Cold hardening processes in the Antarctic springtail, *Cryptopygus antarcticus*: clues from a microarray.

Jelena Purać[1,2#], Gavin Burns[1#], Michael A.S. Thorne[1#], Gordana Grubor-Lajšić[2#], M. Roger Worland[1#] and Melody S. Clark[1#*]

[1]British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge, CB3 0ET, UK

[2]University of Novi-Sad, Faculty of Sciences, Trg Dositeja Obradovića 3, 21000 Novi Sad, Republic of Serbia.

[*]Author for correspondence: Melody S. Clark, British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge, CB3 0ET, UK. Email: mscl@bas.ac.uk

[#]All authors contributed equally to this work.

Abstract

The physiology of the Antarctic microarthropod, *Cryptopygus antarcticus*, has been well studied, particularly with regard to its ability to withstand low winter temperatures. However, the molecular mechanisms underlying this phenomenon are still poorly understood. 1180 sequences (Expressed Sequence Tags or ESTs) were generated and analyzed, from populations of *C. antarcticus*. This represents the first publicly available sequence data for this species. A sub-set (672 clones) were used to generate a small microarray to examine the differences in gene expression between summer acclimated cold tolerant and non-cold tolerant springtails. Although 60% of the clones showed no sequence similarity to annotated genes in the datasets, of those where putative function could be inferred via database homology, there was a clear pattern of up-regulation of structural proteins being associated with the cold tolerant group. These structural proteins mainly comprised cuticle proteins and provide support for the recent theory that summer SCP variation within Collembola species could be a consequence of moulting, with moulting populations having lowered SCPs.

Keywords

Cold tolerance, Collembolan, freeze avoidance, SCP, moulting

Introduction

Low temperature biology, in particular the mechanisms by which organisms survive extreme low temperatures, is of intense interest not only to ecologists, but also in a number of applied medical fields (Fuller, 1999). One of the most amenable and studied group of organisms is the arthropoda where such physiological processes are well documented (Salt, 1961; Lee and Denlinger, 1991; Sømme, 1995; Sinclair et al, 2003a). These organisms have evolved two distinct strategies to survive sub-zero temperatures termed freeze tolerance and freeze avoidance (Cannon and Block, 1988; Block, 1990 see also Bale, 1993 for more ecologically-refined schemes of classificiation). A third mechanism, protective dehydration, has more recently been described but is perhaps less common. (Worland, 1996; Holmstrup et al. 2002). Delineation into these groups is often performed using the measurement of the temperature at which the insect's body fluids freeze, normally termed the supercooling point (SCP).

Freeze avoidance is the main strategy used by Antarctic terrestrial microarthropods (Block 1990, 1991; Cannon and Block, 1988; Convey, 1996), which routinely have to survive over wintering temperatures below -20°C. For these organisms the SCP is equal to their lower lethal temperature. This is not the case for all arthropods as some have been shown to die before they freeze (chill susceptible) (Bale, 2002). Other factors such as the cooling rate can affect the measured SCP and the lower lethal temperature (Worland, 2005). Cold tolerance also varies seasonally, with higher mean values in the summer for active feeding animals than for winter acclimated, cold hardened individuals (Cannon and Block, 1988). Seasonal cold-hardening involves a

number of physiological and biochemical processes that slowly increase supercooling ability and enhance survival at low temperatures. Such mechanisms include the removal or deactivation of ice nucleating agents, accumulation of cryoprotectants and thermal hysteresis proteins (Knight & Duman, 1986; Sømme L, 1982; 1995; Zachariassen, 1992). This gives rise to a bimodal distribution of SCPs with a high group (less cold tolerant) and a low group (more cold tolerant) (Cannon, 1983; Schenker, 1984; Rothery and Block, 1992; Worland & Convey, 2001, Block, 1982, 1984; van der Woude 1987, Sinclair et al 2003b).

This is particularly evident during the growing season in the summer when, perhaps surprisingly, a proportion of the population still exhibit low SCPs. This phenomenon has been extensively studied, but its exact nature remains unclear, i.e. whether it is an adaptation to environmental conditions or part of the animals natural lifecycle. It has been correlated in some studies with the presence or absence of gut contents (Sømme & Conradi-Larsen, 1977; Young & Block, 1980; Sømme, 1981; Block, 1982; Cannon, 1983; Cannon & Block, 1988; Rothery & Block, 1992). It has also been suggested that some insects retain a low SCP during the summer, providing insurance against unexpected freezing temperatures (Chown and Klok, 1998). In inland continental areas where there is a year-round risk of sub-zero temperatures, springtails show permanent cold tolerance even at the "height" of summer (Sømme, 1985). Recently, the effect of moulting on SCP has been observed in Collembola (Worland, 2005; Worland et al, 2006, Worland and Convey, 2008). Moulting might be expected to depress the SCP, because in Collembola the mid-gut and its entire contents are shed during moulting (Thibaud, 1968) resulting in the expulsion of potential ice nucleators

present in the animals gut. In this case cold tolerance can be viewed as a natural part of the animals life cycle rather than an adaptation *per se*.

In general, species typified by a bimodal SCP distribution show a population shift from high to low SCPs as winter approaches (Cannon and Block, 1988), but SCPs in some Antarctic species can vary or be altered diurnally (Sinclair, 2003b) or even within hours (Worland and Convey, 2001) in response to changing environmental conditions. This "rapid cold hardening" is almost certainly a critical factor in the ability of Antarctic terrestrial arthropods to survive variable and unpredictable summer temperatures (Convey, 1997). Although the molecular processes involved in rapid cold hardening and the relationship between seasonal and rapid cold hardening have yet to be characterised, they are thought to be distinct mechanisms (Worland and Convey, 2001).

