

First *de novo* transcriptome analysis of the Antarctic springtail *Cryptopygus terranovus* (Collembola: Isotomidae) following mid-term heat exposure

CLAUDIO CUCINI ¹, CHIARA LEO ², FRANCESCO NARDI ¹, SAMUELE GRECO ³,
CHIARA MANFRIN ³, PIERO G. GIULIANINI ³ and ANTONIO CARAPPELLI ¹

¹Department of Life Sciences, University of Siena, 53100 Siena, Italy

²Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

³Department of Life Sciences, University of Trieste, 34127 Trieste, Italy

claudio.cucini@student.unisi.it

Abstract: Global human activities, such as greenhouse emissions and pollution, are promoting global warming, environmental changes and biodiversity reduction. Pristine environments such as those of Antarctica are not immune to these phenomena, as is noticeable from the increasing pace of the temperature shift registered within the continent in recent decades. In this study, we describe the first *de novo* transcriptome analysis of the endemic Antarctic springtail (= collembolan) *Cryptopygus terranovus* and we evaluate its global gene expression response following a mid-term exposure of 20 days to 18°C. Expression data are compared with wild specimens sampled from their native environment to outline the molecular mechanisms triggered by the thermal exposure. Although individual plasticity in transcript modulation is assessed, several pathways appear to be differentially modulated in springtails subjected to the heat treatment *vs* wild specimens. Through enrichment analysis, we show that protein catabolism, fatty acid metabolism and a sexual response characterized by spermatid development are induced, while carbohydrate consumption, lipid catabolism and tissue development are downregulated in treated samples compared to controls.

Received 3 August 2020, accepted 23 March 2021

Key words: gene expression, RNA-Seq, springtails, thermal stress, transcriptomics, Victoria Land

Introduction

Environmental changes are jeopardizing worldwide biodiversity. The increase in atmospheric CO₂ concentration due to human greenhouse gas emissions is drastically modifying Earth's ecosystems (Convey & Peck 2019). Climate change has caused a significant increase in atmospheric temperature, shifts in precipitation and atmospheric circulation patterns and a higher frequency of extreme events (e.g. droughts, floods or heatwaves; Convey & Peck 2019). In this respect, isolated and highly fragmented areas of our planet, such as the polar regions and hence Antarctica, are likely to be thoroughly affected by such changes in terms of biodiversity composition and loss. The Antarctic terrestrial ecosystem is highly susceptible not only to phenological mismatches, but also to the introduction and establishment of alien species that alter the natural equilibria of native ones (e.g. Convey & Peck 2019, Wauchope *et al.* 2019). In contrast to other southern landmasses, Antarctica is more susceptible to climatic variations due to its environmental conditions. For instance, global warming firstly affects polar regions, leading to ice melting, which contributes to 1) the ice

level increasing, 2) marine chemical properties alterations and 3) the appearance of ice-free areas, which allow for the establishment of non-native species (Convey & Peck 2019) and the intercontinental dispersal of resident biota. In addition, even if Antarctica is protected under international regulations, spatially explicit conservation planning (i.e. the prioritization of planned actions) is still in its infancy (Wauchope *et al.* 2019), whereas anthropogenic disturbance is increasing (Convey & Peck 2019).

It has been suggested that arthropods' thermal sensitivity may follow a latitudinal gradient, with tropical species being more vulnerable to temperature increases than polar biota, because the former are already closer to their upper physiological optima, whereas the latter live well below this limit (Deutsch *et al.* 2008). Moreover, other studies highlighted the importance of micro-environment conditions (e.g. nutrients, humidity, etc.) as the primary variable impacting species wellness and population growth during temperature shifts (Convey *et al.* 2002, Bokhorst *et al.* 2008).

How environmental changes and global warming may affect the resident biota's physiology has been studied,

using molecular approaches, in some Antarctic marine groups, such as molluscs (Reed & Thatje 2015), echinoderms, crustaceans (González-Aravena *et al.* 2018) and bony fishes (Bilyk *et al.* 2018). In comparison, very little is known on how the Antarctic terrestrial fauna endures increasing temperatures (Michaud *et al.* 2008, Everatt *et al.* 2014), and almost no molecular data are available for microarthropods (Everatt *et al.* 2013).

Collembola (= springtails) is one of the most represented groups of strictly terrestrial animals in Antarctica (McGaughan *et al.* 2011). Springtails play a pivotal role in Antarctic soil functioning; therefore, significant efforts are required to understand how these taxa are responding to climate and environmental changes. A significant molecular survey (using microarrays) has been conducted on the effects of cold and dehydration stress on springtail species from the Maritime Antarctic (Burns *et al.* 2010). These should be expanded to include representative species from Continental Antarctica and to study, besides cold tolerance, the physiological and molecular mechanisms activated (if any) with short- and long-term heat exposure. These results will be relevant to more in-depth investigations of how basal Hexapoda respond to global warming and phenological mismatches (Everatt *et al.* 2013).

The Antarctic springtail *Cryptopygus terranovus* is an endemic species distributed along the northern Victoria Land coast. It can be considered an important species for studying how global warming can threaten continental springtails, as it has been recently investigated on morphological and molecular grounds (data not published yet) and its distribution range covers an area where several international research stations are located (Caruso *et al.* 2009). Phylogeographical and phylogenetic studies demonstrated that this species was present in Antarctica and was already well differentiated from other Antarctic springtail lineages during the mid- to late Miocene, implicating it as a long-term resident of the continent (Carapelli *et al.* 2017). As such, *C. terranovus* managed to survive in local refugia during recent glacial cycles, becoming the predominant springtail species along a territory influenced by the volcanic activity of Mount Melbourne (Carapelli *et al.* 2017). Several molecular population genetics studies have been recently conducted on the species (Fanciulli *et al.* 2001, Carapelli *et al.* 2017, Collins *et al.* 2019), and its phylogenetic relatedness to the better-studied and co-generic *Cryptopygus antarcticus* has been assessed (Carapelli *et al.* 2019). Everatt *et al.* (2013) established that *C. antarcticus* is able to survive during short-term heat exposure (72 h at 35°C) when subjected to a slow warming rate. However, the survival rate of *C. antarcticus* decreases during long periods (> 40 days) at 10°C. For these reasons, further investigations are

needed to explore the molecular and physiological mechanisms that are activated during long-term heat exposure and are therefore responsible for, or at least associated with, heat endurance in Antarctic springtails. In this respect, we focused on the identification of pathways that are influenced in *C. terranovus* under artificially warm conditions by means of the first transcriptome analysis performed on this species.

