NGS applications to understand invertebrate biodiversity of Antarctica and mechanisms of gene expression involved in climatic changes

TUTOR: Prof. Antonio Carapelli   DOTTORANDO: Dr. Claudio Cucini

Table of contents

Abstract

Human activities, such as greenhouse emissions and pollution, are leading to 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 temperature shifts and ice-melting registered within the continent in recent decades. To date, many scientists focused on how marine species react to these changes but no molecular data are currently available for continental terrestrial invertebrates and in particular for Collembola (=springtails). Therefore, part of my PhD project was to study the transcriptomic response of the endemic Antarctic springtail *Cryptopygus terranovus* following a mid-term exposure of 20 days at 18°C. Expression data were compared with wild specimens sampled in native environment. Although individual plasticity in transcript modulation was recorded, several pathways appear to be differentially modulated: protein catabolism, fatty acid metabolism and a sexual response characterized by spermatid development were induced, while lipid catabolism was downregulated in treated samples. Moreover, the temperature experienced by these micro-invertebrates is a pivotal parameter to understand these animals' ecology and physiology. However, at present, detailed knowledge of microhabitat physical conditions in Antarctica is limited and biased towards sub-Antarctic and maritime Antarctic regions. To better understand these temperature conditions, it was analysed a year-round temperature data in ponds and soils in an area of the Victoria Land coast, comparing these measurements with air temperatures from the closest automatic weather station. Important difference in temperature dynamics between the air, soil and pond datasets was registered. Ponds were the warmest sites overall, mostly differing with the air temperatures due to their greater thermal capacity, which also influenced their patterns of freeze-thaw cycles and mean daily thermal excursion.

Furthermore, to better understand the biodiversity of Collembola two new mitochondrial genomes of Antarctic springtails were sequenced and analysed. They were employed to revise the entire systematic of the class, its nucleotide composition and genome arrangement by comparing them to all the available sequences deposited in Genbank. In the phylogenetic analysis, with minor exceptions, it was confirmed the monophyly of Poduromorpha and Symphypleona *sensu stricto* (the latter recovered as the most basal group in the springtail phylogenetic tree), whereas monophyly of Neelipleona and Entomobryomorpha was only supported when some critical taxa in these two lineages were excluded. The genome arrangement review allowed to identify four new gene orders (one from the newly sequenced *Tullbergia mixta*), for a total of 16 models. Finally, nucleotide

composition analyses confirmed the low AT bias in Collembola mitochondrial DNA respect to other Hexapoda, and that third codon position is inclined to mutation accumulation, especially in 4-fold amino-amino acids. To ease the process of mitochondrial genome analyses, it was created a web resource named EZmito, a free web server useful to analyse mitochondrial genomes. It is composed of five main tools: (i) EZsplit, useful to download and format sequences directly from the NCBI database; (ii) EZpipe, a pipeline designed to format mitochondrial sequences before the phylogenetic analysis; (iii) EZskew, which helps users to calculate nucleotide biases; (iv) EZcodon, a fast tool which calculates the Relative Synonym Codon Usage of different mtDNA species and (v) EZmix, which recognizes areas of inter molecular similarity indicative of the assembly of chimeric mitochondrial genomes. Interestingly, to date, the most used tool within the hub is EZcodon, followed by EZsplit and EZpipe.

Riassunto

Le attività umane, come ad esempio le emissioni di gas serra e l'inquinamento atmosferico, sono le principali cause relative al riscaldamento globale, ai cambiamenti climatici e alla riduzione della biodiversità. Ambienti incontaminati come l'Antartide non sono immuni da questi fenomeni e gli sbalzi di temperatura stanno caratterizzando il continente, provocandone lo scioglimento dei ghiacciai. Oggigiorno, molti scienziati si sono concentrati su come specie marine antartiche reagiscono a questi cambiamenti ma, attualmente, non sono disponibili dati molecolari che indicano le reazioni di invertebrati terrestri ed in particolare di collemboli. Pertanto, parte del mio progetto di dottorato è stato quello di studiare la risposta trascrittomica del collembolo antartico, endemico della Terra di Vittoria, *Cryptopygus terranovus* a seguito di un'esposizione a medio termine di 20 giorni a 18° C. I dati di espressione sono stati confrontati con esemplari campionati in ambiente esterno e non esposti a nessun tipo di stress abiotico. I risultati ottenuti indicano una elevata plasticità individuale nella modulazione dei trascritti ma alcune pathway caratterizzano differentemente i due gruppi di individui (trattati e controllo). Un aumento del catabolismo proteico, del metabolismo degli acidi grassi e di una risposta sessuale caratterizzata dallo sviluppo degli spermatidi connesso ad una diminuzione del catabolismo lipidico è stato ritrovato nel gruppo trattato ad alte temperature. Inoltre, la temperatura sperimentata da questi microinvertebrati è un parametro fondamentale per comprendere l'ecologia e la fisiologia di questi animali. Tuttavia, ad oggi, la conoscenza dettagliata delle condizioni

abiotiche dei microhabitat antartici è piuttosto limitata e sbilanciata verso le regioni subantartiche e marittime. Per comprendere meglio queste condizioni, sono stati raccolti ed analizzati dati di temperatura registrati per un anno intero in stagni e suoli in un'area vicino alla costa della Terra di Vittoria, confrontando queste misurazioni con le temperature dell'aria rilevate dalla stazione meteorologica automatica più vicina. I risultati di questo studio hanno evidenziato un'importante differenza nei patterns di dati tra aria, suolo e stagno. In generale, a causa della loro maggiore capacità termica, questi ultimi sono stati i siti più caldi, differendo principalmente con le temperature dell'aria, e registrando modelli di cicli di gelo-disgelo più ampi.

Inoltre, per comprendere meglio la biodiversità dei collemboli, sono stati sequenziati e analizzati due nuovi genomi mitocondriali di specie antartiche. Nel dettaglio, questi sono stati impiegati per rivedere l'intera sistematica della classe, la composizione nucleotidica e il gene order dei genomi, attraverso un confronto con tutte le sequenze disponibili depositate in Genbank. Durante l'analisi filogenetica, con piccole eccezioni, sono state confermate la monofilia di Poduromorpha e Symphypleona *sensu stricto* (quest'ultimo identificato come il gruppo più basale), mentre la monofilia di Neelipleona ed Entomobryomorpha è stata supportata solo quando alcuni taxa critici in questi due lignaggi sono stati esclusi. La revisione della disposizione del genoma ha permesso di identificare quattro nuovi gene orders (uno appartenente a *T. mixta*, appena sequenziato), per un totale di 16 modelli diversi. Infine, le analisi della composizione nucleotidica hanno confermato la bassa percentuale di AT nei genomi mitocondriali dei collemboli rispetto ad altri esapodi, e che la terza posizione nelle triplette è maggiormente incline all'accumulo di mutazioni, specialmente negli amminoacidi 4-fold. Per facilitare il processo di analisi dei genomi mitocondriali, è stato creato un web server gratuito: EZmito. EZmito è composto da cinque strumenti principali: (i) EZsplit, utile per scaricare e formattare sequenze direttamente dal database Genbank (NCBI); (ii) EZpipe, una pipeline progettata per formattare sequenze mitocondriali prima dell'analisi filogenetica; (iii) EZskew, che aiuta gli utenti a calcolare i biases di composizione nucleotidica; (iv) EZcodon, uno strumento veloce che calcola l'utilizzo del relativo dei codoni sinonimi per diverse specie e (v) EZmix, che riconosce aree di somiglianza intermolecolare indicative dell'assemblaggio di genomi mitocondriali chimerici. È interessante notare che, ad oggi, lo strumento più utilizzato all'interno dell'hub è EZcodon, seguito da EZsplit ed EZpipe.

1. Introduction

1.1. Antarctica's geography and its geological history, an overview

Located within the Antarctic Circle, Antarctica is the southernmost, most remote, and most inaccessible continent on Earth. Antarctica is isolated from other landmasses by 1,000 km of the Drake Passage (the body of water separating South America from the Antarctic Peninsula) and by 4,000-5,000 km from Australia and South Africa. Due to its geographical location, Antarctica is also the coldest continent on Earth. In fact, more than 99.6% of its area is permanently covered by snow or ice, with few ecological exceptions such as cliffs, nunataks, and seasonally snow and ice-free areas. Most of the terrestrial biodiversity is concentrated in these areas (Convey, 2013, 2017).

During its geological history, Antarctica experienced largely different climatological conditions, with associated different faunas. It was originally part of the Gondwana supercontinent, formed after the fragmentation of Rodina (ca from late Neoproterozoic to early Cambrian). Indeed, during this geological period, Gondwana was located at equatorial latitudes. From its original location, Antarctica began rotating and moving southwards. During this period, it underwent extensive glaciation cycles but never experienced permanent ice covers (McLoughlin, 2001). Due to its first oceanic rift, the supercontinent started to break up around 165 Mya (middle Jurassic) leading to the formation of East (comprising current India, Australia and East Antarctica) and West Gondwana (represented today by South America, Arabia, Africa and West Antarctica landmasses). Around 137-127 Mya (early Cretaceous), East Gondwana separated from the Indian subcontinent, whereas the rest of the landmass continued in its movement southwards. Simultaneously, West Antarctica (namely the Antarctic Peninsula, Ellsworth–Whitmore Mountains block, Thurston Island and Marie Byrd Land) connected to East Antarctica without losing its continuity with South America, thus allowing the movement of terrestrial species towards the southern continent (McLoughlin, 2001). Indeed, at this time, the continent was characterized by a temperate rainforest climate and CO_2 levels reached 3000 ppm (compare to a current figure of 421.13 ppm; https://www.co2.earth/daily-co2). As such, due to these palaeoclimatological features, authors refer to this period as the "Greenhouse World", where biodiversity and ecological interactions flourished, culminating in late Cretaceous before K-Pg extinction (Turner et al., 2014; Condamine et al., 2021). Only 30-35 Mya (ca early Oligocene) Antarctica lost all its terrestrial connections and acquired its status as an isolated landmass. The separation from South America led to the development of circumpolar ocean currents, that in turn accelerated its gradual cooling (Convey, 2017). Due to the drop of CO_2 levels

and the formation of permanent ice sheets, scientists refer to this period as the "Icehouse World", that led Tundra-like ecosystems to gradually disappear until 12-13 Mya (Miocene) (Turner et al., 2014; Convey, 2017). Nevertheless, Pleistocene paleoclimatic data show that the Antarctic climate was linked to the climate observed in the Northern Hemisphere in terms of cycles of warming and cooling associated to modifications of the Earth's tilt around the Sun's orbit. However, even during inter-glacial periods, atmospheric CO_2 concentration was stably below 300 ppm and the maximum temperature peak recorded was 15°C (Turner et al., 2014).

1.2. Antarctica's climate

Antarctica is commonly divided into three main terrestrial biogeographic regions: (i) the sub-Antarctic, (ii) the maritime Antarctic and (iii) Continental Antarctica (Bergstrom et al., 2006; Convey, 2017). More recently, as many as 16 distinct "Antarctic Biogeographic Conservation Regions" (ABCRs) were identified by Terauds and Lee (2016) based on ecological and biogeographic data (Fig. 1). The sub-Antarctic includes all the islands and archipelagos that surround Antarctica. Atmospheric temperatures in this region are on average above the 0°C and precipitations, which range between 2,000 and 3,000 mm per year, are the most abundant within the continent (Bergstrom et al., 2006). The maritime Antarctic consists of the South Shetlands, South Orkney and South Sandwich archipelagos, Bouvetøya and Peter I. islands, as well as the western part of the Antarctic Peninsula. This area is characterized by average summer temperatures above 0°C, although temperature excursion, between the warm and cold season, is substantial. Precipitations account for 400-500 mm per year and are overall limited compared to the sub-Antarctic (Bergstrom et al., 2006). Lastly, the continental Antarctic includes the rest of Antarctica's landmasses and the east coast of the Antarctic Peninsula. This is the area of Antarctica characterized by the harshest climatic conditions, where cold deserts (such as the Dry Valleys) are present. Atmospheric temperatures are generally below 0°C and precipitations are scarce, in the order of 30-40 mm per year (Bergstrom et al., 2006).
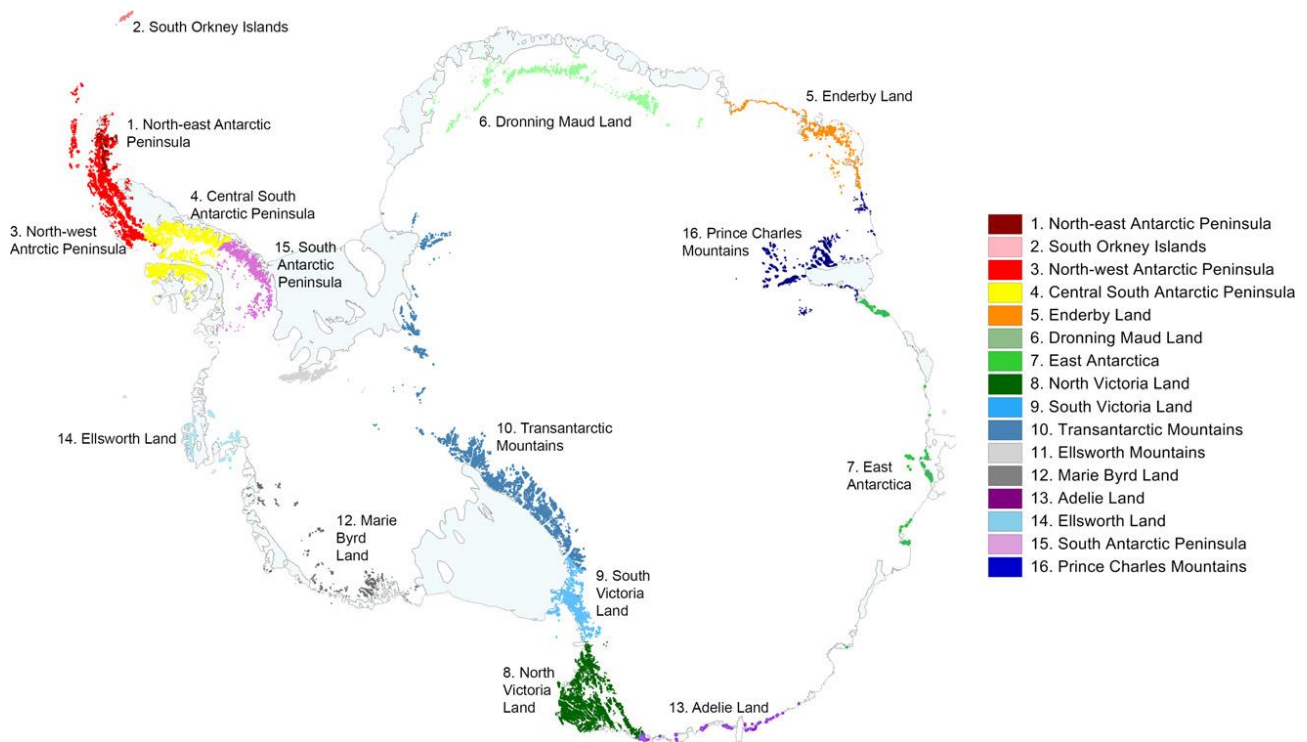
Figure 1. The 16 ABCRs proposed by Terauds and Lee (2016)

In continental and maritime Antarctic, viable habitats for terrestrial biodiversity are mostly restricted to few permanent or seasonally ice-free areas (about the 0.3% of the entire landmass) mainly concentrated in coastal areas (Convey, 2011; Convey et al., 2014; Wauchope et al., 2019). As such, terrestrial Antarctica can be compared to an island-like ecosystem, in which viable habitats can be separated by few meters to many hundreds of kilometres of ice (Convey et al., 2014). In this respect, terrestrial Antarctica, with its isolated and highly fragmented viable areas, is likely to be heavily affected by abrupt environmental changes due to possible abiotic shifts and climate changes (Convey & Peck, 2019).

While climate warming and cooling cycles have always characterized Earth's history, with the advent of the Anthropocene human activities and their impact on the environment determined marked modifications and lead to a reshaping of natural habitats as well as a substantial reduction in biodiversity. In particular, beginning with the industrial revolution, Earth's climate has undergone direct and significant modifications. Human activities determined a substantial increase in environment pollution, leading to a significant release of greenhouse gases and CO_2 rise. The last report of the Intergovernmental Panel on Climate Change (IPCC, 2022) presented dramatic consequences for Earth's future as an outcome of anthropogenic activities, anticipating that the World temperatures will see an

increase of 1.5°C above pre-industrial levels within the next two decades, which could in turn determine a rise in global temperatures by more than 4°C by the end of the century (Hughes et al., 2021).

In this scenario, some of the most dramatic alterations registered are (i) the increase of droughts, floods and heatwaves and (ii) the water cycle intensification. Moreover, global warming has promoted melting of ice covers, in turn determining a rise of sea levels, the release of greenhouse gases by human activities has caused ocean acidifications, and the ozone layer depletion has altered precipitations, especially in the Southern Hemisphere (Lee et al., 2017; Convey & Peck, 2019).

In this context, Antarctica may play a crucial role in the study and the understanding of climate change, as well as the promotion of global awareness to face such changes (Convey & Peck, 2019). As an example, Antarctica was the key to discover the ozone layer depletion, caused by chlorofluorocarbons (CFCs) release into the atmosphere, in the 1970s. Thanks to the success of Montreal Protocol in 1987, an international agreement to reduce chemical emissions, the "ozone hole" is now reducing, despite a recent evidence that some surviving CFCs are slowing down the process (Solomon, 2019). Nonetheless, since human activities trigger ripple effects on climate variations, the ozone depletion might be strictly related to the alteration of the whole climatological mechanism which regulates the south circumpolar zone, namely the Southern Annular Mode (SAM) (Thompson et al., 2011; Polvani et al., 2021). SAM oscillation varies depending on the location and intensity of air pressure, leading to positive (winds intensify towards Antarctica) or negative (winds increase at mid-latitudes) phases that impact over rainfall rates in Antarctica as well as in other western lands (e.g. Australia).