This study focusses on the overwinter survival strategy of the freeze-avoiding springtail *Cryptopygus antarcticus* Willem (Collembola, Isotomidae). This is one of the most abundant and widespread terrestrial arthropod in the Maritime Antarctic and Sub-Antarctic regions (Block, 1982; Usher and Booth, 1986; Convey and Smith, 1997) and has been one of most studied Antarctic microarthropods over the last thirty years (e.g. Ewing, 1922; Gressitt, 1967; Janetschek, 1967; Tilbrook, 1970, 1977; Cannon and Block, 1988; Worland and Convey, 2001). It shows a bimodal SCP distribution with distinct High (-8ºC to –10ºC) and Low (–18ºC to –30ºC) groups (Sømme and Block, 1982, Rothery and Block, 1992; Worland and Convey, 2001). Studies on *C. antarcticus* have contributed greatly towards our understanding of cold tolerance, but to date neither this species, nor any other polar arthropod, has been the

subject of sequence-based analyses. In order to redress this deficit and gain an insight into the genome of *C. Antarcticus,* 1180 ESTs were generated and a sub-set of clones used to produce a small microarray. The aim was to use novel tools provided by microarray expression analysis to provide an insight into the underlying biochemical mechanisms behind the summer bimodal SCP populations and increase our understanding of the nature of environmental adaptation in these arthropods in one of the world's most challenging environments.

Materials and methods

Sample collection and preparation:

Antarctic springtails (*Cryptopygus antarcticus*) were collected from wet moss (*Sanionia uncinata* (Hedw.) during the austral summer of 2005 at the British Antarctic Survey's research station at Rothera Point, Adelaide Island (67'34'S, 66'8'W). The SCPs of the animals were measured using a differential scanning calorimeter (DSC) (Mettler Toledo DSC 820). The boundary temperature dividing high and low group SCPs was defined as -15ºC (Worland and Convey, 2001). Animals which survived cooling to -15°C at 1°C/minute were designated as Low Group (LG) animals (SCP below -15°C), and those which died were designated as the High Group (HG, SCP above -15°C). All groups of animals were rapidly frozen in liquid nitrogen and stored at -80°C until required. Live animals maintained with food in a +2°C growth cabinet were used as controls.

cDNA libraries and expressed sequence tag (EST) sequencing:

ESTs were generated from a mixed population of animals. Total RNA was extracted by homogenizing samples in Tri Reagent (Sigma) according to manufacturers instructions. 5µg Poly A+ mRNA was prepared from total RNA using the Oligotex mRNA Kit-Midi kit (Qiagen) according to manufacturers instructions. Directional plasmid cDNA libraries were produced using the pBluescript®II XR cDNA Library Construction Kit (Stratagene). The inserts from each isolated clone in the cDNA library were amplified via PCR, performed in a reaction mixture of 20 µl containing: 1X NH_4 reaction buffer (Bioline), 0.125 µM of each M13 primer (M13 reverse: 5'-AACAGCTATGACCATGAT-3', M13 forward: 5'-GTAAAACGACGGCCAG-3'), 0.084 mM of dNTPs (Bioline), 1.5 mM $MgCl_2$ and 0.5 U of Taq Polymerase (Bioline). After an initial denaturation of 2 min. at 96°C, the thermal cycling consisted of 30 cycles of denaturation at 96°C for 20s, primer annealing at 49°C for 20s and extension at 72°C for 3 min. In the final cycle, the extension was prolonged for 5 min. This reaction was also repeated using amine terminated M13 primers to produce cDNA products to be printed on the microarray. After PCR, all products were diluted with 30µl of sterile distilled water. Enzymatic cleanup of the amplified cDNA was performed using 5µl of non-amine terminated PCR product, 0.4U Shrimp Alkaline Phosphatase (GE Healthcare), 0.6U Exonuclease I (GE Healthcare) in a total volume of 6µl containing Shrimp Alkaline Phosphatase dilution buffer. The reactions were incubated at 37ºC for 30 min and then at 80°C for 10 min to inactivate the enzymes. This 6µl reaction was then used as a template in a sequencing reaction by adding 4 µl of ET-terminator pre-mix (GE Healthcare) and 0.5 µM M13long reverse sequencing primer (5'-AACAGCTATGACCATGATTACG-3'). Thermal cycling conditions were 26 cycles at 20s at 95°C and 2 min and 20 s at 60°C. Sequence

reactions were ethanol precipitated and run on a MegaBACE 1000 capillary sequencer (GE Healthcare) using standard sequence filters and a 100 min run time.

Sequence data analysis:

Sequence fasta files were processed using the script Trace2dbest (Parkinson et al, 2004), which incorporated the phred (Ewing and Green, 1998a; 1998b) and crossmatch (P. Green, unpublished) programmes. A minimum cut-off value of 100bp was applied after quality control processing for sequence database searching and for generating the submission file for dbEST (Boguski et al, 1993) (Accession numbers, dbEST: 55137170-55138349, Genbank: FF278135-FF279314). Tgicl (Pertea et al, 2003) was used for clustering the fasta files, incorporating quality scores. The clusters were database searched using Blastx (Altschul et al, 1997) against the Uniprot/Swissprot and Uniprot/Trembl databases (The Uniprot Consortium, 2007), with matches annotated for all scores with an expect score less than $1e^{-10}$. Sequences with a database match were then further annotated using GO and GOSLIM (The Gene Ontology Consortium, 2000) to the full depth of their significant Blast matches.

Microarray construction and hybridization:

A microarray was produced from seven 96 well clone plates with the highest insert ratios from the 1124 EST bacterial clones to give a total of 672 clones printed in duplicate. The Stratagene SpotReport Alien cDNA Array Validation System was included on the microarray. The microarray was probed with three populations of *Cryptopygus:* control animals (mixed population), high group animals and low group animals. Prior to microarray printing, the probes were prepared as described in Lyne et al, 2003 and printed using a Genetix Q-array robot. Microarrays were incubated in

a saturated sodium chloride humid chamber at room temperature overnight and post-processed according to the manufacturers instructions. PCR amplified ds cDNA probes for all groups were prepared using the protocol described in Petalidis et al (2003) with the addition of 1µl Stratagene Alien mRNA spikes 1-6 into the initial total RNA. The Cy5 and Cy3 samples were pooled together and 6µg poly dA and 6µg yeast tRNA (Sigma) were added as blocking agents. The samples were ethanol precipitated. Labelled targets were resuspended in 20µl of hybridisation buffer (40% formamide, 5x SSC, 5x Denhardt's solution, 1mM sodium pyrophosphate, 50mM Tris pH 7.4, 0.1% SDS), denatured at 95°C for 5 mins and allowed to cool at room temperature for 5 mins prior to hybridization. The labelled targets for each of the two experimental animal groups were hybridized to the microarray with labelled control group target as a reference. Hybridizations were performed in a humidified incubator at 49ºC overnight. The microarrays were then washed in 2x SSC, 0.1% SDS for 15 min, followed by 2x SSC for 5 min. and finally in 0.1x SSC for 5 min, all at RT. After washing, slides were dried by centrifugation and scanned on a Molecular Devices GenePix 4100 microarray scanner. The design entailed five biological replicates and three technical replicates, one of which was a dye swop, per treatment.