Methods

Sample collection and preparation

Individuals of *C. terranovus* were sampled during the Italian National Antarctic Program (PNRA) expedition in the summer of 2018 at the Mario Zucchelli Station (MZS; Victoria Land, 74°42'02"S 164°06'45"E; at 70 m altitude). Individuals were subdivided in a control (CT) group and a treatment (HE) group. The temperature for the former was 5.4°C (SD = 3.27°C; data registered with a data logger for 20 days at the sampling location). The HE group was collected and stressed in plastic boxes at MZS with stones at 18°C for 20 days under controlled conditions (24 h full daylight, as expected during the Antarctic summer, and 62% relative humidity) and fed *ad libitum* with natural moss before analysis. Therefore, the average temperature difference during the experimental time between the CT and HE groups was ~12.5°C. Both groups were then transferred in RNAlater solution (Sigma-Aldrich) and stored at -20°C for further processing.

RNA extraction and sequencing

Three samples of *C. terranovus* were screened for each of the CT and HE groups (biological replicates). Each sample was composed of a pool of three individuals (i.e. for a total of nine specimens for each of the CT and HE groups), whose RNA was extracted with TRI Reagent® solution (Sigma-Aldrich) following the manufacturer's instructions. The concentration and quality of the RNA were checked by capillary electrophoresis using Agilent High Sensitivity RNA TapeStation on a BioAnalyzer 2100 instrument (Agilent Technologies). Library preparation was performed using Lexogen's SENSE™ mRNA-Seq Kit V2 as per the manufacturer's instructions. All of the libraries were sent to an external sequencing service (ARGO Open Lab Platform for Genome Sequencing, AREA Science Park, Trieste, Italy) to be sequenced on a NovaSeq 6000 system (Illumina; 100 bp paired end reads). Raw sequencing reads are available at the Sequence Read Archive (SRA; National Center for Biotechnology Information (NCBI)) with the BioProject ID PRJNA645792.

De novo transcriptome assembly and annotation

Raw sequences were quality checked using *FastQC* and *MultiQC* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, Ewels *et al.* 2016). In order to tackle possible assembly bias due to overrepresented sequences and therefore reduce assembly times, read files were merged and normalized using *BBNorm* (*BBTools*: <http://sourceforge.net/projects/bbmap>) with the following flags: *min* = 1, *max* = 100, *prefilter* = *t*, *fixspikes* = *t*. De novo transcriptome assembly was performed with the *Oyster River Protocol (ORP)* v.2.3.1 (MacManes 2018) with default settings. Briefly, the protocol performs an initial trimming with error correction and then a *de novo* assembly using three different programs: *Trinity* v.2.8.4 (Grabherr *et al.* 2011), *Trans-ABYSS* v.2.0.1 (Robertson *et al.* 2010) and *SPAdes* v.3.13.0 (Bushmanova *et al.* 2019), merging their results as the final step.

In order to improve completeness assembly, reads were mapped back on the assembled transcriptome and unmapped reads were collected and mapped on the intermediate assemblies generated by *ORP* with *Salmon* v.0.13.1 (Patro *et al.* 2017). Contigs with coverage > 10 were then added to the final transcriptome and sequences < 200 nt were removed. *Cryptopygus terranovus* rRNAs and mtDNAs available on GenBank (NCBI) in February 2020 were downloaded, clustered with *cd-hit-est* v.4.7 (Fu *et al.* 2012) and used as database in a *blastn+* v.2.7.1 (Camacho *et al.* 2009), against which the whole transcriptome was queried in order to filter out high-similarity hits (e-value < 1e-50), minimizing ribosomal and mitochondrial contaminations and hence improving the transcript signal. Final completeness was assessed with *BUSCO* v.3 against the Arthropoda database (Seppey *et al.* 2019). Transcript annotation was performed using the *annotaM* pipeline (<https://gitlab.com/54mu/annotaM>), which annotates the assemblies by querying several databases (UniprotKb, OrthoDB, PFAM) to recover the gene IDs, Gene Ontologies (GOs) and associated pathways.

Mapping, differentially expressed gene identification and hypergeometric tests

Gene expression values were estimated as transcripts per million (TPM) and counted with *Salmon* v.0.13.1 (Patro *et al.* 2017) using the not-normalized reads in mapping-based mode. The *NOISeq* package (Tarazona *et al.* 2015) was used in the *R* environment to identify differentially expressed genes (DEGs) between the two conditions. The *noisegbio* function was applied (Protocol S1) and DEGs with a probability cut-off of 0.9999 and \log_2 -fold change > |2| were kept for statistical investigations. An enrichment analysis based on the hypergeometrical test of the correspondent GOs and PFAMs was performed (Protocols S2 & S3) and only the

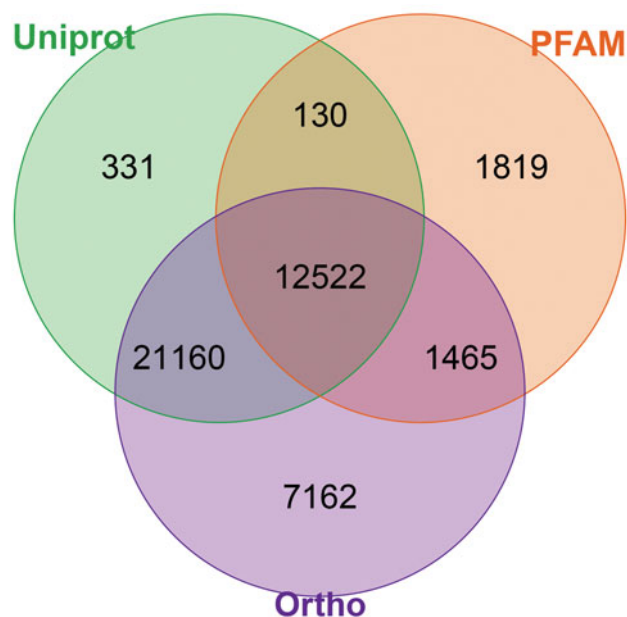


Fig. 1. Venn diagram of the annotated genes shared by and unique to each database. Numbers shown in the graph correspond to the number of contigs in each dataset. Orthologs represent the majority within them. A total of 13% of all annotations are common to all three of the investigated databases.

ones represented by a significant *P*-value (< 5×10^{-4}) and observation counts > 3 were retained.

Results

Wild temperature trend

CT springtails were sampled in the wild environment after 20 days of temperature data acquisition. During this time, registered climate conditions of the soil (recorded every 10 min between 27 December 2017 and 15 January 2018) varied between 0.11°C and 12.60°C, with an average temperature of 5.4°C (SD = 3.27°C). Therefore, a difference of ~12.5°C was recorded between the CT and HE samples.