Interestingly, a slowing down in warming trend was registered since the late 2000s across parts of Antarctica, although temperatures are still higher than in the 1950s (Turner et al., 2016) and models predict a resumption of this tendency by the end of this century (Convey & Peck, 2019; Hughes et al., 2021). Since the middle of the last century, when the first Automatic Wheatear Stations (AWSs) were installed (Fig. 2), authors believed that Antarctica's warming was only to be registered in certain areas, namely around the Antarctic Peninsula (Lee et al., 2017). Nevertheless, it was recently demonstrated that a wide and significant warming took place at the South Pole (Clem et al., 2020; Turner et al., 2020). In a broader perspective, multi-years data indicated that the most significant warming took place at Vernadsky station (Antarctic Peninsula), where annual mean temperatures

increased by 3°C in less than 70 years. While these climatological modifications can have an impact on species viability and population (Kaspari & Valone, 2002; Bokhorst et al., 2008), scientists have also pointed out that standard surface mean air temperatures do not necessarily translate to biologically relevant consequences of temperature change directly (Hrbáček et al., 2020; Robinson et al., 2020; Cucini et al., 2022). Indeed, despite temperature recordings carried out with AWS or satellite infrastructures are relevant for large-scale climatic features, these are not capable of capturing fine-scale microclimatic dynamics (Lembrechts, 2022). It was already demonstrated that Antarctic microhabitats such as soil, freshwater and intertidal spaces (where flora and fauna truly extend) experience different abiotic conditions to those recorded by standard meteorological stations in the open-air (Davey et al., 1992; Quayle et al., 2002; Peck et al., 2005; Convey et al., 2018; Clarke & Beaumont, 2020). This primarily depends on the abiotic factors that affect and characterize the two distinct geographical scales such as radiation, water availability, soil composition, ground topography, etc (Woods et al., 2015; Hrbáček et al., 2020). Moreover, in sub-nivean areas snow acts as a thermal insulator preventing drastic temperature drop and allowing invertebrates, plants and microorganisms to be metabolically active even if air temperature is substantially lower (Convey et al., 2015, 2018).

In this respect, Next Generation Sequencing strategies, especially RNA-seq, can help scientists to investigate and assess the physiological effects and response to mutated conditions by simulating thermal stress in controlled environments and directly assessing which biological pathways are modulated. Molecular approaches have been employed to understand physiological responses in many Antarctic marine groups such as molluscs, echinoderms, crustaceans and bony fishes (Reed & Thatje, 2015; Bilyk et al., 2018; González-Aravena et al., 2018) but very little effort has been done to understand how the Antarctic terrestrial fauna reacts to increasing temperatures. While some molecular studies have highlighted the physiological response of the Antarctic midge *Belgica antarctica* to dehydration (Goto et al., 2011; Teets & Denlinger, 2014), overall knowledge on how Continental Antarctica hexapods may survive and react to rising temperatures is limited to work by Everatt and colleagues (2013) that studied the resistance to abiotic factors in *Alaskozetes antarcticus* and *Cryptopygus antarcticus*.
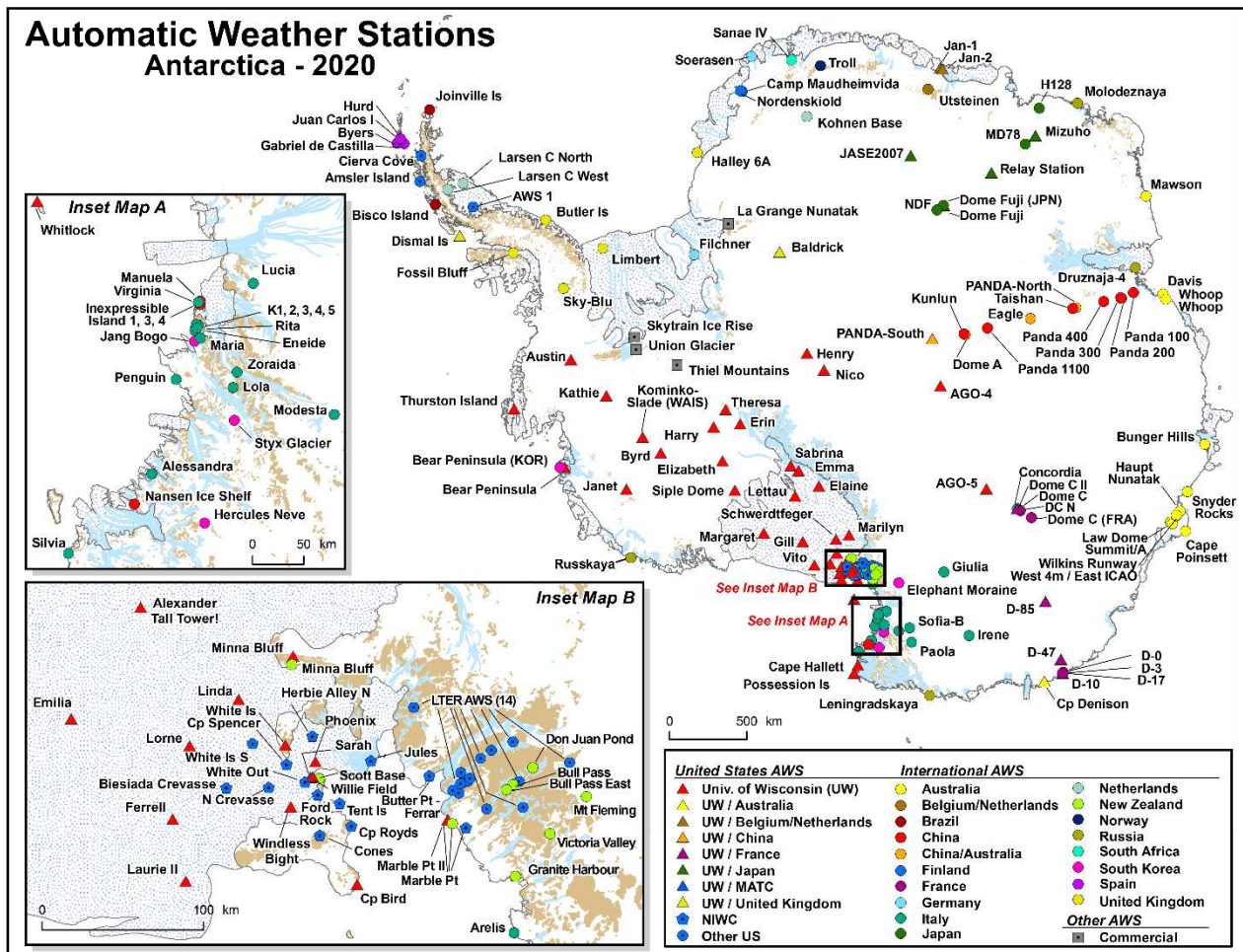
Figure 2. Automatic Weather Stations (AWSs) distribution in Antarctica

1.3. Antarctic terrestrial biodiversity

Terrestrial ecosystem development is restricted to permanently or seasonally snow- and ice-free regions, where biologically compatible environmental conditions are present (Convey et al., 2014). Key actors of biodiversity in this extreme polar ecosystem are autotrophic microorganisms, such as cyanobacteria and algae, which play a pivotal role in primary colonization. Bacteria communities, fungi and protozoans are indeed fundamental to stabilize minerals in the soil allowing for the establishment of micro-invertebrates' and plants' communities (Convey, 2017).

Despite some species of vertebrates (seabirds and mammals) breed on the sub- and maritime Antarctic islands (thus contributing to their nutrient input) only a few of them are endemic to the sub-Antarctic. Thus, most of the species dominating the terrestrial ecosystem across the entire continent are invertebrates (Convey, 2017).

The sub-Antarctic is the richest biogeographical region in terms of its biological components thanks to its milder climatic and environmental conditions. Here flora and fauna are similar to those observed in South America and New Zealand (Convey, 2017). Terrestrial biodiversity includes insects (Diptera and Coleoptera), micro-arthropods (Acari and Collembola) and micro-invertebrates (Nematoda, Tardigrada, Rotifera). Its flora is similar to the Arctic tundra, even though the sub-Antarctic lacks a permafrost layer (Convey, 2017).

On the other hand, maritime and Continental Antarctica are characterized by a reduced terrestrial biodiversity due to the harsher environmental conditions that limit flora and fauna expansion and diversification. Micro-arthropods and micro-invertebrates are the dominant components of biodiversity in the maritime Antarctica, with higher insects being represented by only two endemic chironomid species (Diptera, Hexapoda). This regions' flora is also restricted to turf-forming mosses, with only two angiosperm species being endemic. (Bergstrom et al., 2006; Convey, 2017). Lastly, the continental Antarctic is characterized by an even more reduced ecosystem, since its fauna is limited to few micro-invertebrates and micro-arthropods species, as well as some lichens and mosses which dominate the flora biodiversity (Bergstrom et al., 2006; Convey, 2017). In detail continental Antarctic animals are represented by Acari, Collembola, non-marine Crustacean, Nematoda, Rotifera, Tardigrada and parasitic insects (Mallophaga) (Convey, 2017).

Due to its geographic location, isolation and geological history, Antarctic biodiversity is the legacy of Gondwanan communities that survived glaciations in refugia and expanded during interglacial periods, managing to survive, adapt and establish (e.g., Bergstrom et al., 2006; Convey, 2017). Nonetheless, as a consequence of global warming and the increase of human traffics, non-native species have established in Antarctica, threatening the endemic biodiversity. To date, only a few species colonized the Antarctic Peninsula and the continent, but several introductions have been registered in the sub- and peri-Antarctic zones and are currently monitored by scientists (Convey & Peck 2019).

In order to preserve Antarctic biodiversity, it is necessary to study, investigate and understand species phylogeny and their distribution patterns. For this reason, since springtails are one of the most abundant groups in the Southern landmass, efforts to understand their taxonomy and systematic are necessary.

1.4. Collembola phylogeny and systematic

Collembola (springtails) are the most diverse group amongst basal Hexapoda, accounting for over 7,000 species worldwide and inhabiting almost every terrestrial habitat on Earth: from tropical deserts to ice-free areas of the polar regions. Soil, litter, and ground vegetation are their natural habitats (Hopkin, 1997), and here they play a key role in the decomposition of organic matter. Their first appearance was recently dated to the Silurian (443–419 million years ago), suggesting that their early diversification into major lineages may have taken place in parallel with the evolution of vascular plants and the processes of primary soil formation (Leo et al., 2019).

While springtail monophyly has never been questioned, the phylogenetic relationships of the group in the context of the Arthropoda have been a long-debated topic, with several hypotheses proposed during the last 20 years (Nardi et al., 2003; Luan et al., 2005; Carapelli et al., 2007). Traditionally Collembola have always been classified as hexapods because of many characters common to 'true insects' such as the body segment organization, the presence of six leg and the terrestrialization (Henning, 1981; Kristensen, 1998). However, in early 2000s, using mitochondrial DNA (mtDNA) as phylogenetic marker Nardi et al. (2003) proposed Hexapoda as diphyletic, with Collembola being the sister taxon of some crustacean *plus* insects. Further works which employed the same methodological strategy, but amplifying the original dataset, corroborated this hypothesis (e.g. Carapelli et al., 2007). In contrast, other authors, employing a different set of nuclear genes (Timmermans et al., 2008) and entire transcriptomes (Misof et al., 2014) or by modifying the original methodological strategy (Delsuc et al., 2003), challenged Nardi et al. (2003) hypothesis, proposing Hexapoda as monophyletic, founding supports by morphological and ecological characters.

At the beginning of Collembola systematics, distinct lineages were classified based on gross body morphology and Collembola were subdivided into two suborders: Symphypleona *sensu lato* (with a globular body shape and fused body segments) and Arthropleona (with an elongated body and well-defined segments) (Hopkin, 1997). With the introduction of molecular data, progresses were made towards resolving the group's internal relationships, leading to the acceptance of four main orders: the Arthropleona were subdivided into Entomobryomorpha and Poduromorpha, while Symphypleona *s.l.* was subdivided into Symphypleona *sensu stricto* and Neelipleona (D'Haese, 2002). At this point, inter-class relationships became the new object of debate amongst collembolan taxonomists, with different competing hypotheses being proposed based on nuclear and/or mitochondrial DNA

markers (Gao et al., 2008; Yu et al., 2016; Leo et al., 2019; Sun et al., 2020; Schneider et al., 2011).
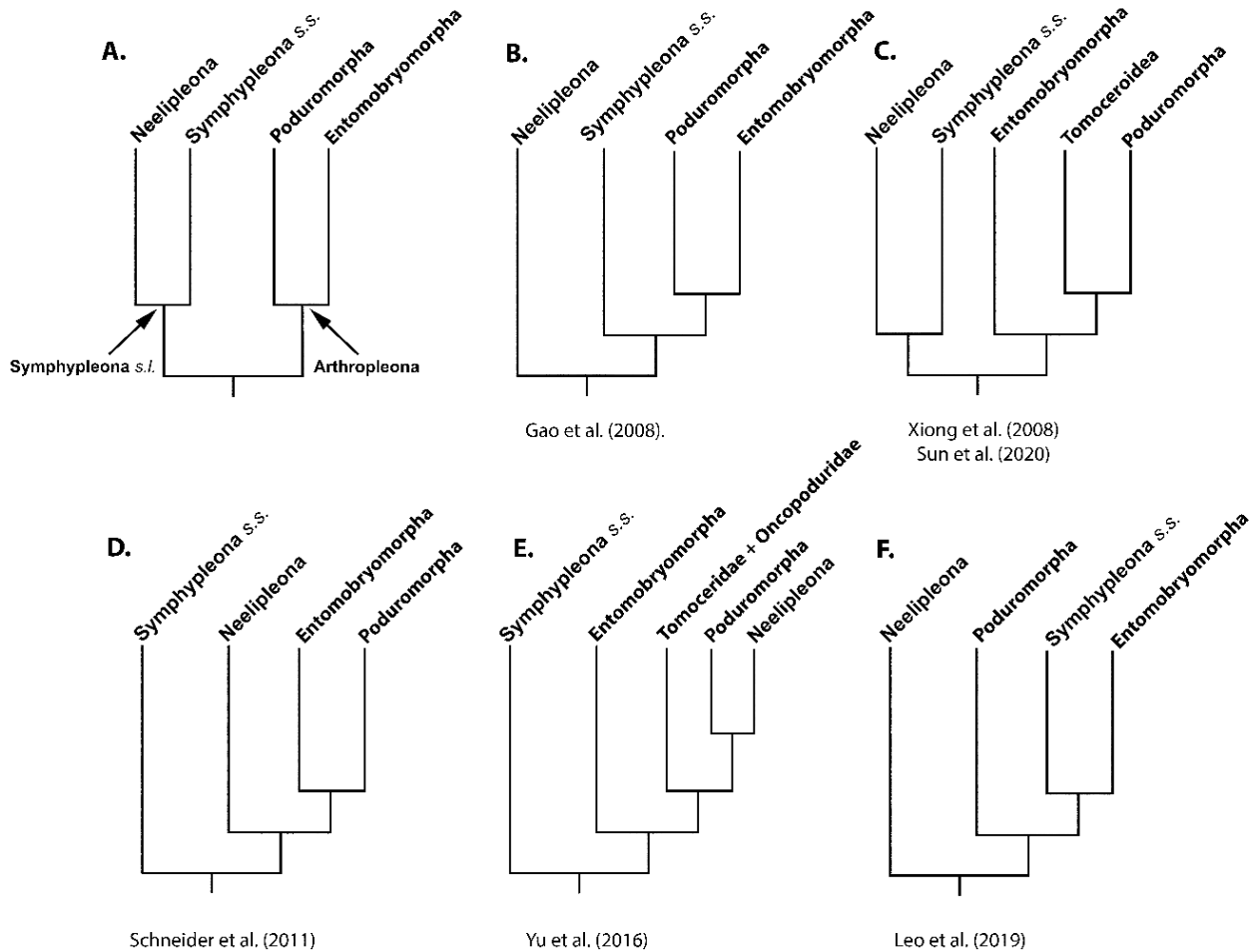
Figure 3. Collembola phylogenetic relationships proposed by different authors. (A) Traditional view; (B) Based on ribosomal markers; based on ribosomal and mtDNA protein coding genes; (D) Based on *16S*, *cox1* and *28S* markers; (E) Based on ribosomal markers; (F) Based on mitochondrial protein coding genes.

In this context, complete mitochondrial genomes appear to be a suitable marker to unravel phylogenetic relationships at the order to species level and have become very popular in the last two decades (Nardi et al., 2003; Carapelli et al., 2019). Several features make mitogenomes an interesting genetic marker: they lack recombination, they are uniparentally inherited, they are haploid and, finally, they can be easily sequenced using NGS technologies. Furthermore, their gene conservation minimizes the possibility that paralogous genes are sampled, and the order of genes along the molecule (gene order) provides an independent set of characters that can complement sequence analyses (Bernt et al., 2013a; Boore, 1999; Boore and Brown, 1998). In June 2020, before this thesis project, twelve gene orders were known in Collembola (Fig. 4).