Microarray data analysis:

The images were analysed using the GenePix 6.0 software (Molecular Devices). After gridding and segmentation, visual inspection was used to flag and exclude anomalous spots. The R (R Development Core Team, 2005) Limma package (Smyth, 2004, 2005; Richie et al, 2007; Smyth and Speed, 2003; Smyth et al, 2003) was used for background subtraction, normalisation and to determine the differentially expressed clones at an FDR adjusted p-value of 0.01. GO enrichment was determined

by a proprtion test, at a p-value of 0.01, between the number of clones representing a GO term on the chip compared to the number of differentially expressed clones representing the same GO term in a given list. The array design is housed at ArrayExpress, accession number: A-MEXP-1128, and the experiments: E-MEXP-1569.

Q PCR:

To validate the microarray results, 3 clones were chosen for Q-PCR analysis: CR_C02P0_01G10: Control: CF: CGCTCTTCATTTAATGGGGTT, CRev: AAACGACATTCGAGTACTTCAT; CHK1: CHKF: AGGGAGCCTGAGAAACGGCT CHKRev: CTGGCACCAGACTTGCCCT; Cuticle7F: CAAGATGAACTTGGTCAAGCAT, cuticle7Rev: CTGATAAGAGCATCAGCATGA. The RSq and efficiency values for each of these primer sets was: 0.970/115.5%; 0.991/150.8% and 0.990/102.2% respectively. All genes were amplified using specific primers, Brilliant SYBR® Green QPCR Master Mix (Stratagene) and an MX3000P Q-PCR machine (Stratagene). PCR conditions were as follows: 95°C 10 minutes, 40 cycles of 95°C 30 seconds, 60°C 1 minute and 72°C for 45 seconds with a final dissociation curve step as per manufacturers recommendations. The plate set-up for each Q-PCR experiment consisted of the 3 animal treatments (control, High and Low Groups) amplified for all 3 primer pairs in triplicate. The whole experiment was repeated in duplicate. Primers were validated and the results analysed as described in Clark et al (2008) using the method of Pfaffl (2001) which incorporates the efficiency of the primers as a factor in the equation.

Results

EST analysis

Sequence data was generated for 1180 ESTs. Clustering produced 777 putative transcripts, of which almost 60% had no annotation when searched against the TrEMBL and SwissProt databases (Table 1). The two largest clusters produced matches against insect mitochondrial genomes (Table 2), with the smaller clusters producing matches against genes involved in energy production (ATP synthase, ADP/ATP translocase, Cytochrome C oxidase) and skeletal functioning (Troponin). GOSLIM analysis of the EST clones indicates that the RNA was populated by metabolically active genes, with the IDs for cellular process, transport, response to stimulus, regulation of biological process and metabolic process predominating the listings (Table 3).

Microarray analysis

672 clones were chosen for the array, of which 214 clones (40%) had BLAST matches against the Swiss-Prot/TrEMBL databases (Table 1). The results of the clustering mirrored those for the ESTs with the largest cluster (of 71 clones) producing matches against the giant springtail mitochondrial sequence and the second largest cluster (7 clones) matching an ADP/ATP translocase.

Clones within 29 clusters were significantly upregulated (Table 4) in the Low Group when compared to the control sample. Of these, 18 clusters showed no significant sequence similarity against the databases. It was possible to assign at least some functionality to 11 of the clusters, with cuticle and exoskeleton-related proteins predominating (3 out of 11 clusters) and indeed this was the only overarching

description to be duplicated within this gene table. Q-PCR on one cuticle clone verified that this was upregulated in the Low Group compared to the control animals (by 1.3 fold). It should be noted that there were multiple clusters for cuticle proteins, representing a multigene family, only one of these was surveyed by Q-PCR. When the High Group samples were compared to the controls, clones within 6 clusters were significantly up-regulated (Table 5). All had associated Blast data, with 2 out of the 6 (33%) being identified as putatively involved in the respiration and energy metabolism pathways (NADH dehydrogenase and Cytochrome c oxidase). Several of the up-regulated clusters occurred in common between both the Low and the High Group (CHK1: checkpoint homologue, NADH dehydrogenase, secretory protein and senescence associated protein), but it should be remembered that these groups were each compared to controls. Q-PCR results verified that CHK-1 was upregulated by 2.5 fold in the High Group compared to control animals. When the Low and High Group were directly compared, 30 clusters were significantly up-regulated, of which 13 had putative function assigned via Blast sequence similarity searching (Table 6). Structural protein genes predominated with 38% either associated with the cuticle or muscle and a further 23% involved in membrane trafficking.

When the results are considered in terms of which overarching functions are proportionally more up regulated in one sample compared to another, then production of cuticle protein (Low Group) and protein kinase activity (High Group) (Table 7) are highlighted. When the results of the GO annotation of the Low Group compared to the High Group were analysed, then cuticle proteins predominate along with hydrolase activity, which is coupled to catalysing transmembrane movement of substances.