De novo assembled transcriptome

Paired-end Illumina sequencing resulted in ~14–41 million reads per sample (Table SI). After the filtering steps described in the 'Methods' section, the *de novo* assembled transcriptome had a total of 64,054 contigs with an average length of 369.25 nt (Table SII). The N50 statistic, defined as the sequence length of the shortest contig at 50% of the total transcriptome length (useful for assessing the mean quality of a transcriptome), was 345 nt, and the *BUSCO* report against the Arthropoda database indicated the completeness as acceptable (as *C. terranovus*

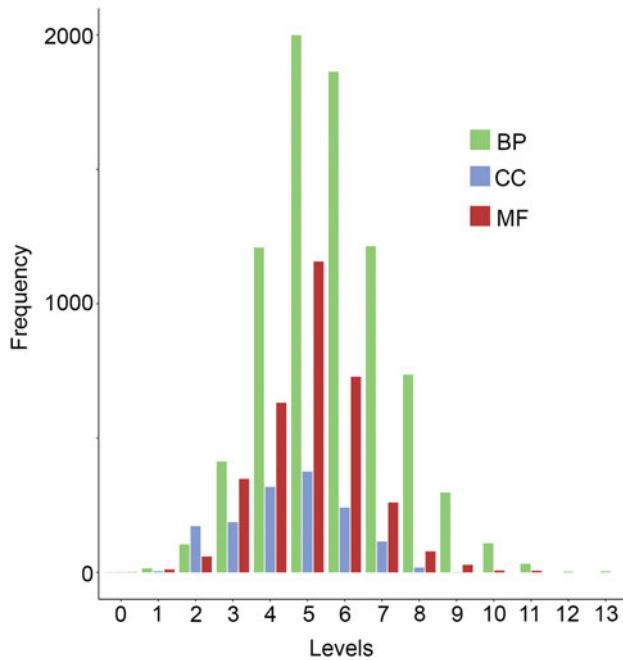


Fig. 2. Levels of representation of Gene Ontology terms for biological processes (BPs) in green, cellular components (CCs) in blue and molecular functions (MFs) in red. The maximum resolution was achieved between the 5th and 13th levels (78.3% of Gene Ontology terms assigned).

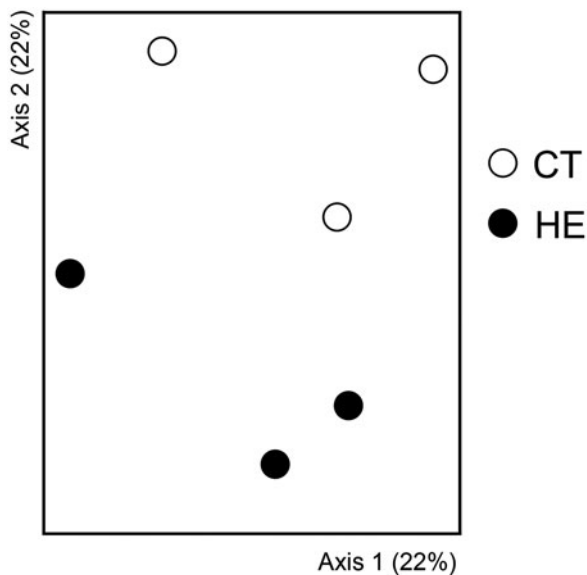


Fig. 3. Principal component analysis of control (CT) group (white dots) and heated (HT) group (black dots) samples.

is neither a model organism nor phylogenetically close to one of them) while also evidencing a certain degree of assembly fragmentation: 29.4% complete (21.7% single copy, 7.7% duplicated), 47.1% fragmented and 23.5% missing *BUSCOs*. The GC content in the final assembled transcriptome was $45.53 \pm 6.08\%$.

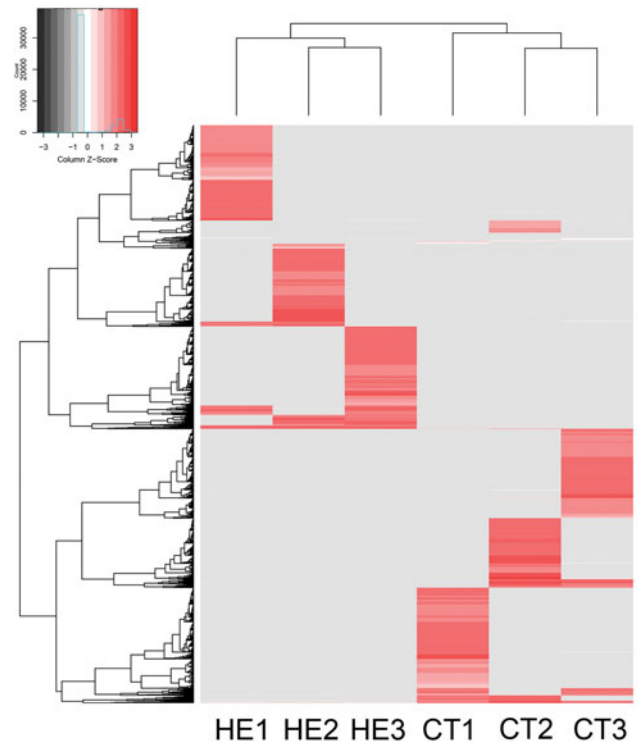


Fig. 4. Heatmap summarizing the \log_{10} -transformed ratio of differentially expressed genes between the control (CT) group and the heated (HT) group obtained with the *NOISEq* package using a probability threshold of 0.9999.

Functional annotation

A total of 45,303 genes were annotated in at least one of the used databases, leaving 29.3% of transcript unsuccessfully annotated (Table SII). In detail, 42,309 genes were OrthoDB hits, 34,143 genes were Uniprot-Swissprot hits and 15,936 genes contained at least one PFAM domain (Fig. 1). Following the GO annotation, a total of 12,747 different GO terms were recovered, of which 7998 represent biological processes (BPs), 3317 represent molecular functions (MFs) and 1432 represent cellular components (CCs). GO terms were distributed across 13 different GO levels. Between the 5th and 13th levels, there were 78.3% GO terms for BPs, 52.3% terms for CCs and 68.2% terms for MFs, indicating that a good resolution was achieved by the functional annotation process (Fig. 2).

