium containing soil. *Environ. Sci. Technol.* **42**, 8152–8157 (2008).
40. Novais, S. C., De Coen, W. & Amorim, M. J. B. Transcriptional responses in *Enchytraeus albidus* (Oligochaeta): Comparison between cadmium and zinc exposure and linkage to reproduction effects. *Environ. Toxicol. Chem.* **31**, 2289–2299 (2012).
41. Novais, S. C. *et al.* Differential gene expression analysis in *Enchytraeus albidus* exposed to natural and chemical stressors at different exposure periods. *Ecotoxicology* **21**, 213–224 (2012).
42. Huang, F., Wong, X. & Jan, L. Y. International Union of Basic and Clinical Pharmacology. LXXXV: calcium-activated chloride channels. *Pharmacol. Rev.* **64**, 1–15 (2012).
43. De Liz Oliveira Cavalli, V. L. *et al.* Roundup disrupts male reproductive functions by triggering calcium-mediated cell death in rat testis and Sertoli cells. *Free Radic. Biol. Med.* **65**, 335–346 (2013).
44. Milan, M. *et al.* Ecotoxicological effects of the herbicide glyphosate in non-target aquatic species: Transcriptional responses in the mussel *Mytilus galloprovincialis*. *Environ. Pollut.* **237**, 442–451 (2018).
45. Roeder, T. Tyramine and octopamine: ruling behavior and metabolism. *Annu. Rev. Entomol.* **50**, 447–477 (2005).
46. Roeder, T. Octopamine in invertebrates. *Progress in Neurobiology* **59**, 533–561 (1999).
47. Rubinstein, C. D. & Wolfner, M. F. *Drosophila* seminal protein ovulin mediates ovulation through female octopamine neuronal signaling. *Proc. Natl. Acad. Sci. USA* **110**, 17420–5 (2013).
48. Adamo, S. Norepinephrine and octopamine: linking stress and immune function across phyla. *Invertebr. Surviv. J.* **5**, 12–19 (2008).
49. Chen, K. *et al.* Transcriptome and molecular pathway analysis of the hepatopancreas in the Pacific white shrimp *Litopenaeus vannamei* under chronic low-salinity stress. *Plos One* **10** (2015).
50. Webster, U. M. T. & Santos, E. M. Global transcriptomic profiling demonstrates induction of oxidative stress and of compensatory cellular stress responses in brown trout exposed to glyphosate and Roundup. *BMC Genomics* **16**, 32 (2015).
51. Contardo-Jara, V., Klingelmann, E. & Wiegand, C. Bioaccumulation of glyphosate and its formulation Roundup Ultra in *Lumbricus variegatus* and its effects on biotransformation and antioxidant enzymes. *Environ. Pollut.* **157**, 57–63 (2009).
52. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
53. Portier, C. J. & Clausen, P. Re: Tarazona *et al.*: Glyphosate toxicity and carcinogenicity: a review of the scientific basis of the European Union assessment and its differences with IARC. *Arch. Toxicol.* **91**, 3195–3197, <https://doi.org/10.1007/s00204-017-1962-5> (2017).
54. Dutra, B. K., Fernandes, F. A., Failace, D. M. & Oliveira, G. T. Effect of roundup?? (glyphosate formulation) in the energy metabolism and reproductive traits of *Hyalella castroi* (Crustacea, Amphipoda, Dogielinotidae). *Ecotoxicology* **20**, 255–263 (2011).
55. Diogo, J. B., Natal-Da-Luz, T., Sousa, J. P., Vogt, C. & Nowak, C. Controversies and Solutions in Environmental Sciences (Editor: Henner Hollert) Tolerance of Genetically Characterized *Folsomia candida* Strains to Phenmedipham Exposure A comparison between reproduction and avoidance tests\*. *J Soils Sediments* **7**, 388–392 (2007).
56. Kühnlein, R. P. Thematic review series: Lipid droplet synthesis and metabolism: from yeast to man. Lipid droplet-based storage fat metabolism in *Drosophila*. *J. Lipid Res.* **53**, 1430–6 (2012).
57. Cortés, V. A. *et al.* Molecular Mechanisms of Hepatic Steatosis and Insulin Resistance in the AGPAT2-Deficient Mouse Model of Congenital Generalized Lipodystrophy. *Cell Metab.* **9**, 165–176 (2009).
58. Mesnier, A. *et al.* The transcriptional effects of PCB118 and PCB153 on the liver, adipose tissue, muscle and colon of mice: Highlighting of Glut4 and Lipin1 as main target genes for PCB induced metabolic disorders. *Plos One* **10** (2015).