Discussion

These experiments detail the first expression sequence data produced for the Antarctic Collembola *C. antarcticus*. Over half of the EST clones had no significant matches when searched against the two different databases (TrEMBL and SwissProt) and therefore it was not possible to even putatively assign function to 60% of the data (Table 1). This situation is not unusual when producing sequence data from a non-model organism. Although *Drosophila melanogaster* was fully sequenced in 2000 (Adams et al, 2000), in the current build of ensembl, still only 33.8% of the protein coding gene complement have been characterised, this figure decreases dramatically when considering the more recently targeted insects *Anopheles gambiae* and *Aedes aegypti* (10.5% and 4.9% of protein coding genes respectively) (www.ensembl.org). The vast majority of data in the public databases is of vertebrate origin and with the emphasis on data-mining for medical purposes, these taxa are where most of the functional annotations lie. Therefore putative functionality of genes in non-model organisms initially has to be drawn from vertebrates by analogy and direct sequence similarity. This clearly is only possible if the genes are highly conserved (many are not, c.f. Family 2 of G-protein couple receptors between vertebrates and invertebrates (Cardoso et al, 2006) and even then with high sequence similarity, functionality may not be conserved (c.f. calcitonin between mammals and fish (Clark et al, 2002)).

However, there was sufficient annotation of this dataset to be able to draw some conclusions about the nature of summer-acclimated cold tolerance in these organisms (Tables 4-7). A number of genes/clones were up-regulated in both groups when

compared to controls, such as NADH, cytochrome c and CHK1: a checkpoint homologue involved in the cell cycle, indicating that both sets of animals were metabolically active. However the relative contribution of these clones to each of the upregulated datasets varies between the High and the Low Group. The High Group comprises almost exclusively of these types of genes (Table 5). This result combined with that of the proportionally increased GO annotation (Table 7), which highlighted serine/threonine kinases, which are largely involved in cell signalling processes indicated a higher level of metabolic activity in this sample of animals. This is in agreement with the general understanding, that the High Group contains representatives of actively growing and reproducing summer animals, in a period which is characterised by increased food consumption and energy demand.

The Low Group, and indeed the Low to High comparison, was dominated by cuticle proteins (Table 4 and 6) and this was again reflected in the proportional GO annotation results (Table 7). This was accompanied by genes involved in membrane transport, a function, which links to secretion and production of new cuticle. Interestingly one of the major genes identified in the Low Group was a 10kDa secretory protein from the Mono Lake bird tick, the function of which is unknown, but was isolated from a salivary gland proteome project (Mans et al, 2008). The second clone of the list from the Low Group showed sequence similarity to the cupin superfamily. These proteins are a diverse family involved in isomerase and epimerase activities involved in the modification of cell walls in bacteria, to non-enzymic storage proteins in plant seeds and transcription factors linked to congenital baldness in mammals. This is possibly the widest range of biochemical functions of any superfamily described to date (Dunwell et al, 2001) and so whilst one of the highest

upregulated clones in the Low Group shows homology to a cupin domain, the protein could do anything, no predictive functions are possible. Clearly an interesting candidate for future studies.

Taking a broad overview of the results, they correlate with the recent research in Collembola (Worland, 2005; Worland et al, 2006, Worland and Convey, 2008) suggesting that moulting could be a process during which SCP is incidentally lowered in summer acclimated animals. The pre-moulting stage in Collembola is characterised by a fasting period in which a new intestinal epithelium is formed. The epicuticle and exocuticle are secreted and just before moulting the old intestinal epithelium is rejected (Humbert, 1979). This process could involve complete evacuation of the gut contents and therefore might be expected to decrease the SCP, by reducing the number of potential ice nucleators in the body of the animal. Collembola moult throughout their lifecycle, with the number of instars ranging from four to more than 50 (Christiansen, 1990), with up to 40% of a population being in the moulting cycle at any time (Leinaas, 1983). However, Worland and Convey (2008) recently showed that the percentage of springtails involved in ecdysis increases with decreasing temperatures, such that at 0°C 80% of the population are potentially involved in ecdysis. Therefore these moulting and non-feeding periods represent a significant percentage of both the juvenile and adult life of a collembolan, with potentially strong modifying effects on their supercooling ability, at least in summer months.

Given this interpretation of the results, it is important to note that physiologically important effects may arise from shifts in gene expression that are not picked up at the significance level of 0.01 adopted in this study. To detect significantly smaller changes in gene expression, a larger study is necessary. It should also be noted that

changes in mRNA transcription levels do not necessarily equate to changes in protein levels. However, they provide an indicator of cellular processes affected by the conditions under study. It should also be noted that whilst meaningful functions and ecological relevance can be ascribed to the genes identified above, a high percentage of clones (approximately 50%) displayed no identifiable homology to known sequences. Some of these may well represent critical genes in the biochemical processes of High and Low Group animals, but more extensive studies, including cloning full-length sequences, will be required for putative domain identification and functional assignment. Of this set of genes, certainly the cluster with sequence similarity to a cupin domain is of future targeted interest.

Production of cDNA was problematic for this organism. Although the reason for this is unknown, the high level of cuticle pigmentation interfering with the enzymic construction process is suspected. Therefore it was decided to construct a small prototype microarray to determine if meaningful differences in gene expression could be identified between environmental samples of interest in this group. This preliminary study provides some very interesting results and justifies further input into this project in the future. The results of this microarray for *C. antarcticus,* provide evidence for a readily identifiable differential gene expression between summer acclimated High and Low Group animals. The results provide support for the theory that SCP variation within *C. antarcticus* could be a consequence of endogenous physiological processes occurring during moulting, at least in the summer (Worland and Convey, 2008). This analysis has provided the basis for further investigations, such as the production of comprehensive cDNA libraries and a high-

density microarray. This will enable us to conduct more detailed studies and hopefully provide further insights into the extreme environmental adaptation of this species.

Acknowledgements

This paper was produced within the BAS Q4 BIOREACH/BIOFLAME core programmes and also contributes to the SCAR EBA programme. JP was sponsored by the EU Sleeping Beauty Consortium: Specific Targeted Research Project, Contract no 012674 (NEST). JP and GG-L are also funded by the MSTD grant 143034, awarded by the Republic of Serbia. The authors would like to thank the NEBC Bio-linux team (http://nebc.nox.ac.uk/) for assistance with the bioinformatics and provision of software packages. Many thanks go to Pete Convey and Bill Block for critical reading of the manuscript.