Differentially expressed genes and enrichment analysis

A total of 7637 significant DEGs were found, of which 3719 were upregulated and 3918 were downregulated in the HE samples (Table SII). Overall, 29% of the total DEGs lacked a meaningful functional annotation. Furthermore, both the principal component analysis and

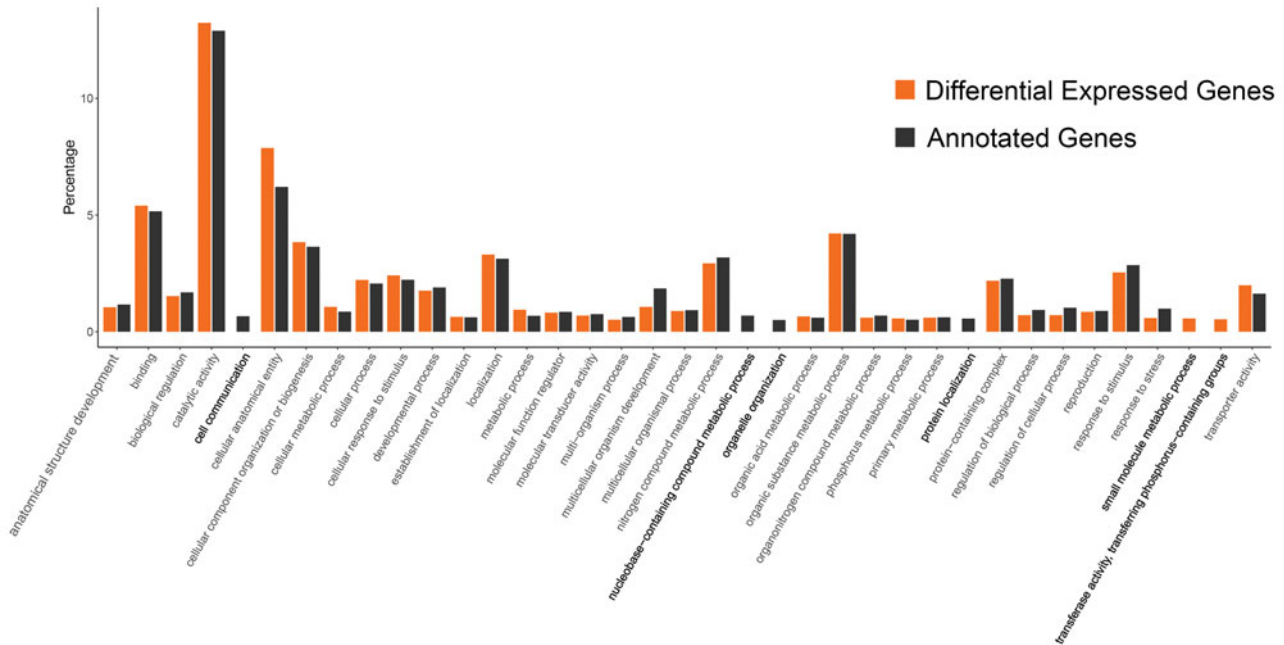


Fig. 5. Abundance differences between the Gene Ontologies of differentially expressed genes (orange) and annotated genes (dark grey) at the 99.5th percentile. The terms that diverge between the two datasets are highlighted in bold.

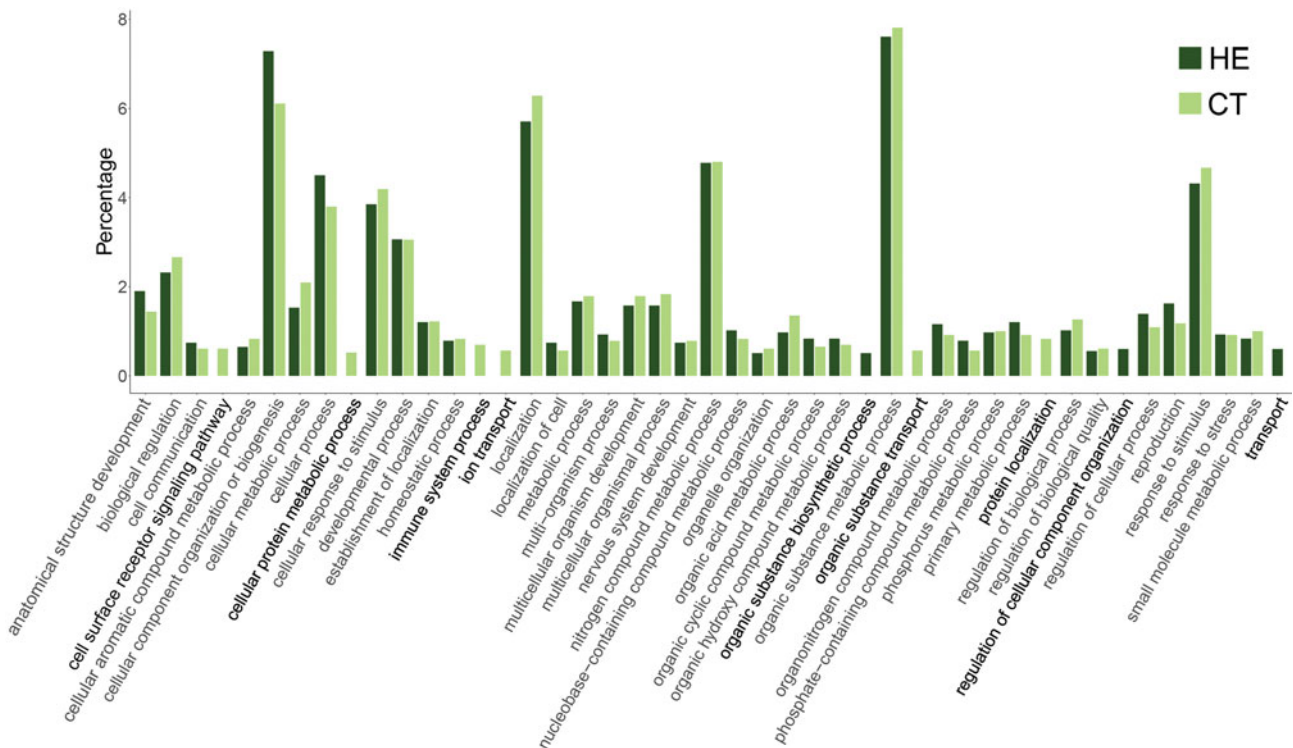


Fig. 6. Biological process terms at the 99.5th percentile compared between the heated (HE) group (dark green) and the control (CT) group (light green). Expression differences between the two conditions are highlighted in bold.

the heatmap displayed a clear clustering between the two conditions (Figs 3 & 4).

A comparative assessment of GO classifications between the overall annotated genes and DEGs identified a limited

number of differences. We considered only the most representative GOs, excluding the rarest (below the 0.5th percentile), and whether, among them, any particular difference was assessed (Fig. 5). The comparative

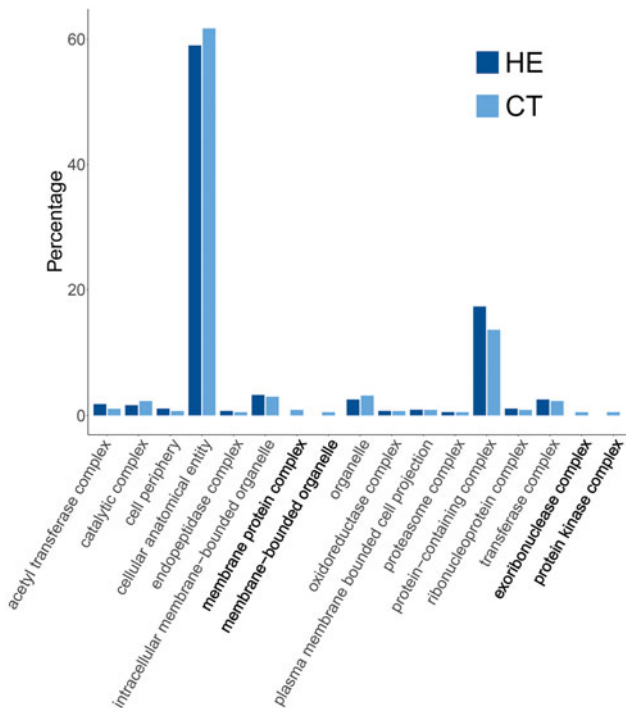


Fig. 7. Cellular component terms at the 99.5th percentile compared between the heated (HE) group (dark blue) and the control (CT) group (light blue). Expression differences between the two conditions are highlighted in bold.