In summary, phylomitogenomics can be a powerful tool not only to study Antarctic species biodiversity and systematics but also to clarify higher rank relationships amongst Collembola.
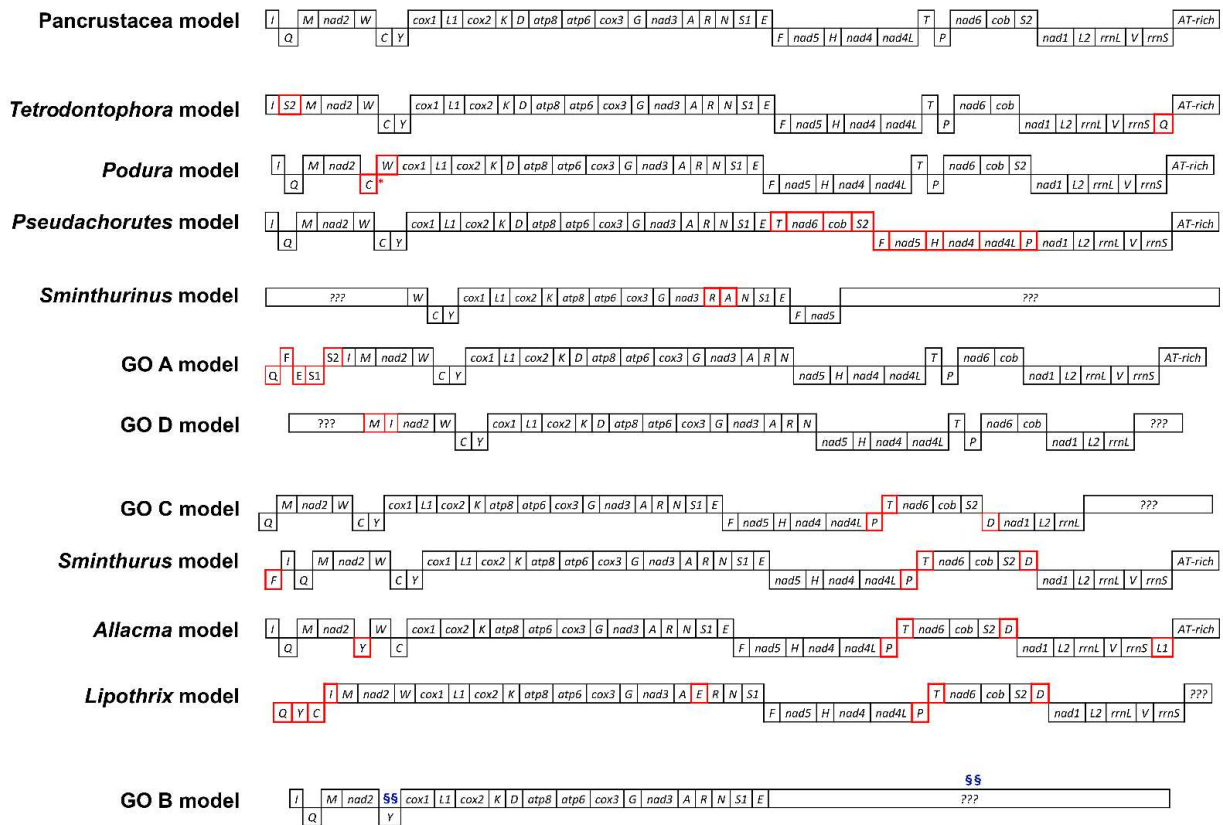
Figure 4. Collembola gene orders available up to June 2020. The Pancrustacea model is considered to be ancestral within springtails gene orders. All derived genome rearrangements, with respect to the Pancrustacean, show translocations as red boxes, deletions as red asterisks (*), and possible translocations as blue section signs (§). Incomplete mitogenome sequences are identified with question marks.

1.5. Aims

In the last decade, the wider application of Next Generation Sequencing strategies beyond the precincts of genomics has allowed ecologists and evolutionary biologists to obtain a multitude of genomic and transcriptomic data in short times and with reduced costs. The application of such technologies can vary from single cell sequencing to multiple animal pools. In this scenario, scientific fields such as evolutionary biology, comparative physiology, phylogeography and related areas grew exponentially.

This thesis is organized in two main lines of enquiry: (i) climate variation and RNA-seq analysis of *Cryptopygus terranovus* (Collembola: Isotomidae) following a mid-term heat

exposure and (ii) re-evaluation of Collembola phylogeny through the use of mitogenomic data.

The first task was addressed by studying the transcriptomic profile of this Antarctic springtail following an experimentally induced thermal stress simulating a possible climatic variation caused by global warming. During this experiment, I also collected and analysed a 1-year temperature series from natural habitats in Victoria Land (Continental Antarctica) in order to characterize the microclimatic conditions experienced by endemic invertebrates and comparing these results with surface air temperatures.

The second part of this project focussed on the application of phylomitogenomics to Collembola by sequencing the complete mitochondrial genome in two additional species of Antarctic springtails and summarising the numerous systematic hypotheses made on this taxon. Furthermore, during the development of this task, I created a bioinformatic pipeline, available as web server, to process mitogenomic data in a fast and easy way, allowing users to obtain results related to Relative Synonym Codon Usage (RSCU) and nucleotide mitochondrial biases (Hassanin et al., 2005).

2. Materials and methods

2.1. RNA-seq analysis of *C. terranovus* following a mid-term heat exposure

2.1.1. Sample collection, stressing conditions and sequencing

Specimens of *C. terranovus* were collected during the Italian National Antarctic Program (PNRA) expedition in the summer of 2018 at the Mario Zucchelli Station (MZS; Victoria Land, 74°42'0" S 164°06'45" E; at 70m altitude; Fig. 5). Individuals were subdivided in a control (CT) group and a treatment (HE) group. The temperature for the former was equal to natural habitat, namely 5.4°C (s= 3.27°C; data registered with a data logger for 20 days at the sampling location). The treated group was sampled and stressed at MZS in a plastic box 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 with natural moss. Both groups were then transferred in RNAlater solution (Sigma-Aldrich) and stored at -20°C for further processing.

Three groups of *C. terranovus* were screened for each of the CT and HE conditions (biological replicates). Each sample was composed of a pool of three individuals (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 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 the libraries were sequenced on a NovaSeq 6000 system (Illumina; 100 bp paired end reads) in ARGO Open Lab Platform for Genome Sequencing (AREA Science Park, Trieste, Italy). Raw sequencing reads are available at the Sequence Read Archive (SRA; National Center for Biotechnology Information (NCBI)) with the BioProject ID PRJNA645792.

2.1.2. De novo *transcriptome assembly, annotation and gene expression analyses*

Raw sequences were initially quality checked using FastQC and MultiQC (Ewels et al., 2016). In order to 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*. The transcriptome was assembled *de novo* using 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 the assembly completeness, 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 shorter than 200 nt were removed. *C. 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 a 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. Completeness of the final assembly 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.

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 *noiseqbio* function was applied and DEGs with a probability cut-off of 0.9999 and log2-fold change > |2| were kept for statistical investigations. An enrichment analysis based on the hypergeometrical test of the correspondent GOs was performed and only the ones represented by a significant p-value ($p < 5 \times 10^{-4}$) and observation counts > 3 were retained.

2.2. Microhabitat temperature record in Victoria Land

2.2.1. Site description and data collection

In order to study microhabitat temperatures that influence micro-invertebrate species at a biologically meaningful geographic scale in Victoria Land (Continental Antarctica) temperatures were recorded for one entire year starting from the 2019 PNRA expedition (the following year with respect to activities described in section 2.1.1). Microhabitat temperature records were obtained using four temperature data loggers (Aquatic 2 TG-4100, Tinytag) of 50mm height and 51mm diameter, with a recording range of -40°C to 70°C and a resolution of 0.01°C. Temperatures were recorded every hour in the period comprised between January 25th 2019 and January 24th 2020. Data loggers were positioned in two distinct areas

close to the Italian research station MZS, namely Campo Antenne (three specific locations; 74°42.075' S, 164°06.003' E) and Oasi (one location; 74°41.599' S, 164°06.055' E) (Fig. 5). The three locations at Campo Antenne included two in soil under rocks (Campo Antenne 1 and Campo Antenne 2) and one in a freshwater pond (small pond Campo Antenne). In Oasi, temperatures were recorded in a freshwater pond only (Oasi). These sites were chosen in order to monitor environmental conditions experienced by Acari, Collembola, Tardigrada and Rotifera, with special emphasis on *C. terranovus* populations that are endemic of this area (see section 2.1). The two data loggers recording soil temperatures were positioned below small-sized (< 50 cm) white/grey granite rocks exposed amongst the gravel and coarse-grained soil that characterize the area. The undersides of these stones can be colonized by type I hypolithic communities (Cowan et al., 2010) that provide the perfect habitat for the springtail *C. terranovus* (Carapelli et al., 2017) and the mite *Stereotydeus delicatus* (Brunetti et al., 2021). The two data loggers recording freshwater temperatures were placed on the bottom of shallow permanent ponds of ~30 m diameter and 40 cm depth that remained partly covered by ice during the summer period. The pond benthos comprised sediments and rocks of different sizes (ranging from 2 to 30 cm diameter), providing the typical microenvironment for tardigrade and rotifer species (Cesari et al., 2016; Dartnall, 2017). Standard meteorological data from the closest AWS, Eneide Station, at 82m above sea level and 300m from MZS, provided air temperatures. The distances between the AWS and the two study sites were 379 and 657m, respectively (Fig. 5). Data and information regarding the AWS were obtained from the PNRA's 'MeteoClimatological Observatory at Mario Zucchelli Station (MZS) and Victoria Land' (http://www.climantartide.it).
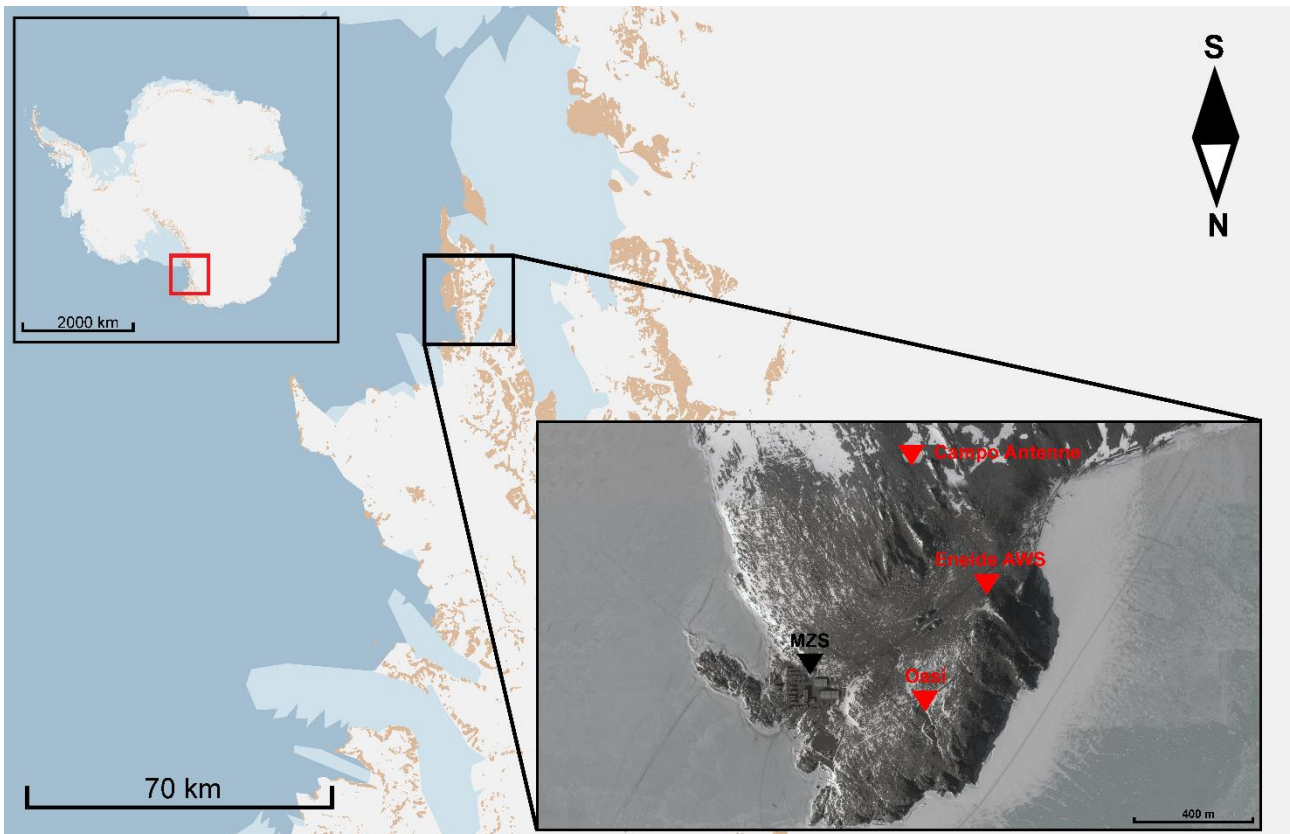
Figure 5. Sampling localities in Victoria Land. *C. terranovus* specimens used for RNA-seq analyses (paragraph 2.1) were collected near the Italian research station Mario Zucchelli Station (MZS; in black). Temperature data (paragraph 2.2) were collected at sites marked in red.

2.2.2. Data analyses

Temperature data were imported, analysed and plotted into R v.3.4.4. Cumulative degree days were calculated as the sum of mean daily temperature above 0°C multiplied by the number of days with that mean temperature following Convey et al. (2018). The thaw period, which is useful for assessing the duration of positive temperature, was calculated here as the time between thawing (> 1°C) at the beginning of the summer and refreezing (< -1°C) at its end. The 'zero curtain period' was calculated as the period when the temperature remained between -1°C and +1°C, regardless of any larger variability in external temperature, at the onset and fall of the warm season, following Convey et al. (2018).

2.3. EZmito, a web resource for a fast analysis of mitochondrial genomes

In order to provide a simple and automated workflow to analyse complete mitochondrial genomes, it was created the EZmito hub. This was presented as a freely usable web server (http://ezmito.unisi.it) able to execute the basic data preparation and visualization steps for mitochondrial DNA data analyses. The web server, written using three programming

languages (R, python and bash) is divided into five distinct tools: EZsplit, EZpipe, EZskew, EZcodon and EZmix.

2.3.1. EZsplit

The goal of EZsplit is to manage genome files downloaded from data banks, namely NCBI. EZsplit requires, as input file, a compress archive containing the Genbank multiple species nucleotide coding sequences features in FASTA format *plus* its corresponding file TinySeq XML. The program 'splits' the coding sequences following the annotations contained in the XML file, producing 13 distinct resulting output files (corresponding to the 13 mitochondrial Protein Coding Genes [PCGs]) that can be employed for the EZpipe tool.

2.3.2. EZpipe

EZpipe is designed to prepare mitochondrial PCGs data sets to be used in phylogenetic analyses. The tool requires in input a compressed archive of FASTA files corresponding to individual PCGs (as in the EZsplit output). Each file must contain nonaligned sequences with unique taxon names (common to all files) as sequence headers. Moreover, user can choose two additional input parameters: the appropriate genetic code following NCBI designations and the number of nucleotide codon positions that should be analysed (e.g. 3 for all codon positions, 2 for first and second codon positions only).

Input files undergo a first quality check to assess if the data provided are correctly formatted and suitable for downstream analyses. The analysis continues, but warnings are displayed if: (i) duplicated sequences are found within the data set; (ii) the length of a sequence differs by more than two standard deviations from the mean length of all sequences (e.g. if a gene is truncated); (iii) gaps are found within sequences – in this case, they will be automatically removed; and (iv) sequences display a truncated end codon – which will be automatically removed. On the other hand, an error is displayed, and hence the analysis stopped if: (i) the input archive is not prepared as required; (ii) input files are not in FASTA format; (iii) duplicated sequence IDs are present within a file; (iv) non-IUPAC nucleotides are observed in a sequence; and v) stop codons are present within a sequence (i.e. not counting terminal stop codons).

After sanity check, sequences in each file are retro-aligned using RevTrans (Wernersson, 2003) based on the genetic code designated by the user. Hyper-variable regions of the alignment are further discarded using Gblocks (Castresana, 2000), removing full codons through *strict* and *codon* option. Finally, single-gene alignments are concatenated in R through the *concatenateAlignments* function and converted in the PHYLIP format.

Furthermore, a PartitionFinder2 configuration file is created where a starting partitioning scheme is designed subdividing the final alignment by gene and by codon. These two latter files, the concatenated alignment in PHYLIP format and the PartitionFinder2 configuration file, are the main output of EZpipe. Moreover, three log files are written with a different level of detail, of which the log.txt file includes information relevant for the user and can provide important information for troubleshooting.

EZpipe does not include neither a model selection step nor a tree-building step because of hosting server computing limitations. Nevertheless, the two resulting files can be readily used in downstream applications. The concatenated alignment and the PartitionFinder2 configuration file are meant to be directly used as input for PartitionFinder2 (Lanfear et al., 2016) to optimize the partitioning scheme and associated evolutionary models.