References

Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F. et al. 2000. The Genome Sequence of *Drosophila melanogaster*. Science 287, 2185-2195.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucl Acids Res. 25, 3389-3402.

Bale, J.S., 1993. Classes of insect cold hardiness. Funct. Ecol. 7, 751–753.

Bale, J.S., 2002. Insects and low temperatures: from molecular biology to distributions and abundance. Philos. Trans. R. Soc. Lond. B Biol. Sci. 357, 849-862.

Bassuk, J.A., Iruela-Arispe, M.L., Lane, T.F., Benson, J.M., Berg, R.A., Sage, E.H., 1993. Molecular analysis of chicken embryo SPARC (osteonectin). Eur. J. Biochem. 218, 117–127.

Blaxter, M., 2002. Opinion piece. Genome sequencing: time to widen our horizons. Brief. Funct. Genomics Proteomics 1, 7-9.

Block, W., 1982. Cold hardiness in invertebrate poikilotherms. Comp. Biochem. Physiol. A Comp. Physiol. 73A, 581-593.

Block, W., 1984. A comparative study of invertebrate supercooling at Signy island, maritime Antarctic. Br. Antarct. Surv. Bull. 64, 67–76.

Block, W., 1990. Cold tolerance of insects and other arthro- pods. Philos. Trans. R. Soc. Lond. B Biol. Sci. 326, 613-633.

Block, W., 1991. To freeze or not to freeze? Invertebrate survival of. sub-zero temperatures. Funct. Ecol. 5, 284-290.

Boguski, M.S., Lowe, T.M.J., Tolstoshev, C.M. 1993. dbEST – database for 'expressed sequence tags'. Nat Genet. 4, 332-333.

Cannon, R.J.C., 1983. Experimental studies on supercooling in two Antarctic micro-arthropods. J. Insect Physiol. 29, 617-624.

Cannon, R.J.C., Block, W., 1988. Cold tolerance of microarthropods. Biol. Rev. (Camb.) 63, 23–77.

Cardoso,J.C.R., Pinto, V.V., Vieira, F.A., Clark, M.S., Power, D.M. 2006 Evolution of secretin family GPCR members in metazoa. BMC Evol Biol. 6, 108.

Chown, S.L., Klok, C.J., 1998. Interactions between desiccation resistance, host-plant contact and the thermal biology of a leaf-dwelling sub-antarctic caterpillar, *Embryonopsis halticella* (Lepidoptera: Yponomeutidae). J. Insect Physiol. 44, 615-628.

Christiansen, K.A., 1990. Insecta: Collembola. In: Dindal, D. (Ed.), Soil Biology Guide. John Wiley & Sons, New York, pp. 965-995.

Clark, M.S., Bendell, L., Power, D.M., Warner, S., Elgar, G., Ingleton P.M. 2002. Calcitonin: characterisation and expression in a teleost fish, *Fugu rubripes*. J Mol Endocrinol. 28, 111-123.

Clark, M.S., Edwards, Y.J.K., Peterson, D., Clifton, S.W., Thompson, A.J., Sasaki, M., Suzuki, Y., Kikuchi, K., Watabe, S., Kawakami, K., Sugano, S., Elgar, G., Johnson, S.L., 2003 *Fugu* ESTs: New resources for transcription analysis and genome annotation. Genome Res. 13**,** 2747-2753.

Clark, M.S; Fraser, K.P.P.F; Peck, L.S. (2008) Antarctic marine molluscs do have an HSP70 heat shock response. Cell Stress Chaperones. 13. 10.1007/s12192-008-0014-8.

Convey, P., 1996. The influence of environmental characteristics on the life history attributes of Antarctic terrestrial biota. Biol. Rev 71, 191–225.

Convey, P., 1997. How are the life history strategies of Antarctic terrestrial invertebrates influenced by extreme environmental conditions? J. Therm. Biol. 22, 429–440.

Convey, P., Smith, R.I.L., 1997. The terrestrial arthropod fauna and its habitats in northern Marguerite Bay and Alexander Island, maritime Antarctica. Antarct. Sci. 9, 12– 26.

Dunwell, J.M., Culham, A., Carter, C.E., Sosa-Aguirre, C.R., Goodenough, P.W. 2001. Evolution of functional diversity in the cupin superfamily. Trends Biochem Sci. 26, 740-746.

Ewing, B., Green, P. 1998a. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8, 175-185.

Ewing, B., Green, P. 1998b. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8,186-194.

Ewing, H.E., 1922. Note on the occurrence and distribution of Antarctic land Arthropods (Springtails Mites: Collembola and Acarina). Ent. News 33, 76-79.

Fuller, B.J., 1999. Organ preservation: The profit and loss account of using hypothermia to maintain viability. Transplant. Rev. 13, 55-66.

Gressitt, J.L., 1967. Introduction. Entomology of Antarctica. In: Gressitt, J.L. (Ed.), Antarctic Research Series. American Geophysical Union, Washington, D.C. , pp. 1–33.

Holmstrup, M., Bayley, M., Ramlov, H. 2002. Supercool or dehydrate? An experimental analysis of overwintering strategies in small permeable arctic invertebrates. Proc Natl Acad Sci USA. 99, 5716-5720.

Humbert, W., 1979. The midgut of Tomocerus minor Lubbock (Insecta, Collembola): ultrastructure, cytochemistry, ageing and renewal during a moulting cycle. Cell Tissue Res. 196, 39-57.

Janetschek, H., 1967. Arthropod ecology of South Victoria Land. In: Gressitt, J.L. (Ed.), Antarctic Research Series. American Geophysical Union, Washington, D.C. , pp. 205–293.

Knight, C.A., Duman, J.G., 1986. Inhibition of recrystallization of ice by insect thermal hysteresis proteins: a possible cryoprotective role. Cryobiology 23, 256–262.

Lee, R.E., Denlinger, D.L., 1991. Insects at LowTemperature. Chapman and Hall, New York.