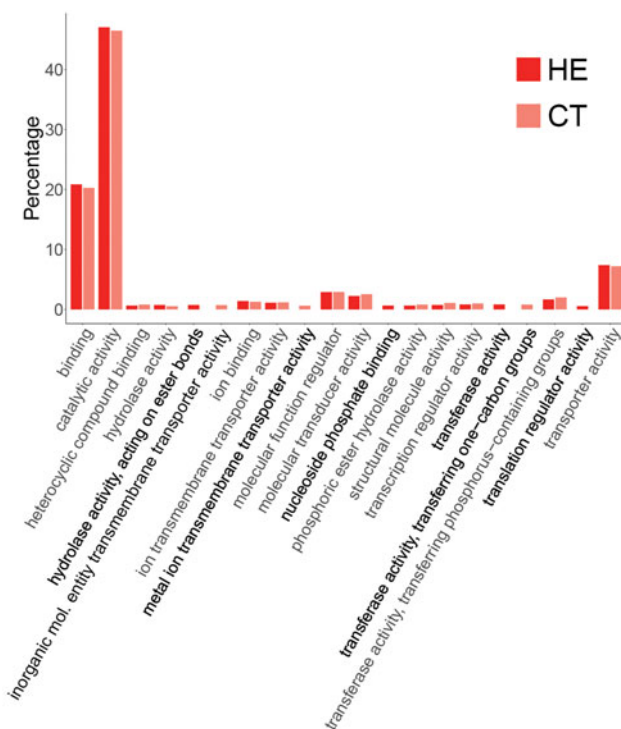


Fig. 8. Molecular function terms at the 99.5th percentile compared between the heated (HE) group (dark red) and the control (CT) group (light red). Expression differences between the two conditions are highlighted in bold.

measurement of the most expressed BPs, CCs and MFs between the HE and CT springtails showed no remarkable divergences, except for a frequency bias of a few GO (Figs 6–8).

The GO enrichment analysis on the annotated DEGs revealed the presence of 39 significant terms, classified as 19 MFs, 15 BPs and 5 CCs (Fig. S1). Overall, the majority of the enriched GOs displayed a noticeable expression difference between the HE and CT samples, except for some that did not show a clear dissimilarity between the two groups, probably due to a similar transcript co-expression. On the other hand, the PFAM enrichment analysis disclosed the presence of five significant domains. Among these, one was considered clearly perturbed because of the limited divergence between the two conditions (Fig. S2).

Discussion

The effects of thermal stress on polar terrestrial arthropods is still a relatively unexplored field of study. Although previous research shows contrasting outcomes on animal survival, micro-environmental conditions apparently play a pivotal role in population prosperity (Bokhorst *et al.* 2008). Nutrients and predators were identified as the two major factors that influence polar invertebrate populations' viability in a situation characterized by rising temperatures (Bokhorst *et al.* 2008). However, the effect of thermal stress at the gene level has never been investigated in detail. In this study, we used a *de novo* transcriptome analysis to identify genes and pathways specifically modulated following mid-term exposure to higher temperatures.

The lack of a reference genome or transcriptome hampered the analysis due to the production of partial and/or fragmented transcripts (Voshall & Moriyama 2018). Nevertheless, the fraction of the annotated transcripts was remarkable, and only a small fraction of contigs remained unassigned. In detail, 71.7% of the transcriptome was annotated in at least one of the considered databases, and 27.6% (12,522) of this was characterized by the three main databases used (Uniprot, OrthoDB and PFAM) (Fig. 1). This suggests that although *C. terranovus* is phylogenetically distant from most model species and especially from those with exhaustive genomic data available, protein families and domains are sufficiently conserved, allowing for solid automatic annotation.

The differential expression analysis revealed that only subtle dissimilarities could be observed between the two conditions, at variance with the fairly high level of inter-individual variability. This latter finding is probably associated with the remarkably plastic gene expression of this taxon, as registered in *C. antarcticus*, where

subgroups of the main population managed to endure long periods of thermal stress better than others, despite the survival rate continuously decreasing (Everatt *et al.* 2013). On the other hand, the mortality rate of both the CT and HE groups in captivity was well below 10%, and some egg clusters of *C. terranovus* were observed at the bottom of the box containers where specimens were reared.

In the comparative analysis of the 99.5th percentile of genes annotated in both conditions, a similar distribution of GO terms was observed, with few exceptions, indicating high intra-condition variability due to transcript expression plasticity. Despite this outcome, gene expression enrichment analysis provided significant insights into the modulation of gene expression following mid-term thermal stress.

Carbohydrate metabolism

GOs related to the sugar catabolic processes (GO:0004034, GO:0004553, GO:0046373, GO:004656), probably induced during glycolytic activity, appear to be underrepresented in HE compared to CT samples. Significantly, the transcripts involved in this process were identified as part of a digestion pathway (e.g. β -galactosidase, β -glucuronidase or aldose-1 epimerase) (Table SIII), probably being induced by an optimal nutritional status in the wild environment. These enzymes are, indeed, parts of pathways involved in carbohydrate hydrolysis or glycolytic activities (e.g. aldose-1 epimerase). However, carbohydrate catabolism is generally a typical defence mechanism adopted by Antarctic arthropods to endure extreme cold temperatures (Teets & Denlinger 2014). In *C. antarcticus*, three main compounds are used as cryoprotectants: trehalose, glucose and glycerol (Elnitsky *et al.* 2008). Despite our CT samples expressing a different carbohydrate catabolism pathway for trehalose and glucose, it is possible to hypothesize that this metabolic pathway may be activated as a response to low temperatures as an alternative mechanism that is parallel or connected to cryoprotectant production. In this respect, subsequent studies are necessary to investigate which cryoprotectant is used in the continental Antarctic springtail *C. terranovus* and to compare its carbohydrate profile with the co-generic species *C. antarcticus*.