2.3.3. EZskew

EZskew is able to calculate nucleotide biases by protein-encoding gene, strand, and codon position. It takes as input a compressed archive including two folders (J and N), each containing a FASTA file of nonaligned protein-encoding genes in the expected orientation. Moreover, the user must choose the appropriate genetic code as an input parameter. An initial sanity check is performed as described in 2.3.2. Then, each PCG is padded with gaps and concatenated to produce three final data matrices: full genome data, J-strand only, and N-strand only. Concatenated matrix files are used for the calculation of compositional skews following Hassanin et al. (2005). AT%, is calculated over the entire PCG concatenated matrix, while AC% and GT% are calculated over the concatenated J- and N-strands, respectively. These values are plotted in a single histogram by genome. Codon position biases, namely the AT and CG skews, are calculated over the J- and N-strand separately for first, second, and third codon positions using formulas:

$$ATskew = \frac{A - T}{A + T}$$

$$CGskew = \frac{C - G}{C + G}$$

and plotted separately for each codon position. A text-based tabular output is also produced, as well as log files. In the attempt to produce publication-quality graphics, the level of detail in figures automatically adapts to the number of genomes being analysed.

2.3.4. EZcodon

EZcodon is designed to calculate amino acid frequencies and RSCU over different mtDNA genomes. It requires the same EZskew input files. Initial steps are those described for EZskew in 2.3.3 to obtain full genome data in coding orientation. This matrix is used to calculate the frequency of each amino acid as well as the RSCU table using the CAI python package (Lee, 2018). Non-conventional amino acids (following IUPAC nomenclature) are deleted during this step. Two final graphical outputs are produced: (i) amino acid frequencies in all genes as well as in genes encoded in the J- and N-strand separately; and (ii) RSCU values for each genome. A text-based output and log files are also produced. As for EZskew, the level of detail in output figures automatically adapts as a function of the number of genomes being analysed.

2.3.5. EZmix

EZmix helps users to recognize areas of inter molecular similarity that may be indicative of the assembly of chimeric mitochondrial genomes. The tool produces a graphical output where such regions are highlighted. EZmix was designed for projects where a non-barcoded library including pooled DNAs from different species is sequenced as a pool (such as Cucini et al., 2020, 2021b; Nardi et al., 2003). While a correct assembly should avoid the possibility that sequences from different species are mixed in one assembled candidate genome, this possibility cannot be ruled out *a priori*. As such, visualizing similarities between the complete set of candidate genomes may be a useful step in the quality control of assemblies. Noteworthy, areas of inter molecular similarity may arise from phylogenetic relatedness (not indicative of an assembly error) or from chimeric assemblies (indicative of an error). The program requires a FASTA file with whole candidate mitogenome sequences, the percentage of similarity to be used as threshold and the minimal length for a match. After a sanity check similar to the one described in 2.3.2 EZmix identifies regions of similarity between every pair of sequences through the use of BLASTn and plots the results in a colour-coded output.

2.4. Phylomitogenomics of Collembola

2.4.1. K. klovstadi *and* T. mixta *sample collection, mitochondrial DNA sequencing and assembly*

Specimens of *T. mixta* were collected in January 2003 at Harmony Point, Nelson Island (South Shetland Islands; 62°13'20.7" S, 58°46'54.0" W) during the 2002/2003 field season of the British Antarctic Survey (BAS). Specimens of *K. klovstadi* were collected in January

2019 at Redcastle Ridge (Continental Antarctica; 72° 26' 25.0" S, 169° 56' 32.0" E) during the PNRA expedition in the summer of 2018. Total DNA was extracted using the QIAmp® UCP DNA kit from pools of 10–15 individuals of each species. Extractions were normalized, pooled, and sequenced at Macrogen Europe using a TruSeq Nano DNA chemistry on a NovaSeq 6000 platform (Illumina, San Diego, CA, USA), together with additional samples not part of this thesis project.

Resulting reads were quality controlled using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and assembled using NOVOPlasty v.3.8.3 (Dierckxsens et al., 2017). The assembly step was carried out using the public available *cox1* sequences for *K. klovstadi* downloaded from the Bold System (accession number: GBCO0087/88/89/90-06) and the *T. mixta* partial sequence downloaded from GenBank (accession number: KF982833) as seeds. Initial assemblies were obtained under default settings. These were employed as a filter to create a subset of the reads using the *filter_reads.pl* script from the same package. Subsequently, final assemblies were produced, based on this enriched library, using the full data. Resulting contigs were compared to MEGAHIT (Li et al., 2015) assemblies and manually curated to produce full mitochondrial genome assemblies. Single base ambiguities were resolved on a majority rules basis by remapping using bbmap v.38.84 (sourceforge.net/projects/bbmap/) and visualizing alignments in IGV v.2.8.2 (Robinson, 2011). The possibility of a chimeric assembly was excluded using EZmix (Cucini et al., 2021a). Genomes were preliminarily annotated using Mitos (Bernt et al., 2013b) and manually curated based on known annotations of closely related species. The new mitogenomes have been deposited in GenBank (accession numbers: MW238521 and MW238520), while raw data have been deposited at the SRA archive at the NCBI portal (accession number: SRP289641).

2.4.2. Phylogenetic analyses and tree topology tests

All available complete or semi-complete springtail mitogenomes were downloaded from GenBank (NCBI) in June 2020 together with those of three outgroups, *Daphnia pulex* (Crustacea, Branchiopoda), *Japyx solifugus* (Diplura, Japygidae) and *Trigoniophthalmus alternatus* (Microcoryphia, Machilidae) and preprocessed using the EZsplit program (Cucini et al., 2021a). The final dataset was obtained by adding *K. klovstadi* and *T. mixta* sequences along with those presented by Cicconardi et al. (2017) and Leo et al. (2019). Published mitogenomes which lacked annotations (Godeiro et al., 2020) were automatically annotated using Mitos (Bernt et al., 2013b). A manual revision of genomes which displayed unusual features (long spacers, non-canonical gene order) was carried out and genes that appeared

truncated compared to reference sequences were discarded. PCGs were formatted according to EZpipe instructions and submitted to the web server (Cucini et al., 2021a).

The final matrix was divided in blocks by codon position, gene family and strand, as described by Leo et al. (2019) and submitted to PartitionFinder v.2.1.1 (Lanfear et al., 2016) through the CIPRES Science Gateway (Miller et al., 2010) to determine the best partitioning scheme and associated evolutionary model. GTR+I+Γ was the one selected for all partitions except for the *atp8*, *nad6* and the first codon position of *nad5* and *nad4L*, for which the GTR+Γ was selected. Data matrices, as well as the optimal evolutionary model and partitioning scheme produced by PartitionFinder, were analysed in MrBayes 3.2.6 (Ronquist et al., 2012) using 8 chains and 100 million generations. The final output was summarized in tracer v.1.7.1 (Rambaut et al., 2018) excluding the first 25% as burn-in. The final tree was visualized using Figtree v.1.4.4 (Suchard et al., 2018).

To evaluate the support for some groups that were recovered as polyphyletic in the analysis, constrained tree topologies were investigated with 10,000 replicates through IQ-TREE under the RELL (Re-Estimated Log Likelihood) approximation method (Kishino et al., 1990). Six different hypotheses were compared following Sun and colleagues (2020): (i) best unconstrained tree; (ii) monophyletic Neelipleona; (iii) monophyletic Tomoceridae; (iv) monophyletic Hypogastruridae; (v) monophyletic Paronellidae; (vi) monophyletic Entomobryidae.

2.4.3. Nucleotide biases and Ka/Ks ration

Nucleotide biases were investigated thanks to EZskew in EZmito web server (Cucini et al., 2021a). Final output matrices were divided into J-strand and N-strand oriented genes and, nucleotide skews were calculated as described in section 2.3.3 for first (J1 and N1), second (J2 and N2) and third codon positions of 2-fold-degenerate sites (2J3 and 2N3) and third codon position of 4-fold-degenerate sites (4J3 and 4N3). Outliers were defined as those sequences which showed a value beyond three times the interquartile range from the 25–75% quartiles. AT percentage was assessed on complete mtDNA sequences and compared with all Hexapoda mitochondrial genomes present in the organelle database (NCBI) in June 2020 through the EZskew web resource (Cucini et al., 2021a)

To check for positive or negative selection, the non-synonymous to synonymous mutation ratio (Ka/Ks) was evaluated on the same alignment after outgroups removal. The alignment was processed as described in section 2.4.2 to obtain a phylogenetic tree that was used to test for selection in EasyCodeML (Gao et al., 2019) using the *preset* running mode with a

site model setup. The Ka/Ks ratio was manually controlled and accepted for all models that showed a likelihood ratio test (LRT) with a significant p-value (≤ 0.5).

3. Results

3.1. RNA-seq analysis of *C. terranovus* following a mid-term heat exposure

3.1.1. C. terranovus' *heat exposure*

Springtails belonging to the control group were sampled in the wild environment after 20 days of data recording. Recorded climatic conditions of the soil (registered every 10 min between 27 December 2017 and 15 January 2018) varied between 0.11°C and 12.60°C, with and average temperature of 5.4°C (s=3.27°C). Hence, a difference of about 12.5°C was assessed between the CT and HE group. However, to better understand the yearly temperature trend which influence these species, during the following PNRA expedition (2019-2020) temperature data from these microhabitats were acquired and analysed apart as described in section 3.2.

3.1.2. De novo *transcriptome assembly*

Illumina sequencing produced ~14–41 million reads per sample (Tab. 1). The *de novo* assembled transcriptome accounted for a total of 64,054 contigs with an average length of 369.25 nt with a N50, defined as the sequence length of the shortest contig at 50% of the total transcriptome length was 345 nt.

Since *C. terranovus* is neither a model organism nor phylogenetically close to a model organism, the BUSCO report against the Arthropoda database indicated an acceptable completeness though highlighting a certain degree of fragmentation: 29.4% of genes recovered as complete (21.7% single copy, 7.7% duplicated), 47.1% fragmented and 23.5% missing BUSCOs. The final GC content in of the transcriptome was 45.53 ± 6.08%.

Label	Sample	Number of raw reads
CT1	Control_1	41.4 million
CT2	Control_2	16.8 million
CT3	Control_3	17.4 million
HE1	Heated_1	18.8 million
HE2	Heated_2	15.8 million
HE3	Heated_3	14.7 million

Table 1. Number of reads per sample and sample ID for control (CT) and heated (HE) conditions

3.1.3. 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 (Tab. 2). 42,309 genes were catalogued with an OrthoDB hit, 34,143 genes were annotated in Uniprot-Swissprot and 15,936 genes contained at least one PFAM domain (Fig. 6).
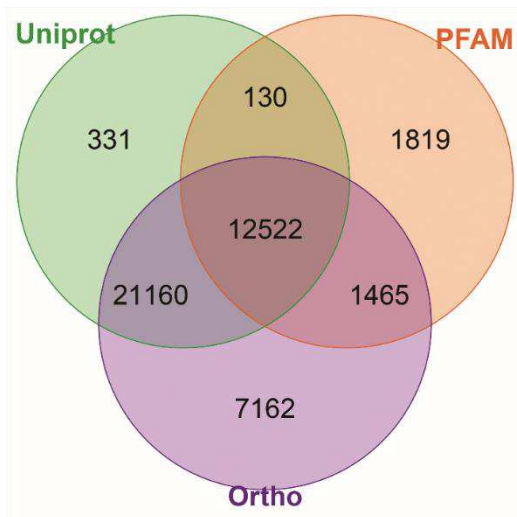

Figure 6. Venn diagram of the annotated genes shared by and unique to each database. Numbers shown in the graph correspond to the number of contigs in each dataset. Orthologs represent the majority within them. The 13% of all annotations are common to all three databases.

The GO annotation retrieved 12,747 terms, of which 7,998 represented biological processes (BPs), 3,317 were catalogued as molecular functions (MFs) and 1,432 as cellular components (CCs). GO terms covered 13 different GO levels. In detail, most of the terms were catalogued in between the 5th and 13th levels: with 78.3% GO terms for BPs, 52.3% terms for CCs and 68.2% terms for MFs, indicating that the functional annotation led to a good resolution (Fig. 7).

3.1.4. Differentially expressed genes and enrichment analysis

The differential expression analysis retrieved a total of 7,637 differentially expressed transcripts, of which 3,719 upregulated and 3,918 downregulated in the HE samples (Tab. 2). However, 29% of the total DEGs lacked a meaningful functional annotation. A significant difference in transcript expression between HE and CT samples was also confirmed by the principal component analysis and the heatmap, as both displayed a clear clustering of the two conditions (Figs. 8,9).

Overall, the comparison of non-rare GOs (namely above the 0.5th percentile) between the annotated genes and DEGs resulted in a limited number of differences (Fig. 10). Similarly, the comparative assessment of the most expressed BPs, CCs and MFs between the HE and CT conditions did not show remarkable divergences, except for a frequency bias of a few GOs (Figs. 11-13).
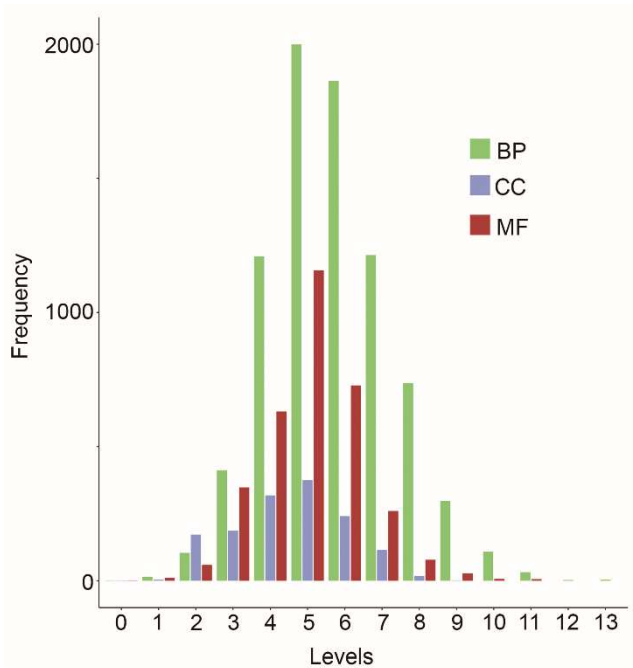

Figure 7. Gene Ontology terms' level of representation for biological processes (BPs) in green, cellular components (CCs) in blue and molecular functions (MFs) in red. The maximum resolution was achieved between the 5th and 13th levels.

|	Number of transcripts
Total transcriptome	64,039
Annotated transcriptome	45,303
Differential expressed genes	7,637
HE up-regulated genes	3,719
HE down-regulated genes	3,918
Annotated differentially expressed genes	5,425