Leinaas, H.P., 1983. Synchronized moulting controlled by communication in group-living Collembola. Science 219, 193–195.

Lyne, R., Burns G., Mata, J., Penkett, C.J., Rustici, G., Chen, D.R., Langford, C., Vetrie, D., Bahler, J. 2003. Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility and processing of array data. BMC Genomics. doi:10.1186/1471-2164-4-27.

Mans, B.J., Andersen, J.F., Francischetti, I.M.B., Valenzuela, J.G., Schwan, T.G., Pham, V.M., Garfield, M.K., Hammer, C.H., Ribeiro, J.M.C. 2008. Comparative sialomics between hard and soft ticks: Implications for the evolution of blood-feeding behaviour. Insect Biochem Mol Biol. 38, 42-58.

Parkinson, J., Anthony, A., Wasmuth, J., Schmid, R., Hedley, A., Blaxter, M., 2004. PartiGene--constructing partial genomes. Bioinformatics 20, 1398-1404.

Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., Lee, Y., White, J., Cheung, F., Parvizi, B., Tsai, J., Quackenbush, J. 2003. TIGR Gene

Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics. 19, 651-652.

Petalidis, L., Bhattacharyya, S., Morris, G.A., Collins, V.P., Freeman, T.C., Lyons, P.A., 2003. Global amplification of mRNA by template-switching PCR: linearity and application to microarray analysis. Nucleic Acids Res. 31, e142.

Pfaffl, M.W., 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucl. Acids. Res. 29, 2002-2007.

R Development Core Team. 2005. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. http://www.R-project.org.

Richie, M.E., Silver, J., Oshlack, A., Holmes, M., Diyagama, D., Holloway, A., Smyth, G.K. 2007. A comparison of background correction methods for two colour microarrays. Bioinformatics. 23, 2700-2707.

Rothery, P., Block, W., 1992. Characterizing supercooling point distributions. Cryo. Lett. 13, 193–198.

Salt, R.W., 1961. Principles of insect cold hardiness. Annu. Rev. Entomol. 6, 58-74.

Schenker, R., 1984. Effects of temperature acclimation on cold hardiness of Antarctic micro-arthropods. Revue d'Ecologie et de Biologie du Sol 21, 205–220.

Schwarzbauer, J.E., Spencer, C.S., 1993. The Caenorhabditis elegans homologue of the extracellular calcium binding protein SPARC/osteonectin affects nematode body morphology and mobility. Mol. Biol. Cell 4, 941–952.

Sinclair, B.J., Vernon, P., Jaco Klok, C., Chown, S.L., 2003a. Insects at low temperatures: an ecological perspective. Trends Ecol. Evol. 18, 257-262.

Sinclair, B.J., Jaco Klok, C., Scott, M.B., Terblanche, J.S., Chown, S.L., 2003b. Diurnal variation in supercooling points of three species of Collembola from Cape Hallett, Antarctica. J. Insect Physiol. 49, 1049-1061.

Smyth G.K. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 3, 3.

Smyth, G.K. 2005. Limma: linear models for microarray data. In Bioinformatics and computational biology solutions using R and Bioconductor, Gentleman, R., Carey, V., Dudoit, S., Irizarry, R., Huber, W. (eds). Springer, New York. P 397-420.

Smyth, G.K., Michaud, J., Scott, H. 2005. The use of within-array replicate spots for assessing differential expression in microarray experiments. Bioinformatics. 21, 2067-2075.

Smyth, G.K., Speed, T.P. 2003. Normalization of cDNA microarray data. Methods. 31, 265-273.

Sømme, L., 1985. Ecological studies on Cryptopygus sverdrupi from Muhlig-Hofmannfjella, Antarctica. Cryo. Lett. 4, 392-393.

Sømme, L., Conradi-Larsen, E.M., 1977. Cold-hardiness of collembolans and oribatid mites from windswept mountain ridges. Oikos 29, 118–126.

Sømme, L., 1981. Cold tolerance of alpine, Arctic, and Antarctic Collembola and mites. Cryobiology 18, 212-220.

Sømme, L., 1982. Supercooling and winter survival in terrestrial arthropods. Comp. Biochem. Physiol. 73A, 519-543.

Sømme, L., Block, W., 1982. Cold hardiness of Collembola at Signy Island, maritime Antarctic. Oikos 38, 168–176.

Sømme, L., 1995. Invertebrates in Hot and Cold Arid Environments. Springer-Verlag, Berlin.

The UniProt Consortium. 2007. The Universal Protein Resource (UniProt). Nucl Acids Res 2007, 35:D193-197.

The Gene Ontology Consortium. 2000. Gene Ontology: tool for the unification of biology. Nat Genet. 25, 25-29.

Thibaud, J.M., 1968. Cycle de tube digestif lors de l'intermue chez les Hypogastruridae (Collemboles) épigés et cavernicoles. Revue d'Ecologie et de Biologie du Sol 4, 647–655.

Tilbrook, P.J., 1970. The biology of *Cryptopygus antarcticus*. In: Holdgate, M.W. (Ed.), Antarctic Ecology. Academic Press, London, pp. 908-918.

Tilbrook, P.J., 1977. Energy flow through a population of the col- lembolan Cryptopygus antarcticus. In: Llano, G. (Ed.), Adaptations within Antarctic ecosystems. Smithsonian Institution, Washington DC., pp. 935-946.

Usher, M.B., Booth, R.G., 1986. Arthropod communities in a maritime Antarctic moss-turf habitat: multiple scales of pattern in the mites and Collembola. J. Anim. Ecol. 55, 155–170.

van der Woude, H.A., 1987. Seasonal changes in cold hardiness of temperate Collembola. Oikos 50, 231–238.

Worland, M.R., 1996. The relationship between water content and cold tolerance in the Arctic collembolan *Onychiurus arcticus* (Collembola: Onychiuridae). Eur. J. Entomol. 93, 341-348.

Worland, M.R., Convey, P., 2001. Rapid cold hardening in Antarctic microarthropods. Funct. Ecol. 15, 515-524.