59. King-Jones, K., Horner, M. A., Lam, G. & Thummel, C. S. The DHR96 nuclear receptor regulates xenobiotic responses in *Drosophila*. *Cell Metab.* **4**, 37–48 (2006).
60. Hirotsune, S. *et al.* Graded reduction of Pafah1b1 (Lis1) activity results in neuronal migration defects and early embryonic lethality. *Nat. Genet.* **19**, 333–339 (1998).
61. Hebbar, S. *et al.* Ceramides And Stress Signalling Intersect With Autophagic Defects In Neurodegenerative *Drosophila* blue cheese (bchs) Mutants. *Sci Rep* **5**, 15926 (2015).
62. Niwa, R. & Niwa, Y. S. The fruit fly *Drosophila melanogaster* as a model system to study cholesterol metabolism and homeostasis. *Cholesterol* **2011**, 176802 (2011).
63. Silva-Neto, M. A. C. *et al.* Isolation of a calcium-binding phosphoprotein from the oocytes and hemolymph of the blood-sucking insect *Rhodnius prolixus*. *J. Biol. Chem.* **271**, 30227–30232 (1996).
64. Tufail, M. & Takeda, M. Molecular characteristics of insect vitellogenins. *Journal of Insect Physiology* **54**, 1447–1458 (2008).
65. Provost-Javierand, K. N. & Rasgon, J. L. 20-hydroxyecdysone mediates non-canonical regulation of mosquito vitellogenins through alternative splicing. *Insect Mol. Biol.* **23**, 407–416 (2014).
66. Webster, U. M. T., Shears, J. A., Moore, K. & Santos, E. M. Identification of conserved hepatic transcriptomic responses to 17 $\beta$ -estradiol using high-throughput sequencing in brown trout. *Physiol. Genomics* **47**, 420–431 (2015).
67. Yeo, I.-K. & Mugiya, Y. Effects of extracellular calcium concentrations and calcium antagonists on vitellogenin induction by estradiol-17B in primary hepatocyte culture in the rainbow trout *Oncorhynchus mykiss*. *Gen. Comp. Endocrinol.* **105**, 294–301 (1997).
68. Le, T. H. *et al.* Effects of glyphosate and methidathion on the expression of the Dhb, Vtg, Arnt, CYP4 and CYP314 in *Daphnia magna*. *Chemosphere* **79**, 67–71 (2010).
69. Cuhra, M., Traavik, T. & Bøhn, T. Clone- and age-dependent toxicity of a glyphosate commercial formulation and its active ingredient in *Daphnia magna*. *Ecotoxicology* **22**, 251–262 (2013).
70. Druart, C., Gimbert, F., Scheifler, R. & de Vaufléury, A. A full life-cycle bioassay with *Cantareus aspersus* shows reproductive effects of a glyphosate-based herbicide suggesting potential endocrine disruption. *Environ. Pollut.* **226**, 240–249 (2017).
71. Monastiriotti, M. Distinct octopamine cell population residing in the CNS abdominal ganglion controls ovulation in *Drosophila melanogaster*. *Dev. Biol.* **264**, 38–49 (2003).
72. Monastiriotti, M., Linn, C. E. & White, K. Characterization of *Drosophila* tyramine beta-hydroxylase gene and isolation of mutant flies lacking octopamine. *J. Neurosci.* **16**, 3900–3911 (1996).
73. Kawano, T. *et al.* *C. elegans* mig-6 encodes papilin isoforms that affect distinct aspects of DTC migration, and interacts genetically with mig-17 and collagen IV. *Development* **136**, 1433–1442 (2009).
74. Kramerova, I. A. *et al.* Papilin in development; a pericellular protein with a homology to the ADAMTS metalloproteinases. *Development* **127**, 5475–5485 (2000).
75. Qiao, M. *et al.* Transcriptional profiling of the soil invertebrate *Folsomia candida* in pentachlorophenol-contaminated soil. *Environ. Toxicol. Chem.* **34**, 1362–1368 (2015).
76. Smit, C. E., Stam, E. M., Baas, N., Hollander, R. & Van Gestel, C. A. M. Effects of dietary zinc exposure on the life history of the parthenogenetic springtail *Folsomia candida* (Collembola: Isotomidae). *Environ. Toxicol. Chem.* **23**, 1719–1724 (2004).
77. Addison, J. A. Safety Testing of Tebufenozide, a New Molt-Inducing Insecticide, for Effects on Nontarget Forest Soil Invertebrates. *Ecotoxicol. Environ. Saf.* **33**, 55–61 (1996).