Table 2. Summary of the transcript numbers processed during the functional analysis of *C. terranovus* transcriptome. HE= heated samples.
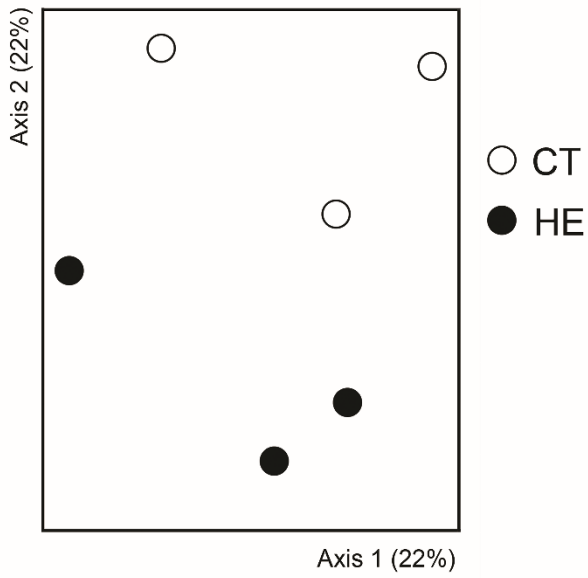
Figure 8. Principal component analysis of control (CT) and heated (HE) group.
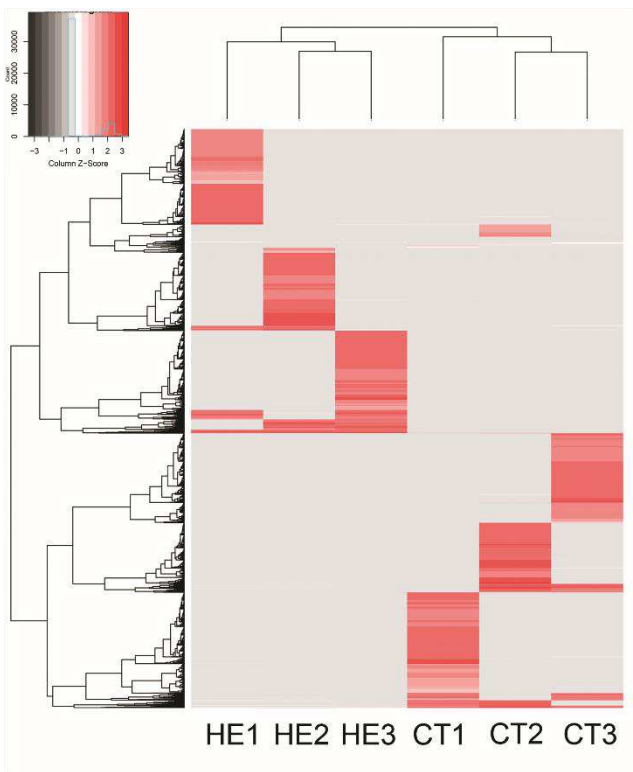
Figure 9. Heatmap summarizing differentially expressed genes between the control (CT) group and the heated (HT) group using a the $\log_{10}$-transformed ratio. The statistical analysis was conducted with the NOISeq package using a probability threshold of 0.9999.
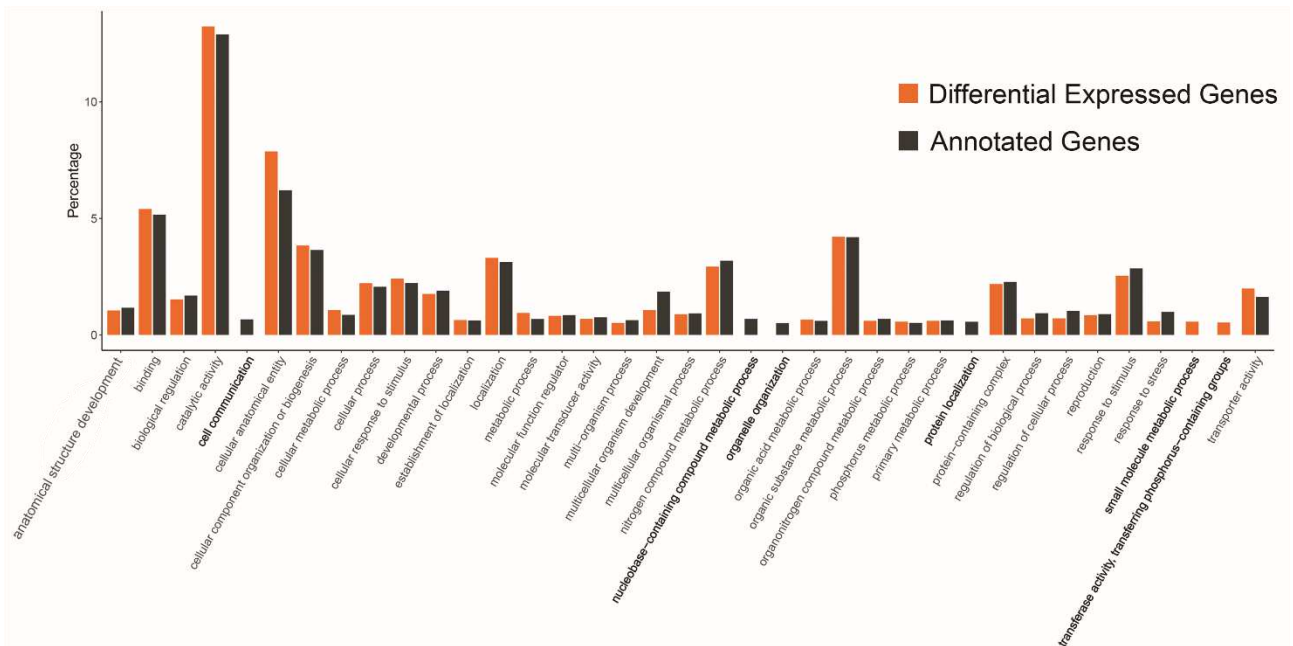
Figure 10. Differentially expressed (orange) and annotated genes (dark grey) Gene Ontology abundance differences at the 99.5th percentile. Diverging terms between the two datasets are highlighted in bold.
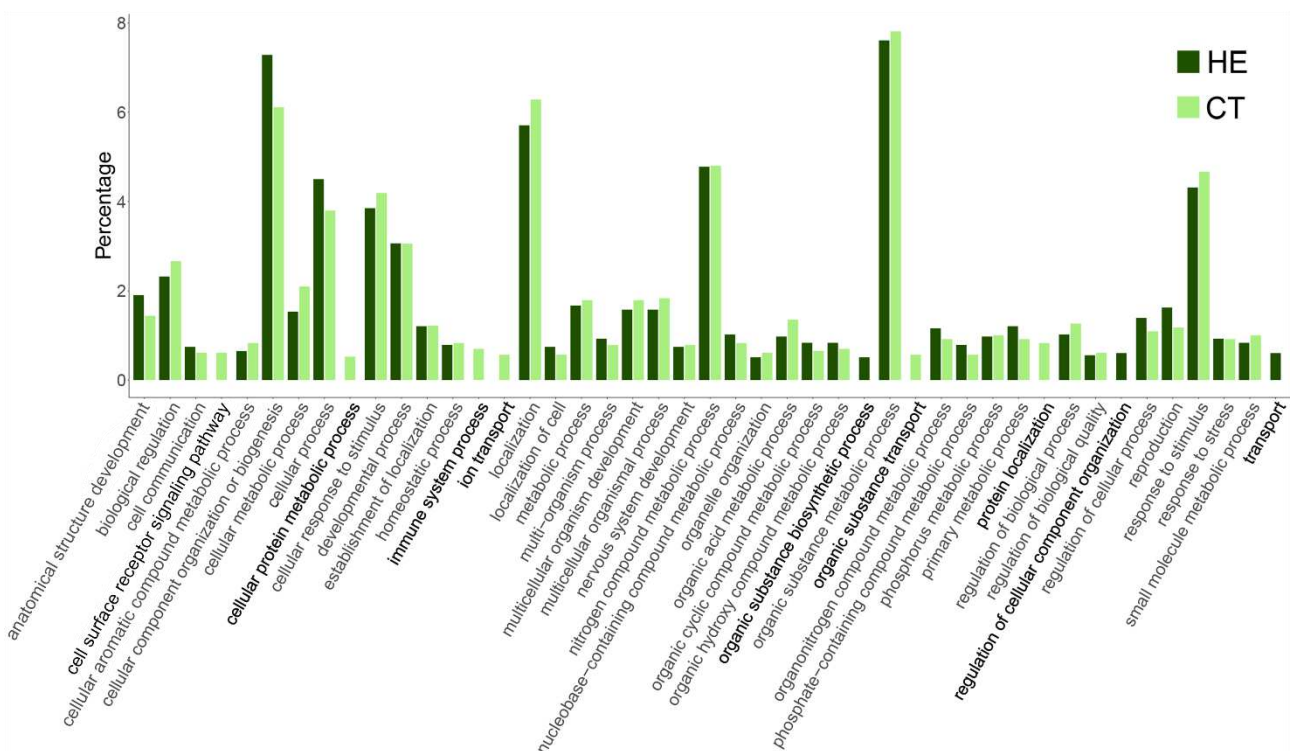
Figure 11. Biological processes terms comparison between the heated (HE) group (dark green) and the control (CT) group (light green) at the 99.5th percentile. Expression differences between the two conditions are highlighted in bold.

28
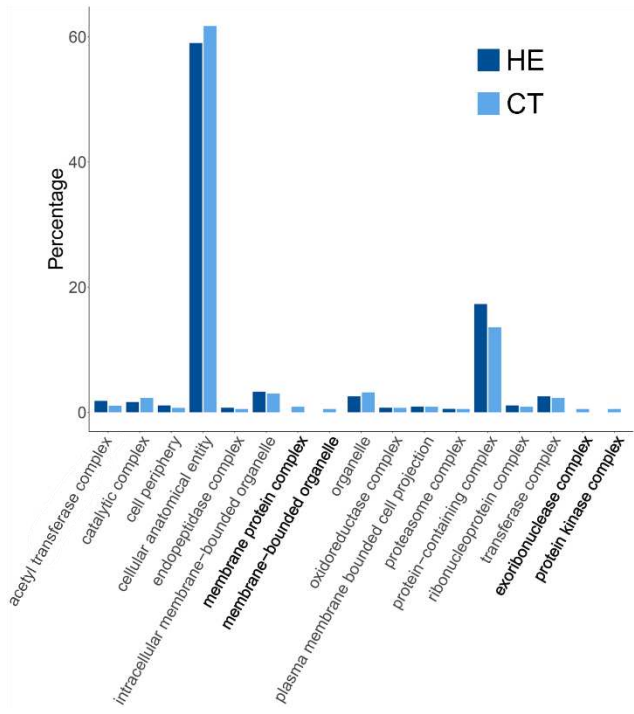
Figure 12. Cellular component terms comparison between the heated (HE) group (dark blue) and the control (CT) group (light blue) at the 99.5th percentile. Expression differences between the two conditions are highlighted in bold.
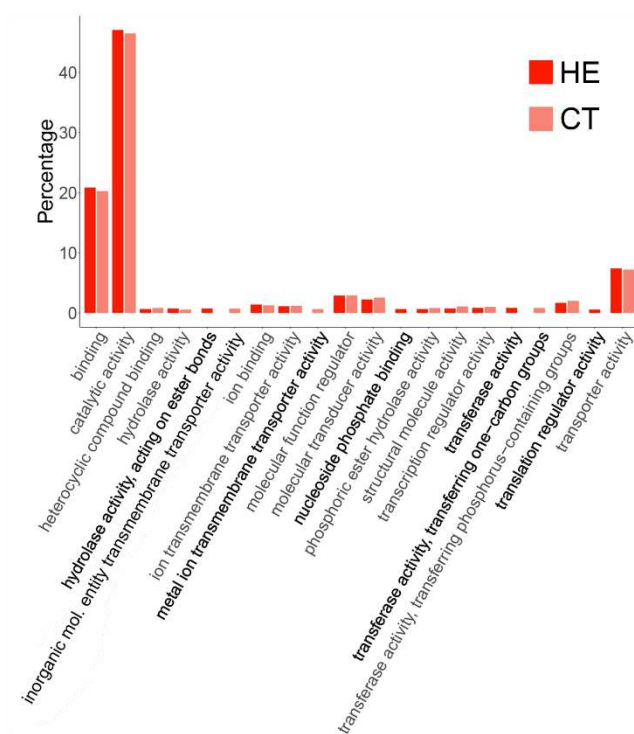

Figure 13. Molecular function terms comparison between the heated (HE) group (dark red) and the control (CT) group (light red) at the 99.5th percentile. Expression differences between the two conditions are highlighted in bold.

The enrichment analyses disclosed significant differences (Fig. 14) with 39 significant terms (19 MFs, 15 BPs and 5 CCs) retrieved as differentially represented. Although some of them did not show a readily visible dissimilarity between the two groups at face value, probably because of a similar transcript co-expression, the majority of enriched GOs were characterized by a noticeable expression difference (Fig. 14).
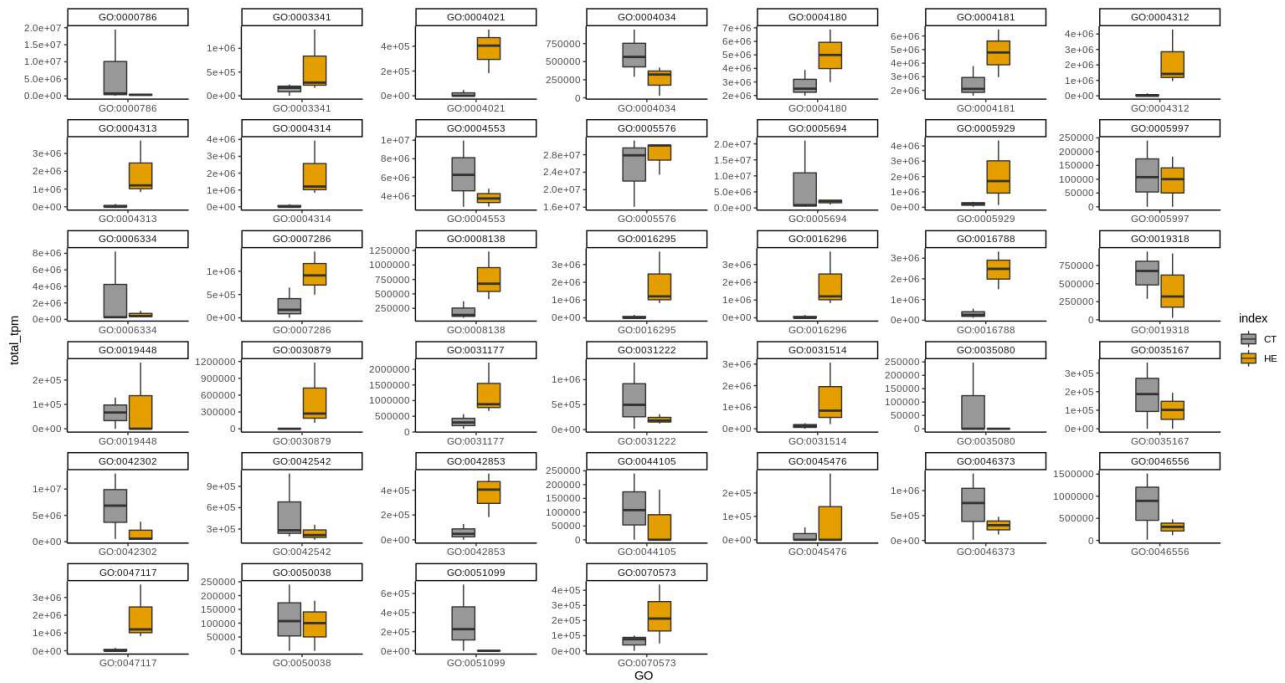
Figure 14. Enrichment analyses boxplots of GO terms between the two springtails conditions: control (CT) in grey and treated (HE) in orange.

3.2. Microhabitats temperatures in Victoria Land

Following one whole year of temperature data recording, four temporal phases were identified (Fig. 15). From November to February (Antarctic summer) average daily temperatures of soils and ponds ranged between -5°C and +6°C. In March (Antarctic autumn) temperatures rapidly dropped with maxima and averages below 0°C and -10°C respectively, with almost no positive temperatures recorded. From April to September (Antarctic winter) the lowest temperatures were observed, with daily averages ranging from -18°C to -22°C. In contrast, in October (at beginning of spring) temperatures began increasing with a substantial and rapid increment in November (Fig. 15).
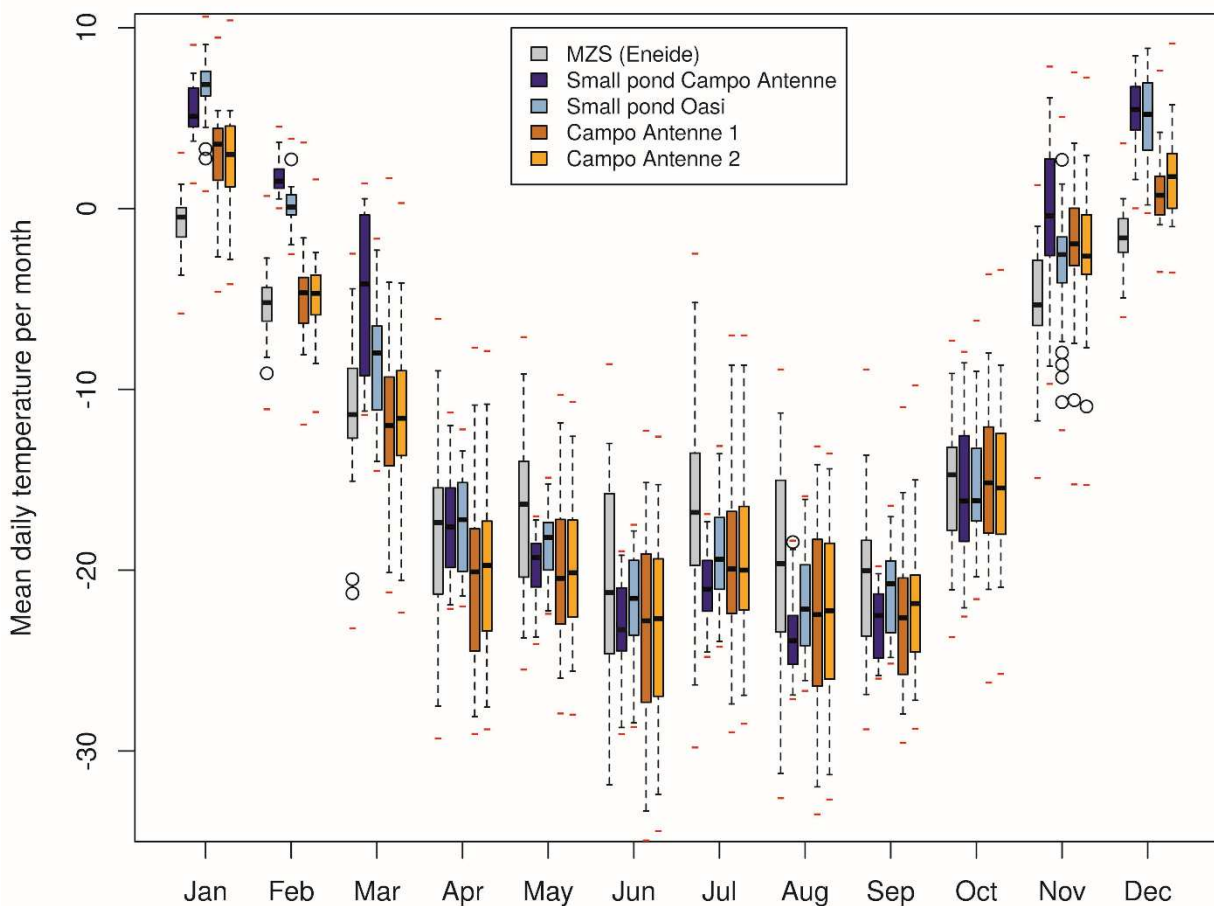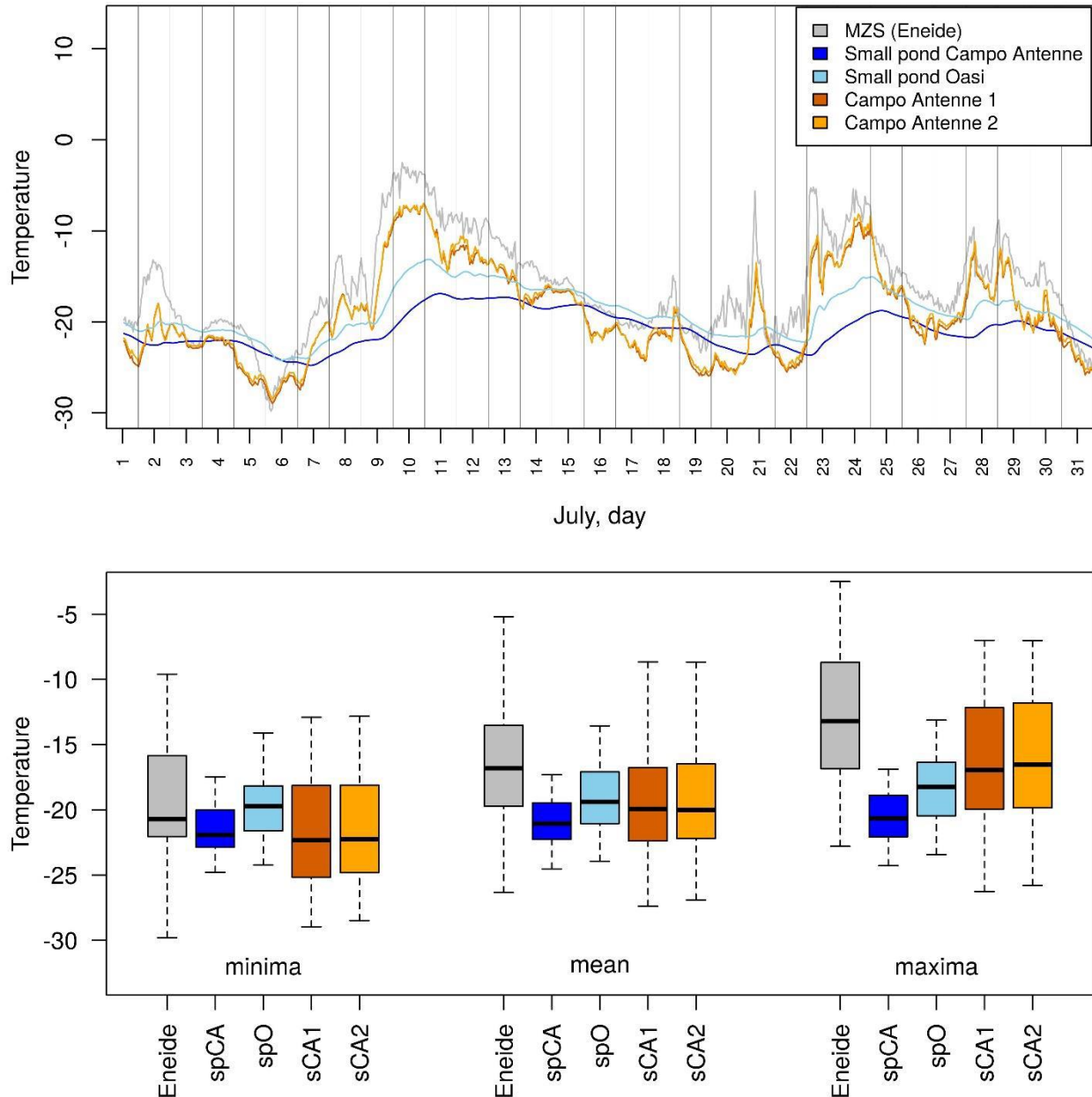