Worland, M.R., 2005. Factors that influence freezing in the sub-Antarctic springtail *Tullbergia antarctica*. J. Insect Physiol. 51, 881-894.

Worland M. R., Convey P., 2008. The significance of the moult cycle to cold tolerance in the Antarctic collembolan *Cryptopygus antarcticus.* J. Insect Physiol. (in press)

Worland, M.R., Leinaas, H.P., Chown, S.L., 2006. Supercooling point frequency distributions in Collembola are affected by moulting. Funct. Ecol. 20, 323-329.

Young, S.R., Block, W., 1980. Experimental studies on the cold tolerance of *Alaskozetes antarcticus*. J. Insect Physiol. 26, 189-200.

Zachariassen, K.E., 1992. Ice nucleating agents in cold-hardy insects. In: Somero, G.N., Osmond, C.B., Bolis, C.L. (Eds.), Water and Life. Springer-Verlag, Berlin, pp. 261-281.

Table Legends

Table 1: EST data statistics for the original EST dataset and specifically for those clones used for the microarray. “Discovery” is defined as the number of singletons in each library defined by library size. “Diversity” is defined as the number of different “genes” (clusters) each library contributed, divided by the library size (Clark et al, 2003). Reads are those submissable after quality clipping and greater than 100bp. The threshold for Blastx significance is $1e^{-10}$.

Table 2: Characterisation of the 7 largest clusters from the EST dataset.

Table 3: GOSLIM annotation for the EST dataset, detailing Biological function and Biological process.

Table 4: Microarray analysis: Genes up regulated in the Low Group animals compared to control animals. Only clusters which matched database entries have been included. Where there was clearly a match to a putative protein or a domain, a specific Blast ID has not been assigned.

Table 5: Microarray analysis: Genes up regulated in the High Group animals compared to control animals.

Table 6: Microarray analysis: Genes up regulated in the Low Group animals compared to High Group animals.

Table 7: GO enrichment results (Biological function) for both the Low, High and Low compared to High Group animals. A = Proportion of clones for this GO annotation in the dataset. B = Proportion of clones for this GO annotation in the gene lists for up-regulated genes. A-B = proportional enrichment of clones associated with these GO terms in the expression data. NB. There were no significant results for the High Group animals compared to the Low Group animals.

	EST dataset	Chip
#Clones		672
# Reads	1180	531
# singletons	647	308
# clusters	130	62
# putative transcripts	777	370
Avg cluster size	4.1	3.6
Largest cluster	92	71
# clusters with 2 ESTs	85	40
# clusters with 3 ESTs	19	15
# clusters with 4-5 ESTs	12	5
# clusters with 6-10 ESTs	9	1
# clusters with >10 ESTs	5	1
Discovery	0.55	0.58
Diversity	0.66	0.7
# (%) with significant SwissProt hits	397 (34%)	181 (34%)
# (%) with significant trembl hits	462 (39%)	214 (40%)
# (%) with no hits	713 (60%)	315 (59%)

Table 1

Contig ID	Size	e-value	database	Database ID	Organism	Common name	Description
CL1Contig5	92	$3e^{-8}$	nr/nt	DQ021427	*Onychothemis testacea*	Dragonfly species	12S rRNA, tRNA-Val, 16S rRNA genes; mitochondrial
CL1Contig3	56	$1e^{-36}$	nr/nt	AF272824.1	*Tetrodontophora bielanensis*	Giant springtail	mitochondrion, complete genome
CL2Contig1	13	$4e^{-58}$	TrEMBL	Q6BDT1	*Orthetrum triangulare melania*	Dragonfly species	ATP synthase A chain
CL1Contig4	12	$8e^{-72}$	TrEMBL	Q9B2J1	*Chrysomya putoria*	African latrine blowfly	Cytochrome c oxidase polypeptide III
CL4Contig1	11	$1e^{-137}$	TrEMBL	Q6VQ13	*Apis mellifera*	Honeybee	ADP/ATP translocase
CL1Contig1	10	$3E^{e20}$	SWISSPROT	P19351	*Drosophila melanogaster*	Fruit fly	Troponin T, skeletal muscle
CL3Contig1	10	$1e^{-119}$	SWISSPROT	P00400	*Drosophila yakuba*	Fruit fly	Cytochrome c oxidase subunit I

Table 2

GO ID	Category	Description	
Biological Function			
GO:0003774	F	motor activity	11
GO:0016874	F	ligase activity	22
GO:0005515	F	protein binding	307
GO:0016787	F	hydrolase activity	147
GO:0045182	F	translation regulator activity	19
GO:0004871	F	signal transducer activity	10
GO:0015267	F	channel or pore class transporter activity	8
GO:0003824	F	catalytic activity	63
GO:0005488	F	binding	299
GO:0016740	F	transferase activity	85
GO:0005215	F	transporter activity	39
GO:0008565	F	protein transporter activity	6
GO:0015075	F	ion transporter activity	79
GO:0030528	F	transcription regulator activity	37
GO:0004386	F	helicase activity	11
GO:0016209	F	antioxidant activity	2
GO:0016829	F	lyase activity	15
GO:0016491	F	oxidoreductase activity	98
GO:0004872	F	receptor activity	27
GO:0016853	F	isomerase activity	15
GO:0030234	F	enzyme regulator activity	16
GO:0005198	F	structural molecule activity	103
Biological Process			
GO:0009058	P	biosynthetic process	4
GO:0050896	P	response to stimulus	149
GO:0030154	P	cell differentiation	97
GO:0008152	P	metabolic process	138
GO:0007275	P	multicellular organismal development	126
GO:0009987	P	cellular process	278
GO:0009405	P	pathogenesis	1
GO:0046903	P	secretion	34
GO:0006810	P	transport	151
GO:0050789	P	regulation of biological process	140
GO:0009056	P	catabolic process	9