Protein degradation

Overall, the HE group springtails showed enhanced protein catabolism (GO:0004021, GO:0004180, GO:0004181, GO:0042853, GO:0070573), as the expression of several peptidase transcripts suggests (e.g. *GPTs* (or *ALTs*), *CPN*, *CPQ*, *CNDP2*, *VASH1* and *DPEPI*, homologous to several model organisms) (Table SIII). Many of these protein degradation genes

are well studied in different animals, but they have never been found in association with thermal stress. As a few examples, the alanine-aminotransferase transcript is a well-known molecular target for cellular necrosis in mammalian serum (Washington & Van Hoosier 2012), but it has never been associated with Arthropoda heat exposure; the several carboxypeptidases and dipeptidase 1 detected herein act in the peptide degradation pathway and are conserved in many animal lineages, but they have never been related to cellular thermal stress. Heat exposure can cause protein damage and misfolding, leading to increased protein turnover (thus possibly explaining the upregulation of genes involved in protein degradation) or rearrangement following heat shock protein (HSP) activation. In addition, these latter molecules play a fundamental role in thermal stress responses and have been detected in all of the studied living organisms, categorized as 1) inducible (e.g. Hsp70) and 2) constitutive (e.g. Hspc70) forms (Roelofs *et al.* 2008). In the Antarctic environment, the larvae of *Belgica antarctica*, the fish *Trematomus bernacchii* and the ciliate *Euplotes focardii* constitutively express such chaperones, which allow for year-round defence against frequent temperature shifts (Rinehart *et al.* 2006, Teets & Denlinger 2014). Similarly, *C. terranovus* specimens subject to mid-term thermal stress seem to activate protein catabolism in the absence of the typical upregulation of HSP production. Indeed, during differential expression analyses, several transcripts annotated as HSPs were revealed to be expressed in the CT samples, as well as in the HE samples. Thus, we can assume that *C. terranovus*, similarly to *B. antarctica* larvae (i.e. the closest phylogenetic Antarctic species whose molecular chaperons have been deeply investigated), expresses HSPs constitutively and not as a response to variable environmental temperature (Rinehart *et al.* 2006). Nevertheless, our findings are based on collembolan individuals sampled during the Antarctic summer (i.e. the warmest season), although it would be of interest to study HSP expression in specimens approaching the coldest season for comparison.

Lipid activity

Lipid metabolism pathways appear to be upregulated following heat treatment. Previous studies indicated that dehydrated animals (i.e. the dipteran *B. antarctica* and the springtail *Folsomia candida*; Timmermans *et al.* 2009, Teets *et al.* 2012) display downregulation of fatty acid metabolism, similar to what happened in the *C. terranovus* CT samples. Our samples were collected in the wild in the absence of detailed humidity monitoring, and therefore it is not possible to assess whether this downregulation might be the consequence of dehydration stress. Moreover, in our analysis, no GO

linked to a similar description was found in the CT group. Nonetheless, Thorne *et al.* (2010) found a similarly enhanced lipid metabolism in the Antarctic benthonic fish *Harpagifer antarcticus* after thermal exposure. These authors proposed that similar pathways are enhanced in order to rearrange cellular membranes after a temperature shock. Heated *C. terranovus* specimens showed both fatty acid synthase and lipase to be differentially expressed (Table SIII), indicating that the metabolism of these molecules was overexpressed. While *C. terranovus* is not phylogenetically related to *H. antarcticus*, they both upregulate these same pathways, possibly indicating an evolutionary convergence in cellular lipid rearrangement to endure heat exposure. However, more evidence should be gathered in order to assess this correspondence, and so more in-depth investigations are required.

Spermatid development

Despite evidence highlighting that aberrant environmental conditions can cause a reduction in the production of gametes, a process that is by no means restricted to springtails (Dallai 2014), stressed *C. terranovus* specimens did not show this expected reduction. Indeed, the stressed group was characterized by increased expression of genes related to a GO term designated for spermatid development (GO:0007286) and of several transcripts annotated as alanine-aminotransferase, the activity of which is highly increased in *Drosophila nigromelanica* sperm and ovarian cells during their development (Schneider & Chen 1981). Significantly, several transcripts related to gametogenesis were discovered in the treated animals (homologs to *Ccdc63*, *SPAG6* and *poe* studied in model organisms), while, on the contrary, increased expression of a male infertility gene (*CDYL*) was observed in the CT group (Table SIII). This may suggest that the extreme environmental conditions simulated for HE samples did not negatively influence sexual development but, on the contrary, fostered spermatid production. This process could be related to the breeding conditions (inside a box container), where the proximity of individuals may enhance chemical signal production to promote aggregation and therefore stimulate sexual behaviour and spermatid production. However, the link between thermal exposure and sexual development must be clarified in the future in order to determine whether or not this long-term condition impacts sexual activity in continental Antarctic collembolans.

Conserved protein domains

The PFAM enrichment analysis disclosed a limited number of outcomes regarding the conserved protein

domains in the *C. terranovus* specimens. In fact, a pentapeptide family domain and a protein of unknown function were found to be overrepresented in the HE group. The former is an amino acid repeat found in a number of different protein families and therefore it cannot be associated with any specific target.

Similarly, a domain of unknown function was more abundant in the CT samples. This was found to be associated to the YLP motif in several *Drosophila* proteins, and apart from the probable interaction of YLP motifs with dozens of tyrosine kinase enzymes, no further insights can be gained.

In addition, a mucin-like protein domain was characteristic of the CT groups. These proteins, which are well characterized in vertebrates, were found to be synthesized in several *Drosophila* tissues as well (i.e. salivary glands, midgut, Malpighian tubules, etc.) and often in relation with development (Syed *et al.* 2008). An additional PFAM-enriched term, whose expression appeared to be disturbed but was slightly more abundant in the CT group, was marked as 'insect cuticle protein'. In line with the mucin-like protein domain, it may indicate an association with tissue development. These two latter pieces of evidence, circumstantial as they are, suggest an increase in body development in the CT group that may be related to springtail wellness in wild environments, where development can proceed with regularity, at variance with individuals in stressed conditions. Nevertheless, these preliminary findings must be deepened in further investigations comparing how springtail development and hormonal levels vary between stressed and wild specimens.

Conclusions

Although Antarctic marine animal groups have been investigated in some detail in order to understand how gene expression is modified during environmental changes, data on Antarctic terrestrial fauna are scant. In this study, we analysed (to our knowledge for the first time) the transcriptome of an Antarctic collembolan in a situation of prolonged heat exposure. We showed that control specimens display an overrepresentation of pathways such as carbohydrate catabolism, reduced lipid metabolism activity and enhancements in tissue development, which have been also found to be modulated in other model organisms under experimental conditions. Moreover, we described some biological mechanisms that appear to be activated in thermally stressed *C. terranovus* individuals, where protein catabolism and fatty acid activity are promoted probably in response to protein misfolding and fresh nutrient availability. Surprisingly, we discovered that spermatid development appears to be improved in

animals subject to heat stress, presumably because, during the experimental breeding condition, the proximity of individuals inside the box containers promoted chemical exchange favouring animal aggregation and, hence, a sexual response.