Figure 15. Mean daily temperature distribution by month. Boxes indicate the interquartile range, whiskers indicate quartiles ±1.5 interquartile range, black marks indicate monthly medians, red marks indicate point (hourly) absolute minima and maxima and empty circles indicate outliers.

The absolute minimum temperature (-35.8°C) was registered by the AWS whereas the absolute maximum (11.9°C) was recorded in Oasi pond during summer (Tab. 3). Although annual means differences were overall limited (up to ~2°C; Tab. 3), large variations across sites and between microhabitats and air temperatures were observed in specific periods of the year (Fig. 15). In detail, during summer a strong dissimilarity was registered between Oasi and Eneide AWS (7.5°C) while in winter and spring a general uniformity was observed (0°C between Campo Antenne 2 and Oasi in July and 0.34°C between the ponds in October). Moreover, the Eneide AWS generally registered lower temperatures than the ponds and soils during summer/autumn but marginally warmer or comparable temperatures during winter/spring. This last result is at odd with the observation that microhabitats, at variance with the AWS, are covered by snow and hence are expected to be thermally buffered. In order to investigate this aspect further, July temperatures (the month where this effect is more evident) were considered in more detail (Fig. 16). Temperature distribution showed coordinated fluctuations, but temperature rates of change were observed to be different across sites. Indeed, in days when a rapid temperature increase was observed in all the localities (e.g. 9, 21 and 23 July), the AWS registered the quickest and highest rates of temperature change, followed by the soil and pond sites. Similarly, also the comparison of daily minima/means/maxima followed the same trend. While minima were similar across sites, means and maxima were slightly higher at the AWS in comparison with the other localities (Fig. 16). This result suggests that thermal buffering, due to snow cover and the thermal capacities of soil and water, may have led to the decrease in average temperatures, since it did not allow soil and pond sites to gain thermal energy from short warm periods in winter.

| Locality	Microhabitat	Abs. minimum (°C)	Abs. maximum (°C)	Annual mean (°C)
Campo Ant. 1	Soil	-34.9	9.4	-13.2 (-33.3 to 5.4)
Campo Ant. 2	Soil	-34.4	10.4	-13.1 (-32.4 to 5.7)
Campo Ant.	Small pond	-29.1	11.0	-11.4 (-28.7 to 8.4)
Oasi	Small pond	-28.7	11.9	-11.3 (-28.5 to 9.1)
Eneide AWS	Air	-35.8	3.6	-12.7 (-31.9 to 1.3)

Table 3. Absolute minimum, maximum, annual mean (with daily ranges, namely minimum and maximum of mean daily temperatures) for each of the monitoring sites and the Eneide automatic weather station (AWS). AWS recording site is located at 2m from the ground. Soil refers to soil under rock sites. Abs = Absolute.

Figure 16. Temperatures registered in July 2019 at the 5 sites, detail of hourly measurements (upper panel). Distribution of daily minima/means/maxima at the 5 sites (lower panel) Eneide: MZS (Eneide); spCA: Small pond Campo Antenne; spO: Small pond Oasi; sCA1: Campo Antenne 1; sCA2: Campo Antenne 2.

Differences between sites were also evident in daily temperature excursions (Fig. 17). If Eneide AWS displayed a daily temperature excursion of ±3°C and ±4°C (in the warmest season) and up to ±6°C (in the coldest period), microhabitat sites followed the opposite trend, with the highest daily excursions taking place during summer and the lowest during winter. In detail, soil sites daily excursions from March (autumn) to September (winter) were overall limited, with a range of ±4°C to ±5°C which increased up to ±6.0°C to ±8.2°C from October to February. Values registered at the two pond sites followed the same trend

although with lower variability (±0.8°C and ±1.4°C in autumn and winter and ±3°C to ±5°C in summer).
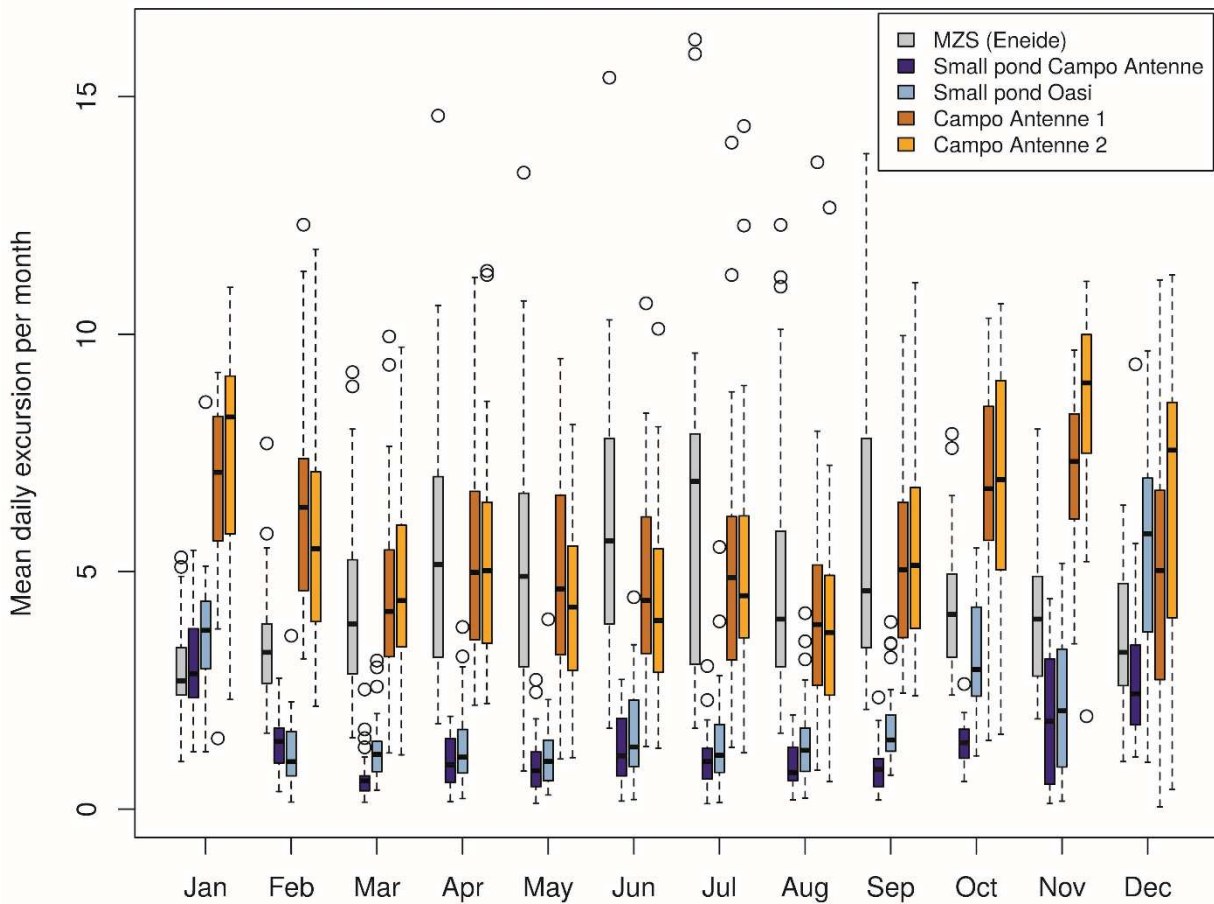
Figure 17. Dily excursions distribution by month. Boxes indicate the interquartile range, whiskers indicate quartiles ±1.5 interquartile range, black marks indicate monthly medians and empty circles indicate outliers. MZS = Mario Zucchelli Station.

During the biologically active season, namely in the warmer months (November–March), temperature trends at the locations varied substantially (Fig. 18). Albeit mean daily temperatures in December and January were consistently above 0°C in all four microhabitats (with few exceptions), peaks of sub-zero temperatures were observed for short time intervals. In the two pond sites, sub-zero temperatures were assessed on only two different days (average 11.5 h per day) at Oasi, and not at Campo Antenne. In the two soil sites, sub-zero temperatures were recorded on 46 and 52 distinct days (Campo Antenne 1 and Campo Antenne 2, respectively) with 11.4 h and 10.1 h as averages. In the Eneide AWS, sub-zero temperatures were recorded on all days except for one for an average of 17.6 h.

Given that temperatures began to increase at the end of winter, the first site which registered this increment was small pond at Campo Antenne in mid-November, followed by the two soil sites and, finally, the pond at Oasi (end of November). The last locality which recorded such variation was the Eneide AWS with the first positive air temperature data only in late December (Fig. 18). From December to January, the general temperature trends were similar across all four microhabitat sites with the exception of mean air temperatures which were generally in the range -2°C to 0°C, with sporadic peaks > 1°C and < -3°C. The soil and pond sites were warmer than the air by ~4°C and 8°C, respectively, ranging between 0°C and 9°C (Fig. 18). At the end of January, soil and pond site temperatures started to gradually decrease. With respect to air and soil sites, where temperatures fell earlier and more rapidly, ponds experienced a slower temperature decline remaining warmer until mid-March.

The sum of cumulative degree days displayed a constant increase at the soil and pond sites until the end of January when it stabilized (Fig. 18), indicating the beginning of autumn. The heating rate differed between the two microhabitat locations since ponds accumulated approximately twice as much energy as the soil sites. On the other hand, the cumulative degree days in air temperature remained almost flat, with only rarely and briefly peaks being above the freezing point.

Similarly, the air temperatures experienced a shorter thaw period (two distinct 7-day periods) with respect to the soil (64–66 days) and pond microhabitats (87–112 days). The difference between microhabitats was due to a substantial shift in the freezing dates (January–March), while spring thaws were almost simultaneous (Tab. 4).

Finally, the zero curtain period was longer in ponds (4–9 thawing days and 14 freezing days) than the soil microhabitats (2 thawing days and 3–4 freezing days) (Tab. 5).
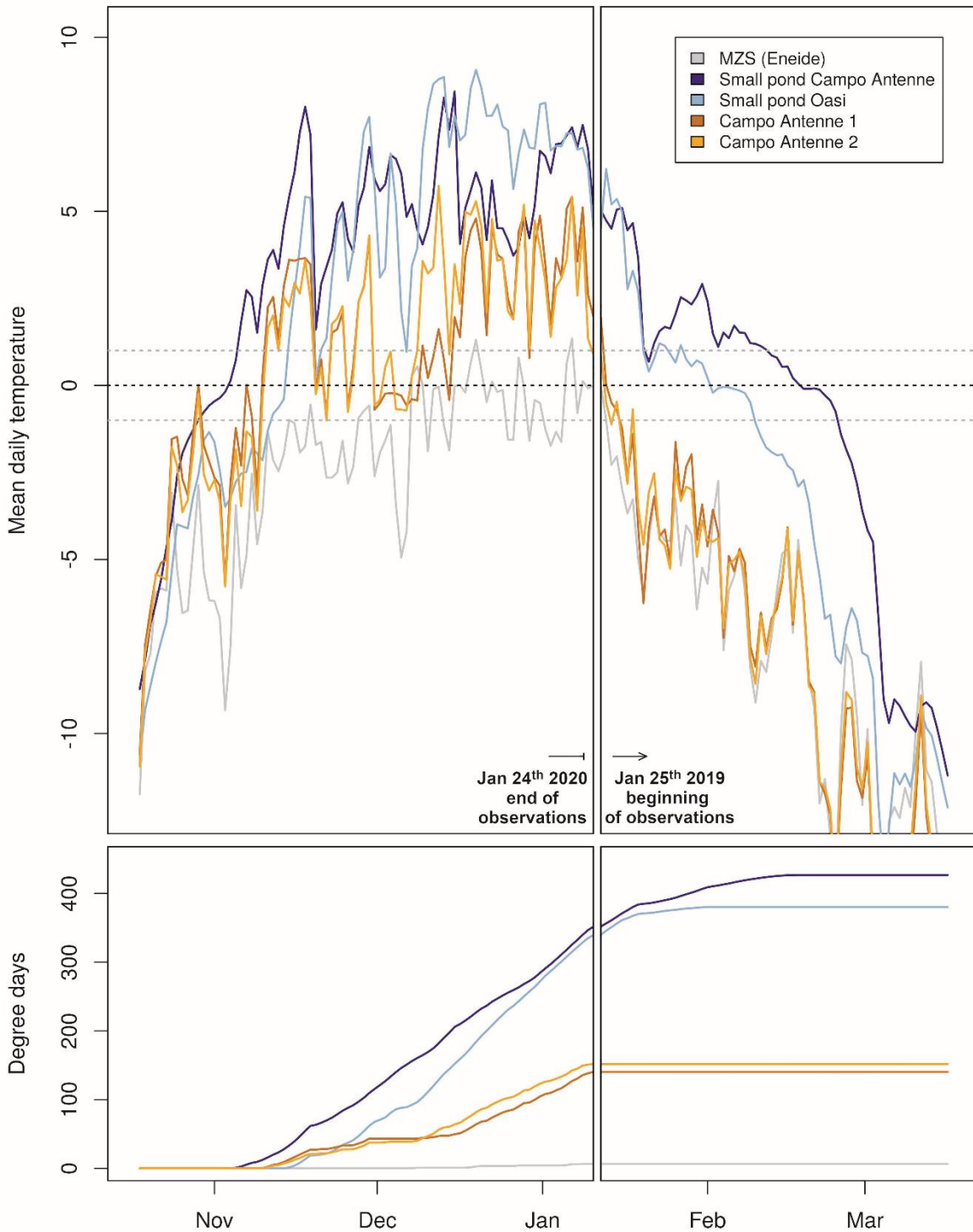
Figure 18. Mean daily temperatures during the biologically active season. Black dotted line indicates 0°C, grey dotted lines indicate ±1°C (upper panel). Cumulative degree days for each site (lower panel). Figure cut separates data collected in 2019 (right) and 2020 (left). Sites are colour coded following the legend in upper panel. MZS = Mario Zucchelli Station.

Site	Thawing (>1°C)	Freezing (<-1°C)	Thawing days
Eneide AWS	*	*	7+7
Small pond Campo Antenne	Nov 20	Mar 11	112
Small pond Oasi	Nov 29	Feb 23	87
Soil under rocks Campo Antenne 1	Nov 25	Jan 29	66
Soil under rocks Campo Antenne 2	Nov 25	Jan 27	64

Table 4. Seasonal thaw and freeze dates, i.e. the first day (at the beginning of the biologically active season) with a mean daily temperature > 1°C and the first day (at the end of the active season) with a mean daily temperature < -1°C. Intervals are given as inclusive of the two extremes.