Table 3

Signature Clone	GenBank ID	P Value	Blast ID	Description	Organism	Blast Score/ P Value
CR_C02P0_01E12	FF278373	6.30E-005	P82166	Cuticle protein 19.8	*Locusta migratoria* Migratory locust	83/4e-16
CR_C03P0_03F06	FF278760	4.88E-004	N/A	Uncharacterised protein containing a Cupin domain	Various: N/A	150/1.9e-14
CR_C03P0_03D08	FF278742	1.66E-003	Q6BDT1	ATP synthase A chain	*Orthetrum triangulare melania* Dragonfly species	227/4e-58
CR_C03P0_03B02	FF278722	2.97E-003	Q0PXZ8	Putative 60s acidic ribosomal protein	*Diaphorina citri* Asian Citrus psyllid	110/5e-23
CR_C03P0_03H07	FF278779	3.36E-003	Q7M497	Exoskeletal protein HACP188	*Homarus americanus* American lobster	106/1e-21
CR_C01P0_04A07	FF278281	3.72E-003	Q28EK6	CHK1 checkpoint homologue	*Xenopus tropicalis* Pipid frog	203/4e-14
CR_C03P0_03C06	FF278735	4.38E-003	Q1HRJ5	Mitochondrial NADH dehydrogenase (Ubiquinone)	*Aedes aegypti* Yellow fever mosquito	211/2.8e-15
CR_C01P0_04A03	FF278278	5.70E-003	Q09JM0	10kDa putative secretory protein	*Argas monolakensis* Mono Lake bird tick	198/9.8e-24
CR_C02P0_02H05	FF278460	5.70E-003	Q16R87	Pupal cuticle protein	*Aedes aegypti* Yellow fever mosquito	106/9e-22
CR_C03P0_03F02	FF278758	8.82E-003	Q2Q1I4	Cytochrome b	*Sclerophasma paresisense* Gladiator species	226/4e-58
CR_C01P0_02A10	FF278159	9.87E-003	N/A	Putative senescence associated protein	Various: *N/A*	95/1e-18

Table 4

Signature Clone	GenBank ID	P Value	Blast ID	Description	Organism	Blast Score/ P Value
CR_C01P0_04A07	FF278281	9.24E-006	Q28EK6	CHK1 checkpoint homologue	*Xenopus tropicalis* Pipid frog	80/4e-14
CR_C03P0_03C06	FF278735	1.14E-004	Q1HRJ5	Mitochondrial NADH dehydrogenase (Ubiquinone)	*Aedes aegypti* Yellow fever mosquito	84/2e-15
CR_C01P0_04A03	FF278278	7.68E-004	Q09JM0	10kDa putative secretory protein	*Argas monolakensis* Mono Lake bird tick	124/4e-27
CR_C03P0_03A09	FF278719	5.65E-003	Q8IS91 Serine protease		*Glossina fuscipes fuscipes* Riverine tsetse fly	72/1e-12
CR_C01P0_02A10	FF278159	7.28E-003	N/A	Putative senescence associated protein	Various: *N/A*	95/1e-18
CR_C04P0_01D06	FF278851	7.97E-003	Q85QR3	Cytochrome c oxidase subunit I	*Gomphiocephalus hodgsoni* Antarctic springtail species	449/1e-125

Table 5

Signature Clone	GenBank ID	P Value	Blast ID	Description	Organism	Blast Score/ P Value
CR_C03P0_03D08	FF278742	5.25E-007	Q6BDT1	ATP synthase A chain	*Orthetrum triangulare melania* Dragonfly species	227/4e-58
CR_C02P0_01E12	FF278373	6.38E-006	P82166	Cuticle protein 19.8	*Locusta migratoria* Migratory locust	83/4e-15
CR_C02P0_02H01	FF278456	1.88E-005	Q16R87	Pupal cuticle protein	*Aedes aegypti* Yellow fever mosquito	106/9e-22
CR_C03P0_03B06	FF278726	3.96E-004	A7UKR5	Sorbitol dehydrogenase	*Pyrrhocoris apterus* Sap sucking bug	644/3.7e-61
CR_C03P0_03F05	FF278759	4.58E-004	Q0PWT7	Putative elongation factor 1-alpha	*Diaphorina citri* Asian Citrus psyllid	181/8e-45
CR_C03P0_03E06	FF278752	5.13E-004	Q95V16	Cuticular protein	*Myzus persicae* Peach-potato aphid	69/7e-11
CR_C01P0_02C04	FF278169	8.68E-004	QPPH7	Putative tropomysin	*Homalodisca coagulata* Glassy winged sharp-shooter	95/1e-18
CR_C01P0_04G05	FF278325	4.05E-003	Q1HR73	GTP-binding ADP-ribosylation factor Arf1	*Aedes aegypti* Yellow fever mosquito	87/3e-16
CR_C01P0_02G05	FF278201	6.46E-003	Q16TR9	Sodium/potassium-dependent ATPase beta-2 sununit	*Aedes aegypti* Yellow fever mosquito	87/3e-16
CR_C03P0_02F06	FF278693	8.73E-003	Q9B2J1	Cytochrome c oxidase polypeptide III	*Chrysomya putoria* African latrine blow fly	306/6e-82
CR_C04P0_01B04	FF278841	9.53E-003	Q175J8	Inhibitor of apoptosis 1, diap1	*Aedes aegypti* Yellow fever mosquito	74/2e-12
CR_C02P0_02E02	FF278429	9.53E-003	A4D010	Beta actin	*Loligo pealeii* Longfin squid	243/2e-63
CR_C03P0_02D10	FF278679	9.53E-003	O96696	Cation-transporting ATPase	*Heliothis virescens* Noctuid moth	228/7e-59

Table 6

GO ID	p value	description	A	B
Low Group				
GO:0008011	0.00982	structural constituent of pupal cuticle (sensu Insecta)	1.94	22.22
GO:0042302	0.00825	structural constituent of cuticle	5.33	33.33
High Group				
GO:0004674	0.00257	protein serine/threonine kinase activity	5.82	42.85
GO:0004672	0.00045	protein kinase activity	4.36	42.85
Low verses High				
GO:0042302	0.00063	Structural constituent of cuticle	5.33	28.27
GO:0016820	0.00809	Hydrolase activity	3.39	19.04

Table 7