The low mortality rate observed in the stressed group, the observation that the animals are able to lay eggs even in unusually high-temperature conditions and the limited extent to which the animals from the HE and CT groups differed in terms of gene expression would lead to the hypothesis that the testing regime (18°C for 20 days) was not stringent enough to observe significant differences. Nevertheless, this temperature is substantially higher than what these animals usually experience in their natural conditions, where average daily temperatures are stably < 10°C, even during the Antarctic summer months (data not shown). This would support the notion that these animals are indeed characterized by ample phenotypic plasticity, which may allow them to survive even significant thermal stress. More severe stressing regimes will be adopted in the future in order to further test this hypothesis over prolonged durations.

In summary, the gene expression of this continental Antarctic springtail should be studied in more detail to: 1) assess which patterns of gene expression are activated during exposure to lower temperatures; 2) evaluate which transcripts are enhanced in short-term heat exposure and to determine whether HSPs are enhanced during this process (as happens in several other groups) or whether these chaperones are constitutively expressed as in *B. antarctica*; and 3) evaluate the extent to which the Antarctic springtail phenology is compromised during long-term heat exposure.

Acknowledgements

We wish to thank Professor Bettine van Vuuren and an anonymous reviewer for their constructive and helpful comments.

Financial support

This study was funded by the Italian National Antarctic Program (PNRA), grant number: PNRA16_00234.

Author contributions

AC collected the samples and supervised the experiment together with PGG; CL performed the experiments, CM extracted the RNA and built the Illumina libraries, CC and SG analysed the bioinformatic data and CC together with FN and AC wrote the paper. All of the authors participated in drafting and critically revising the paper.

Supplemental material

A supplemental methods section including three supplemental bioinformatic codes (Protocols S1–S3), two figures (Figs S1 & S2) and three tables (Tables SI–SIII) will be found at <https://doi.org/10.1017/S0954102021000195>.

References

- BILYK, K.T., VARGAS-CHACOFF, L. & CHENG, C.H.C. 2018. Evolution in chronic cold: varied loss of cellular response to heat in Antarctic notothenioid fish. *BMC Evolutionary Biology*, **18**, 10.1186/s12862-018-1254-6.
- BOKHORST, S., HUISKES, A., CONVEY, P., VAN BODEGOM, P.M. & AERTS, R. 2008. Climate change effects on soil arthropod communities from the Falkland Islands and the Maritime Antarctic. *Soil Biology and Biochemistry*, **40**, 10.1016/j.soilbio.2008.01.017.
- BURNS, G., THORNE, M.A.S., HILLYARD, G., CLARK, M.S., CONVEY, P. & WORLAND, M.R. 2010. Gene expression associated with changes in cold tolerance levels of the Antarctic springtail, *Cryptopygus antarcticus*. *Insect Molecular Biology*, **19**, 10.1111/j.1365-2583.2009.00953.x.
- BUSHMANOVA, E., ANTIPOV, D., LAPIDUS, A. & PRJIBELSKI, A.D. 2019. *maSPAdes*: a *de novo* transcriptome assembler and its application to RNA-Seq data. *GigaScience*, **8**, 10.1093/gigascience/giz100.
- CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T.L. 2009. *BLAST+*: architecture and applications. *BMC Bioinformatics*, **10**, 10.1186/1471-2105-10-421.
- CARAPPELLI, A., LEO, C. & FRATI, F. 2017. High levels of genetic structuring in the Antarctic springtail *Cryptopygus terranovus*. *Antarctic Science*, **29**, 10.1017/S0954102016000730.
- CARAPPELLI, A., FANCIULLI, P.P.P., FRATI, F. & LEO, C. 2019. Mitogenomic data to study the taxonomy of Antarctic springtail species (Hexapoda: Collembola) and their adaptation to extreme environments. *Polar Biology*, **42**, 10.1007/s00300-019-02466-8.
- CARUSO, T., HOGG, I.D., CARAPPELLI, A., FRATI, F. & BARGAGLI, R. 2009. Large-scale spatial patterns in the distribution of Collembola (Hexapoda) species in Antarctic terrestrial ecosystems. *Journal of Biogeography*, **36**, 10.1111/j.1365-2699.2008.02058.x.
- COLLINS, G. E., HOGG, I.D., CONVEY, P., BARNES, A.D. & McDONALD, I.R. 2019. Spatial and temporal scales matter when assessing the species and genetic diversity of springtails (Collembola) in Antarctica. *Frontiers in Ecology and Evolution*, **7**, 10.3389/fevo.2019.00076.
- CONVEY, P. & PECK, L.S. 2019. Antarctic environmental change and biological responses. *Science Advances*, **5**, 10.1126/sciadv.aaz0888.
- CONVEY, P., PUGH, P.J., JACKSON, C., MURRAY, A.W., RUHLAND, C.T., XIONG, F.S. & DAY, T.A. 2002. Response of Antarctic terrestrial microarthropods to long-term climate manipulations. *Ecology*, **83**, 10.1890/0012-9658(2002)083[3130:ROATMT]2.0.CO;2.
- DALLAI, R. 2014. Overview on spermatogenesis and sperm structure of Hexapoda. *Arthropod Structure & Development*, **43**, 10.1016/j.asd.2014.04.002.
- DEUTSCH, C.A., TEWKSBURY, J.J., HUEY, R.B., SHELDON, K.S., GHALAMBOR, C.K., HAAK, D.C. & MARTIN, P.R. 2008. Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 10.1073/pnas.0709472105.
- ELNITSKY, M.A., BENOIT, J.B., DENLINGER, D.L. & LEE, R.E. JR. 2008. Desiccation tolerance and drought acclimation in the Antarctic collembolan *Cryptopygus antarcticus*. *Journal of Insect Physiology*, **54**, 10.1016/j.jinsphys.2008.08.004.
- EVERATT, M.J., CONVEY, P., WORLAND, M.R., BALE, J.S. & HAYWARD, S.A.L. 2013. Heat tolerance and physiological plasticity in the Antarctic collembolan, *Cryptopygus antarcticus*, and mite, *Alaskozetes antarcticus*. *Journal of Thermal Biology*, **38**, 10.1016/j.jtherbio.2013.02.006.