Site	Beginning of active season			Ending of active season		
	Interval dates		Z.C. days	Interval dates		Z.C. days
Small pond Campo Antenne	Nov 12	Nov 20	9	Feb 26	Mar 11	14
Small pond Oasi	Nov 26	Nov 29	4	Feb 10	Feb 23	14
Soil under rocks Campo Ant. 1	Nov 24	Nov 25	2	Jan 26	Jan 29	4
Soil under rocks Campo Ant. 2	Nov 24	Nov 25	2	Jan 25	Jan 27	3

Table 5. Zero curtain (ZC) periods at the beginning and end of the biologically active season. The initiation and end dates and the ZC length (days) are indicated

3.3. Collembola phylogenesis using EZmito as web resource

3.3.1. *K. klovstadi and T. mixta mitogenomes*

EZmix did not retrieve any chimeric region between the two genomes, confirming that the assembly step plus the manual revision produced correct resulting genomes (Fig. 19). *K. klovstadi* and *T. mixta*'s genomes differed by 488 bp in length, with the former being 15,486 bp in length and the latter 14,998 bp length. Both mtDNAs showed the typical metazoan features with 13 PCGs (*atp6, 8; cox1-3; cytb; nad1-6, 4L*), two rRNAs (*rrnL, rrnS*) and, in *K. klovstadi*, a complete set of 22 *tRNAs* (henceforth referred to as *trnX*, where *X* is the amino-acid corresponding letter). On the other hand, *T. mixta* lacks *trnC*, which might represent a true gene loss, although the possibility that its structure may diverge from common *tRNAs* and may not have been recognized cannot be excluded in theory. Canonical methionine start codons (ATA/ATG) characterized the majority of PCGs (8 in *T. mixta* and 6 in *K. klovstadi*). An alternative leucine start codon (TTG) was present in nad4L of both genomes while ATT/ATC (isoleucine) was observed in all other genes. In the *K. klovstadi* mtDNA partial stop codons were frequent (9 out of 13) while only a few were present in *T. mixta* (3

out of 13). Both genomes showed intergenic spacers and overlaps. Concerning spacers, the longest was observed in *K. klovstadi* mtDNA (592bp) between *trnS* and *nad1*, while the longest of *T. mixta* was 155 bp between *rrnL* and *trnV*. Overlaps were, on the contrary, few and restricted to the region around *atp6* and *atp8* in both genomes, plus a 33 bp overlap between *rrnL* and *trnV* in *K. klovstadi*. Gene orders of both species were consistent with the Pancrustacea model (apart from *T. mixta,* which lacks *trnC*). Overall, a bias in nucleotide composition towards A (36.2% and 34.0%) and T (31.6% and 29.1%) was observed with respect to C (18.8% and 24.9%) and G (13.4% and 12.0%) in *K. klovstadi* and *T. mixta* mitogenomes, respectively.

### 3.3.2. Dataset Composition and gene orders

The EZsplit program, used to automatically parse the Collembola and outgroup mitogenomes downloaded from NCBI, took only 1 sec. Mitogenomes without a proper annotation were processed as described in section 2.4.1. Due to the lack of some genes and the presence of long stretches of Ns (undetermined nucleotides), few genomes were identified as incomplete.

A total of 87 species were finally selected and their gene order analysed except for *Lepidocyrtus* sp. (MF716621) which showed an uncertain mitogenome structure and was hence excluded. Most of gene orders herein studied had been already characterized and described (Carapelli et al., 2014; Leo et al., 2019; Dong et al., 2020; Sun et al., 2020). As already described in section 3.2.1, *K. klovstadi* and *T. mixta* display a gene order in line with the presumed ancestral state (Pancrustacea model), apart the missing *trnC* in the latter. Most genomes from Godeiro et al. (2020) display a similar Pancrustacea gene order or a compatible arrangement if incomplete. Some species, i.e. *Seira* ca. *prodiga* 2*, S. downgli, S. tinguara, S. ritae, Entomobrya* sp*. and Tyrannoseira bielanensis* lacked of one or two genes. Pending the need of a confirmation of this gene losses, these were not identified as novel gene orders and were tentatively considered as conforming to the Pancrustacean gene arrangement. A distinct and new gene order was instead observed in *Seira parabiensis*, which showed an *rrnS* deletion and an *rrnL* translocation. Moreover, two species showed a gene order modification in the *trnA-trnR-trnN-trnS-trnE* region: *Trogolaphysa* sp. (MF716607) displayed an inversion and deletion resulting in a *trnE-trnN-trnA-trnR* configuration, whereas *Seira* sp. 3 (MF716612) showed a translocation resulting in a *trnN-trnS-trnR-trnE-trnA* pattern (Fig. 20).
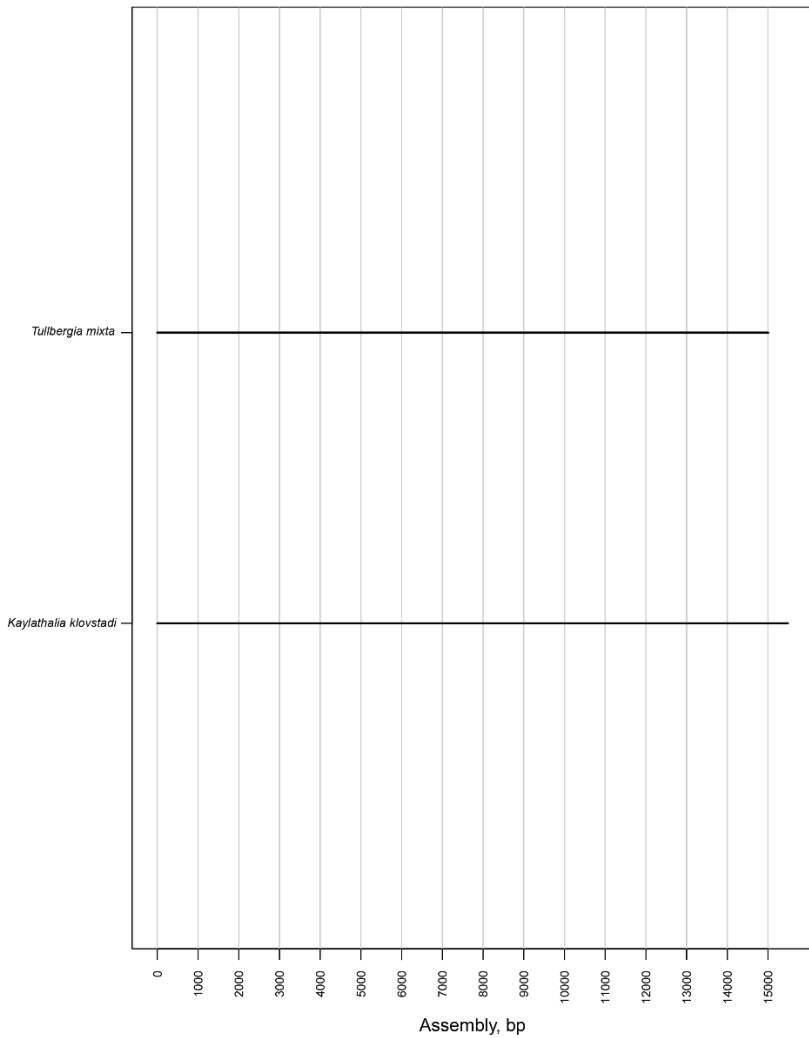
Figure 19. EZmix output obtained from the *T. mixta* and *K. klovstadi* mitogenomes comparison. As the program output shows, no chimeric region was detected between the two sequences.
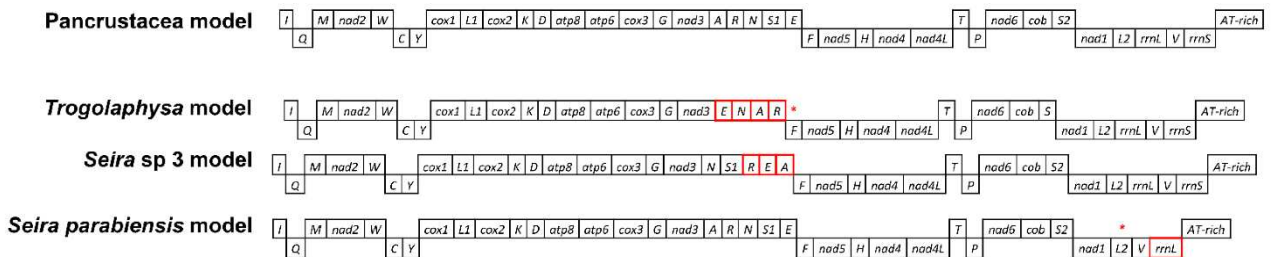
Figure 20. New gene order retrieved from Godeiro et al. (2020) compared to the ancestral Pancrustacean model. All derived gene arrangements, with respect to the basal model, show translocations as red boxes and deletions as asterisks (*).

3.3.3. Phylogenetic analysis and tree topology comparison

Complete or semi-complete mitogenomes of Collembola and outgroups, previously revised manually (see section 2.4.2), were formatted and submitted to EZpipe according to the web service instructions. The output was completed in 101 sec and was used as input for the subsequent phylogenetics analysis. The resulting tree supports the monophyly of Collembola as well as of the orders Symphypleona and Poduromorpha (Fig. 21). Entomobryomorpha and Neelipleona resulted as polyphyletic due to the presence of some outliers: *Neelides* sp., *Novacerus tasmanicus*, *Oncopodura yosiiana* and *Tomocerus qinae*. Nevertheless, with the exception of these four outliers, the majority of Neelipleona and Entomobryomorpha formed two well-supported and clearly identifiable monophyletic clusters. This outcome, at odds with the current taxonomic knowledge of the groups, could have arisen because of (i) a true polyphyly or (ii) as a result of bias in the phylogenetic analysis. However, due to substantial annotation errors found during the previous steps in some genomes taken from the literature, it was not possible to exclude technical problems in the original primary sequence. Indeed, in previous works where such enigmatic taxa were employed (Godeiro et al., 2020; Sun et al., 2020), researchers did not include members belonging to the same order, making it impossible to evaluate the possible relationships of these outlier species. In detail, in the phylogenetic reconstruction deepened in this thesis, *Neelides* sp. appeared as extremely divergent with respect to all other collembolans, being retrieved as the outermost basal taxon distant from the three Neelipleona representatives that correctly form a monophyletic cluster in the Collembola tree. On the other hand, despite the fact that *Oncopodura yosiiana, Novacerus tasmanicus* and *Tomocerus qine* belong to two distinct subfamilies in the same superfamily Tomoceroidea, each clustered with different and unrelated taxa (Fig. 21).

Focusing on the order level, and apart from outliers, Symphypleona appeared as monophyletic and basal with respect to the remaining lineages, which were grouped in a common cluster. This latter was subdivided into two branches: one including Poduromorpha *plus* Neelipleona and the second including all the Entomobryomorpha.

Within the Symplypleona group a polytomy was observed basally. Nevertheless, all the families represented by two or more species (Sminthuridae and Dicyrtomidae) were monophyletic, disclosing that the clade Katiannidae + Sminthurididae was sister group to Sminthuridae. Entomobryomorpha, apart from the yet mentioned outliers, forms a coherent taxon. Isotomidae was monophyletic, while Entomobryidae and Paronellidae were polyphyletic. All Isotomidae genera, with the exception of *Sinella*, were monophyletic. The

whole Poduromorpha order was monophyletic as well and divided in two major lineages: one containing Hypogastruridae, Poduridae and Neanuridae, and the second including Onychiuridae *plus* Tullbergiidae. All the families were monophyletic except for Hypogastruridae, due to the displacement of *Gomphiocephalus hodgsoni*. Though its paraphyly, Hypogastruridae appeared ancestral with respect to the clade formed by Poduridae + Neanuridae.

In general, full support was shown at intermediate and recent nodes with minor exceptions, whereas nodes at the level of order and family showed variable-to-high support (0.84 < p.p. < 1).

In order to evaluate possible alternative topologies, monophyly tests were conducted on ambiguous groups (i.e., taxa recovered as poly- or paraphyletic in the previous analysis). In this evaluation, competing hypothesis could not be rejected in all cases with full confidence. All non-monophyletic groups (Neelipleona, Tomoceridae, Hypogastruridae, Paronellidae and Entomobryidae) were therefore considered as topologically plausible, although suboptimal given our unconstrained analysis (Tab. 6)

3.3.4. Nucleotide Biases and dN/dS Ratio

EZskew analysis, meant to study the nucleotide composition across mitogenomes, was completed in only 15 sec and the resulting matrices were employed to build *ad-hoc* graphics. The AT% bias observed in Collembola was compared with data from all available hexapod mtDNAs. Although some internal variation was observed, Collembola were characterized by a relatively low AT-bias (60.4–74.8%) with respect to other Hexapoda (59.4–88%; Fig. 22).

Collembola nucleotide biases (AT skew and CG skew) were calculated as in section 2.4.3. Variability in skew was sizeable in in four-fold and two-fold degenerate third codon positions and limited in first and second codon positions, following the order 4JN3 > 2JN3 > JN2 > JN1 (Fig. 23). J1 genes were characterized by a weak negative or barely positive CG skew (−0.21 to 0.06), and mainly positive AT skew (−0.04 to 0.15).
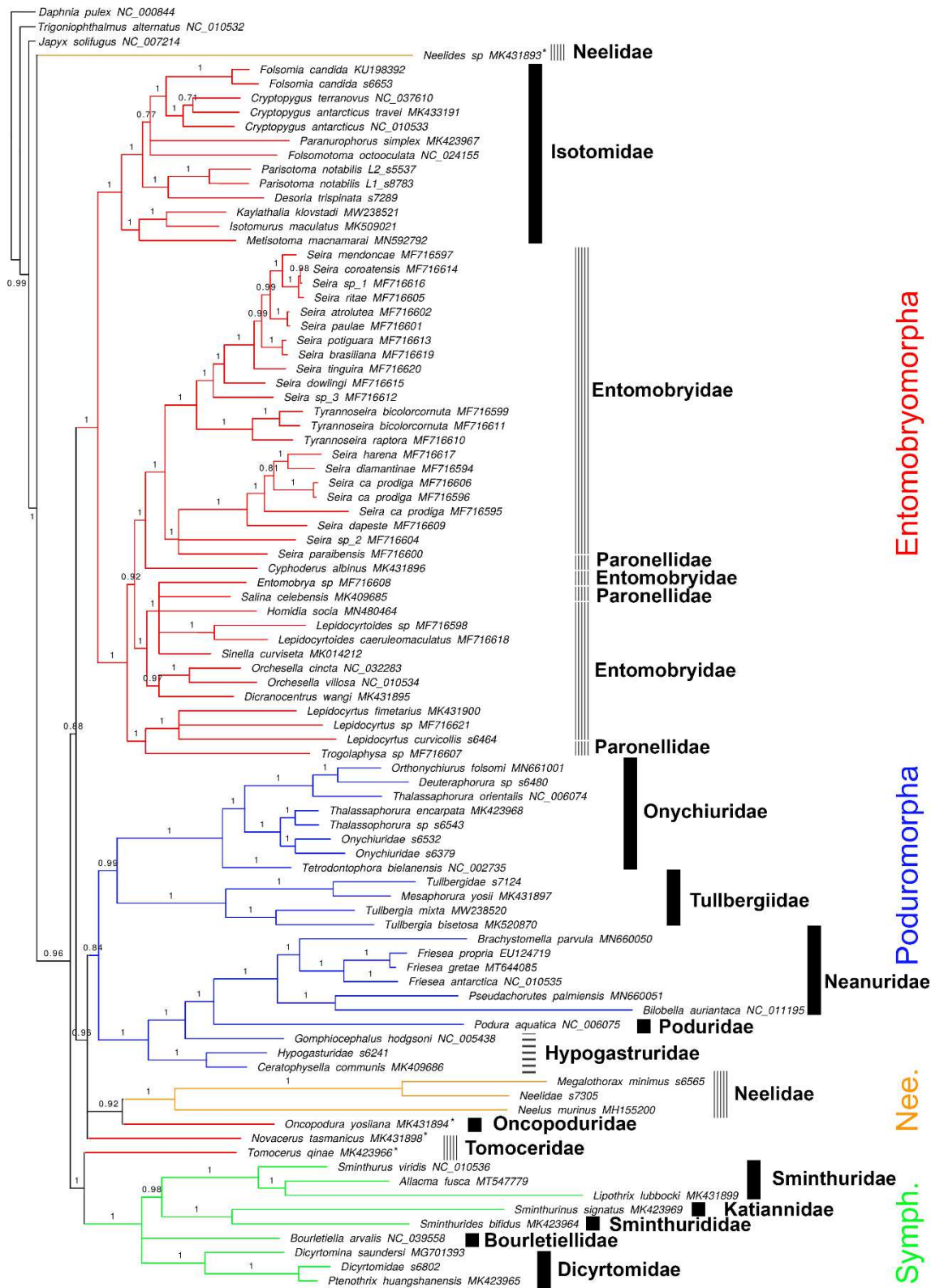
Figure 21. Bayesian phylogenetic reconstruction based on first and second codon positions. Numbers at nodes indicate posterior probabilities. Orders are colour coded. Monophyletic, paraphyletic, and polyphyletic families are indicated with filled, horizontally hatched and vertically hatched lines. Outlier species are highlighted with asterisks. Nodes with support < 0.70 were collapsed. Outgroup nodes not to scale.