78. Koussounadis, A., Langdon, S. P., Um, I. H., Harrison, D. J. & Smith, V. A. Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Sci. Rep.* **5**, 10775 (2015).
79. Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–32 (2012).
80. Schwanhauser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
81. Edwards, F. *et al.* Adaptive resistance of melanoma cells to RAF inhibition via reversible induction of a slowly dividing differentiated state. *Mol Syst Biol* **13**, 1–24 (2017).
82. Pan, C., Kumar, C., Bohl, S., Klingmueller, U. & Mann, M. Comparative proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol. Cell. Proteomics* **8**, 443–450 (2009).
83. Monteiro Silva, G. & Vogel, C. Quantifying gene expression: the importance of being subtle. *Mol Syst Biol* **12**, 1–2 (2016).
84. Kosti, I., Jain, N., Aran, D., Butte, A. J. & Sirota, M. Cross-tissue analysis of gene and protein expression in normal and cancer tissues. *Sci. Rep.* **6**, 24799 (2016).
85. Buszczak, M., Signer, R. A. J. & Morrison, S. J. Cellular differences in protein synthesis regulate tissue homeostasis. *Cell* **159**, 242–251 (2014).
86. Lafontaine, A. *et al.* Proteomic response of *Macrobrachium rosenbergii* hepatopancreas exposed to chlordecone: Identification of endocrine disruption biomarkers? *Ecotoxicol. Environ. Saf.* **141**, 306–314 (2017).
87. Wei, D. *et al.* Comparative proteomic analysis of *Bactrocera dorsalis* (Hendel) in response to thermal stress. *J. Insect Physiol.* **74**, 16–24 (2015).
88. Muñoz-Gómez, A., Corredor, M., Benítez-Páez, A. & Peláez, C. Development of quantitative proteomics using iTRAQ based on the immunological response of *Galleria mellonella* larvae challenged with *Fusarium oxysporum* microconidia. *Plos One* **9** (2014).
89. Leroy, D., Haubruge, E., De Pauw, E., Thomé, J. P. & Francis, F. Development of ecotoxicoproteomics on the freshwater amphipod *Gammarus pulex*: Identification of PCB biomarkers in glycolysis and glutamate pathways. *Ecotoxicol. Environ. Saf.* **73**, 343–352 (2010).
90. Mu, H., Sun, J., Heras, H., Chu, K. H. & Qiu, J. W. An integrated proteomic and transcriptomic analysis of perivitelline fluid proteins in a freshwater gastropod laying aerial eggs. *J. Proteomics* **155**, 22–30 (2017).
91. Piechnik, C. A., Hockner, M., de Souza, M. R. D. P., Donatti, L. & Tomanek, L. Time course of lead induced proteomic changes in gill of the Antarctic limpet *Nacella Concinna* (Gastropoda: Patellidae). *J. Proteomics* **151**, 145–161 (2017).

## Acknowledgements

This study had the support of FCT, through strategic project UID/MAR/04292/2013, project ENVIRONOME (PTDC/AGR-PRO/3496/2012 - POCI-01-0145-FEDER-016773), and grants awarded to T.S. (SFRH/BD/98266/2013), S.N. (SFRH/BPD/94500/2013), and T.N.L. (SFRH/BPD/110943/2015). D.R. received financial support from BE-BASIC project F08.001.03 and EU FP7 Sustainable Nanotechnologies project (SUN, grant number 604305). The authors also acknowledge the help provided by Isabel Vandenberghe, Gonzalez Van Driessche and Riet Vooijs.

## Author Contributions

T.S. and S.C.N.: Data collection, interpretation and manuscript writing; T.N.L.: Ecotoxicological data collection and interpretation; T.d.B. and D.R.: Transcriptomics data collection, interpretation; B.D.: Proteomics data collection and interpretation; S.C.N., M.F.L.L., T.N.L., J.P.S., B.D., T.d.B., D.R., N.M.v.S.: Experimental design and critical review of the manuscript. D.R., M.F.L.L.: financial support.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-29662-6>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018