- EVERATT, M.J., CONVEY, P., WORLAND, M.R., BALE, J.S. & HAYWARD, S.A.L. 2014. Are the Antarctic dipteran, *Eretmoptera murphyi*, and Arctic collembolan, *Megaphorura arctica*, vulnerable to rising temperatures? *Bulletin of Entomological Research*, **104**, 10.1017/S0007485314000261.
- EWELS, P., MAGNUSSON, M., LUNDIN, S. & KÄLLER, M. 2016. *MultiQC*: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, **32**, 10.1093/bioinformatics/btw354.
- FANCIULLI, P.P., SUMMA, D., DALLAI, R. & FRATI, F. 2001. High levels of genetic variability and population differentiation in *Gressittacantha terranova* (Collembola, Hexapoda) from Victoria Land, Antarctica. *Antarctic Science*, **13**, 10.1017/S0954102001000360.
- FU, L., NIU, B., ZHU, Z., WU, S. & LI, W. 2012. *CD-HIT*: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**, 10.1093/bioinformatics/bts565.
- GONZÁLEZ-ARAVENA, M., CALFIO, C., MERCADO, L., MORALES-LANGE, B., BETHKE, J., DE LORGERIL, J. & CÁRDENAS, C.A. 2018. HSP70 from the Antarctic sea urchin *Sterechinus neumayeri*: molecular characterization and expression in response to heat stress. *Biological Research*, **51**, 10.1186/s40659-018-0156-9.
- GRABHERR, M.G., HAAS, B.J., YASSOUR, M., LEVIN, J.Z., THOMPSON, D.A., AMIT, I., *et al.* 2011. *Trinity*: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology*, **29**, 10.1038/nbt.1883.
- MACMANES, M.D. 2018. The *Oyster River Protocol*: a multi-assembler and kmer approach for *de novo* transcriptome assembly. *PeerJ*, **6**, 10.7717/peerj.5428.
- MCGAUGHRAN, A., STEVENS, M.I., HOGG, I.D. & CARAPELLI, A. 2011. Extreme glacial legacies: a synthesis of the Antarctic springtail phylogeographic record. *Insects*, **2**, 10.3390/insects2020062.
- MICHAUD, R.M., BENOIT, J.B., LOPEZ-MARTINEZ, G., ELNITSKY, M.A., LEE, R.E. & DENLINGER, D.L. 2008. Metabolomics reveals unique and shared metabolic changes in response to heat shock, freezing and desiccation in the Antarctic midge, *Belgica antarctica*. *Journal of Insect Physiology*, **54**, 10.1016/j.jinsphys.2008.01.003.
- PATRO, R., DUGGAL, G., LOVE, M.I., IRIZARRY, R.A. & KINGSFORD, C. 2017. *Salmon* provides fast and bias-aware quantification of transcript expression. *Nature Methods*, **14**, 10.1038/nmeth.4197.
- REED, A.J. & THATJE, S. 2015. Long-term acclimation and potential scope for thermal resilience in Southern Ocean bivalves. *Marine Biology*, **162**, 10.1007/s00227-015-2752-3.
- RINEHART, J.P., HAYWARD, S.A., ELNITSKY, M.A., SANDRO, L.H., LEE, R.E. & DENLINGER, D.L. 2006. Continuous up-regulation of heat shock proteins in larvae, but not adults, of a polar insect. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 10.1073/pnas.0606840103.
- ROBERTSON, G., SCHEIN, J., CHIU, R., CORBETT, R., FIELD, M., JACKMAN, S.D., *et al.* 2010. *De novo* assembly and analysis of RNA-seq data. *Nature Methods*, **7**, 10.1038/nmeth.1517.
- ROELOFS, D., AARTS, M.G.M., SCHAT, H. & VAN STRAALLEN, N.M. 2008. Functional ecological genomics to demonstrate general and specific responses to abiotic stress. *Functional Ecology*, **22**, 10.1111/j.1365-2435.2007.01312.x.
- SCHNEIDER, M. & CHEN, P.S. 1981. l-alanine aminotransferase in *Drosophila nigromelanica*: isolation, characterization and activity during ontogenesis. *Insect Biochemistry*, **11**, 10.1016/0020-1790(81)90056-1.
- SEPPEY, M., MANNI, M. & ZDOBNOV, E.M. 2019. *BUSCO*: assessing genome assembly and annotation completeness. In KOLLMAR, M., ed. *Gene prediction: methods and protocols*. Berlin: Springer, 227–245.
- SYED, Z.A., HÄRD, T., UV, A. & VAN DIJK-HÄRD, I.F. 2008. A potential role for *Drosophila* mucins in development and physiology. *PLoS One*, **3**, 10.1371/journal.pone.0003041.
- TARAZONA, S., FURIÓ-TARÍ, P., TURRÀ, D., PIETRO, A.D., NUEDA, M.J., FERRER, A. & CONESA, A. 2015. Data quality aware analysis of differential expression in RNA-seq with *NOISeq R/Bioc* package. *Nucleic Acids Research*, **43**, 10.1093/nar/gkv711.
- TEETS, N.M. & DENLINGER, D.L. 2014. Surviving in a frozen desert: environmental stress physiology of terrestrial Antarctic arthropods. *Journal of Experimental Biology*, **217**, 10.1242/jeb.089490.
- TEETS, N.M., PEYTON, J.T., COLINET, H., RENAULT, D., KELLEY, J.L., KAWARASAKI, Y., *et al.* 2012. Gene expression changes governing extreme dehydration tolerance in an Antarctic insect. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 10.1073/pnas.1218661109.
- THORNE, M.A.S., BURNS, G., FRASER, K.P.P., HILLYARD, G. & CLARK, M.S. 2010. Transcription profiling of acute temperature stress in the Antarctic plunderfish *Harpagifer antarcticus*. *Marine Genomics*, **3**, 10.1016/j.margen.2010.02.002.
- TIMMERMANS, M.J., ROELOFS, D., NOTA, B., YLSTRA, B. & HOLMSTRUP, M. 2009. Sugar sweet springtails: on the transcriptional response of *Folsomia candida* (Collembola) to desiccation stress. *Insect Molecular Biology*, **18**, 10.1111/j.1365-2583.2009.00916.x.
- VOSHALL, A. & MORIYAMA, E.N. 2018. Next-generation transcriptome assembly: strategies and performance analysis. In ABDURAKHMONOV, I.Y., ed. *Bioinformatics in the era of post genomics and big data*. London: IntechOpen, 15–36.
- WASHINGTON, I.M. & VAN HOOSIER, G. 2012. Clinical biochemistry and hematology. In SUCKOW, M.A., STEVENS, K.A. & WILSON, R.P., eds. *The laboratory rabbit, guinea pig, hamster, and other rodents*. Cambridge, MA: Academic Press, 57–116.
- WAUCHOPE, H.S., SHAW, J.D. & TERAUDS, A. 2019. A snapshot of biodiversity protection in Antarctica. *Nature Communications*, **10**, 10.1038/s41467-019-08915-6.