Tree	logL	deltaL	bp-RELL	p-KH	p-SH	p-WKH	p-WSH	c-ELW
Unconstrained	-242.123.285	16.015	0.209 +	0.324 +	0.596 +	0.324 +	0.608 +	0.209 +
Neelipleona	-242.139.976	32.706	0.055 +	0.140 +	0.365 +	0.140 +	0.326 +	0.056 +
Tomoceridae	-242.135.204	27.934	0.073 +	0.194 +	0.426 +	0.194 +	0.431 +	0.073 +
Hypogastruridae	-242.107.270	0.000	0.356 +	0.591 +	1000 +	0.591 +	0.922 +	0.355 +
Paronellidae	-242.127.413	20.143	0.106 +	0.232 +	0.579 +	0.232 +	0.557 +	0.106 +
Entomobryidae	-242.110.257	2.987	0.197 +	0.408 +	0.881 +	0.408 +	0.852 +	0.197 +

Table 6. Tree topology tests across non monophyletic groups of Collembola. Tree column indicates the group under testing and logL the logarithm of the likelihood of the best tree complying to the condtraint. Plus (+) signs indicate for which test the monophyly was accepted. deltaL: logL difference from the maximal logL in the set; bp-RELL: bootstrap proportion using RELL method (Kishino et al., 1990); p-KH: p-value of one sided Kishino-Hasegawa test (1989); p-SH: p-value of Shimodaira-Hasegawa test (1999); p-WKH: p-value of weighted KH test. p-WSH: p-value of weighted SH test. c-ELW: Expected Likelihood Weight (Strimmer & Rambaut, 2002).
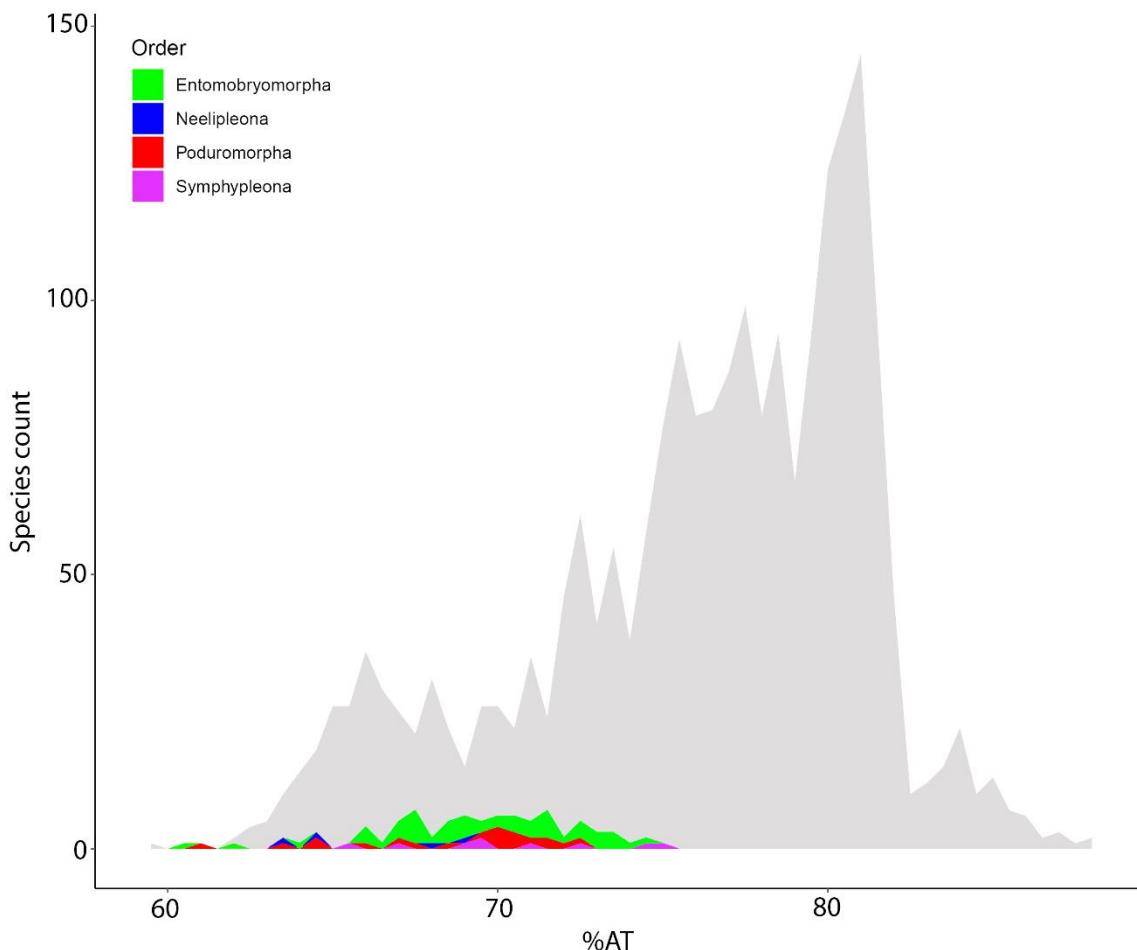

Figure 22. AT% compositional bias within Hexapoda mitogenomes. Collembola genomes are color-coded against a grey Hexapoda dataset.

N1 genes, on the other hand, showed substantial negative CG skew (−0.55 to −0.08) and AT skew (−0.32 to −0.06). *Bourletiella arvalis* appeared as an outlier for the N1 gene bias. Focusing on second codon positions, CG skew was mainly positive in J2 genes (0.19 to 0.35) but not in N2 positions (−0.21 to 0.19). AT skew at second codon positions was negative for both strands with similar ranges (J2: −0.42 to −0.31; N2: −0.56 to −0.22). Two Neelipleona (*Neelides* sp. and *Neelus murinus*) and the Symphypleona *Sminthurinus signatus* were identified as outliers. Two-fold third codon position biases displayed a higher variability across species. Albeit AT skew for 2J3 and 2N3 was comparable (−0.40 to 0.17), CG skew was particularly different, ranging from 0.03 to 0.86 in 2J3 and −0.89 to 0.2 in 2N3. Finally, 4N3 and 4J3 showed the highest variance, with CG skews ranging from 0.85 to −0.73 and AT skews from 0.32 to −0.6. Three outliers (*B. arvalis, Thalassophorura encarpata* and *Thalassophorura* sp.) were observed.

In order to detect the degree of purifying and/or directional selection across springtail mitogenomes, the balance of nonsynonymous substitutions was tested trough the dN/dS ratio. The comparative models tested (M0 vs M3, M7 vs M8 and M8a vs M8) led to a significant LRT p-value with a resulting neutral selection, indicating that a balance for all the examined nucleotides was detected.
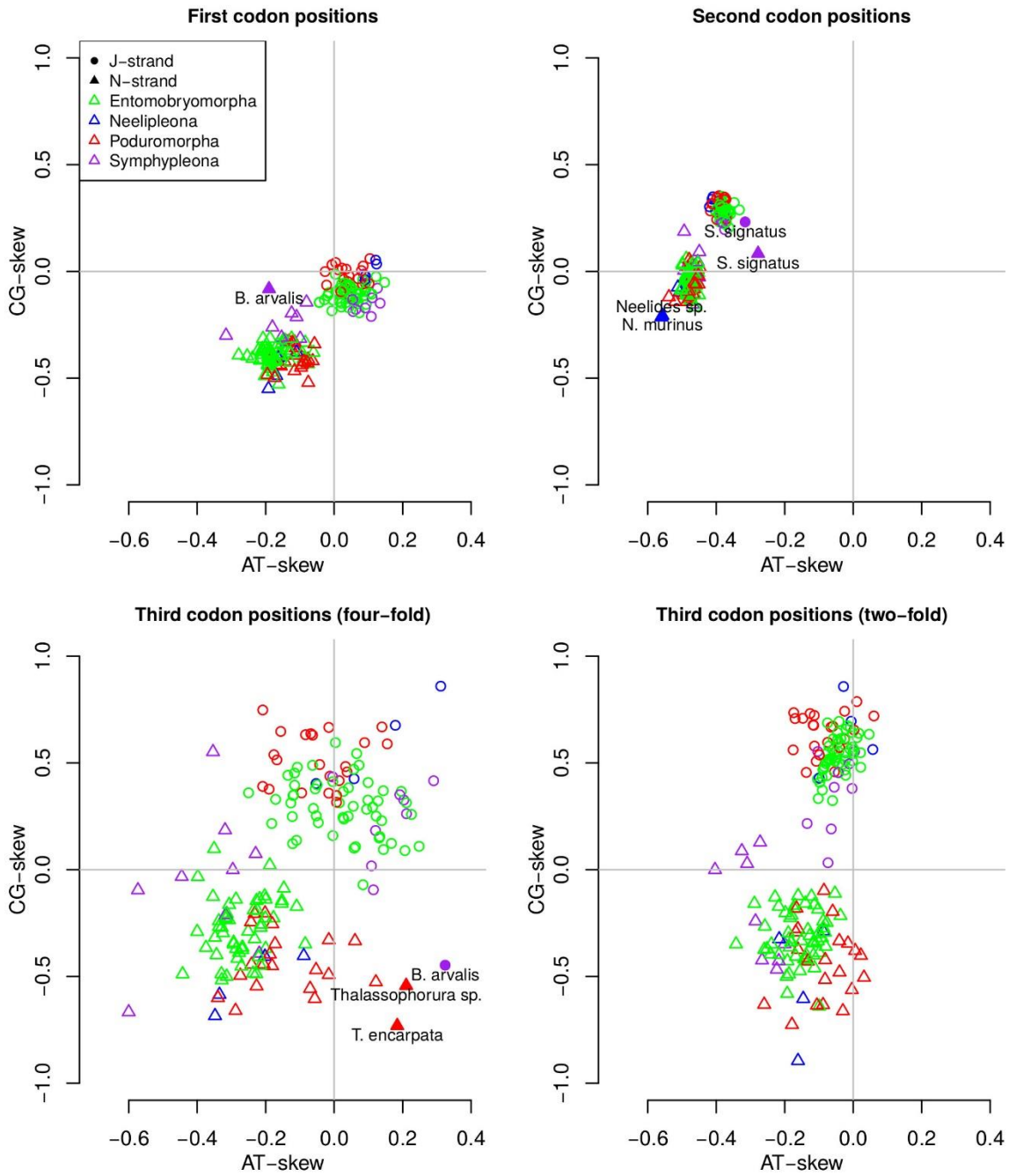
Figure 23. Base skews for all available Collembola mitogenomes. Orders are color-coded and J- and N-strands are coded by shape (see legend). Outliers are identified by species name.

4. Discussion

From the dawn of the industrial age, human activities have led to a substantial increase in environmental pollution, accelerating climatic variations which are now jeopardizing species viability and biodiversity equilibria. In this respect, my thesis work focused on two main issues (i) understanding how a controlled heat exposure could affect *C. terranovus'* transcriptomic response and (ii) evaluating Collembola systematics employing a phylomitogenomic approach trough the addition of two Antarctic species: *K. klovastadi* and *T. mixta*. Both goals were carried out using NGS technologies, powerful multi-disciplinary tools that can be employed to better understand animal physiology as well as explore species biodiversity.

To date, multiple researchers applied RNA-seq to investigate how Antarctic species react to thermal stresses, although invertebrates have been unduly excluded from the list of focal species. Only few ecological studies focused on how Antarctic arthropods' species (the springtail *C. antarcticus* and the mite *A. antarcticus*) adapt to higher temperatures, with nevertheless only a secondary focus on molecular metabolic and physiological modifications (Everatt et al., 2013; Morley et al., 2019; Bahrndorff et al., 2021). In this context, aim of this study was to investigate the effect of thermal stress on *C. terranovus* exposed to mid-term high temperatures, for the first time using a RNA-seq approach. The average temperature to which specimens of *C. terranovus* sampled in the wild (employed as control in the experiment) were exposed under natural conditions was 5.4°C. This figure, measured at the time of the experiment, resulted in line, with minimal variations, to temperatures registered in the subsequent summer (see section 3.4; Fig. 15), confirming that this is the regular temperature experienced by these animals in this period of the year under natural conditions and, reasonably, the temperature to which these animals are acclimated on the long term. The thermal exposure applied during the experiment was +12.4°C with respect to temperatures in the wild. Based on available, and extensive, year-round temperature recordings in the microhabitats in the following year, this difference can be evaluated ex post.

Nevertheless, the thermal exposure applied during the experiment, appears of limited to mid extent compared to a year-round temperature variation of ±45°C from the coldest to the warmest season (Fig. 15). At the same time, an average temperature (exposed samples) of 18°C is expected to be sufficient viable in the mid period, give the results of Everatt et al. (2013). In the latter experiment, authors opted for a mild testing regime (10°C) for a long-period using the congeneric *C. antarcticus* as model species, noticing that surviving strongly

began to decrease after two weeks. The experimental exposure exceeds the maximum temperature that phylogenetically closer animals experienced in Everatt et al. (2013) work (~8°C), as well as the expected temperature in the wild (~10°C). As such, it can be re-evaluated *a posteriori* as a biologically meaningful stressing regime.

Given the absence of a reference genome for the species, and being *C. terranovus* a non-model animal species, its transcriptome was characterized ex-novo. In line with the paucity of genomic tools for the species under scrutiny, and for Collembola at large, as well as the scarcity of available material, transcriptome sequencing produced an assembly on the mid-low range in terms of the most commonly employed quality indicators, characterized by a fair number of partial and fragmented transcripts. However, the annotation rate was remarkable for a non-model species, with ~72% of annotated genes.

The differential expression analysis showed subtle dissimilarities between the two conditions with fairly high levels of inter-individual variability. This outcome might be related to a high plasticity in gene expression of this taxon, as also reported in other polar species (Bahrndorff et al., 2021). As an example, Everatt et al. (2013) found that some subgroups of *C. antarcticus* managed to endure long periods of thermal stress better than others, despite a continuous decrease in the survival rate. If the gene differential expression analysis showed limited differences across conditions, associated with a high intra condition variability, the GO terms distribution disclosed interesting results. These latter were better investigated with a GO enrichment analysis where significant differences between the two groups were assessed. These were primarily related to (i) carbohydrate, (ii) protein, (iii) lipid metabolism and (iv) spermatid development (Fig. 14, Tab. 7).

Class	GO	Up-regulation
Carbohydrate metabolism	GO:0004034	CT
	GO:0004553	
	GO:0046373	
	GO:004656	
Protein catabolism	GO:0004021	HE
	GO:0004180	
	GO:0004181	
	GO:0042853	
	GO:0070573	
Lipid metabolism	GO:0004312	HE
	GO:0004313	
	GO:0004314	
	GO:0016295	
	GO:0016296	
	GO:0016788	
	GO:0031177	
	GO:00471177	
Spermatid development	GO:0007286	HE

Table 7. Gene expression enrichment analysis of GO terms in relation to the *C. terranovus* group up-regulation.

In detail, GO terms related to the sugar catabolic processes appeared underrepresented in HE specimens. These transcripts were identified as part of digestion pathways, perhaps indicating that suboptimal conditions, related to the thermal treatment, repress the glycolytic activity that is, on the other hand, activated by a better nutritional status at physiological temperatures in the wild environment. Noteworthy, carbohydrate catabolism pathways can be also activated by Antarctic arthropods to face extreme cold temperatures (Teets & Denlinger, 2014). For example, *C. antarcticus* is known to produces trehalose, glucose and glycerol as cryoprotectants (Elnitsky et al., 2008). Albeit *C. terranovus* apparently expressed different carbohydrate catabolism pathways with respect to its congeneric, it is nevertheless possible to speculate that this could be a signature of a metabolic activity activated in response to low temperatures, and hence repressed in thermally stressed animals.

On the other hand, the HE group was characterized by an increased activation of protein degradation pathways, with the expression of several different peptidases. Overall, thermal exposure can cause protein misfolding and damage, that in turn can trigger protein turnover

or rearrangement trough the activation of heat shock proteins (HSPs). In the Antarctic environment, these chaperones seem to be constitutive expressed in many species, allowing for a year-round defence against frequent temperature shifts (Rinehart et al., 2006; Teets & Denlinger, 2014). Considering that *C. terranovus* specimens undergoing thermal stress seem to activate protein catabolism in the absence of the typical upregulation of HSP, it can be hypothesized that *C. terranovus* expresses HSPs constitutively, with no evidence of a direct response to variable environmental temperature (Rinehart et al., 2006).

Furthermore, a general upregulation of lipid activity was found in HE samples. Previous research indicated that arthropods subjected to dehydration display a downregulation of fatty acid metabolism (Timmermans et al., 2009; Teets et al., 2012), similar to what it can be observed in *C. terranovus* specimens of the CT group. These latter were sampled in the wild without a specific monitoring of humidity, and it was not possible to assess if this downregulation could be interpreted as a consequence of dehydration stress. The current observations, nevertheless, find a stark parallel with what observed in the Antarctic benthonic fish *Harpagifer antarcticus* following thermal exposure, where it was possible to observe an increase in lipid activity associated with the rearrangement of cellular membranes (Thorne et al., 2010). In a similar way, heated samples of *C. terranovus* showed an activation of the same metabolic pathways which might be shared, even if phylogenetically distant, with the Antarctic fish.

Finally, HE individuals also displayed an enrichment of the GO term related to spermatid development. This might suggest that the environmental conditions experimented by treated samples fostered spermatid production. This outcome may be related to the breeding conditions (inside a box container), where the physical proximity of individuals may have triggered chemical signal production promoting sexual behaviour and spermatid production. Another possible hypothesis could be related to the higher temperature that, mimicking or exceeding conditions in the good season, could have triggered the mating behaviour of these individuals, that is restricted to the good season.

In order to get a better insight on the actual temperatures experienced by Antarctic springtails under physiological condition, and to make reasonable hypotheses on the temperatures that might be experienced in critical heating conditions in the wild, it is mandatory to collect temperature data from the very microhabitat where these species thrive, at variance with standard meteorological records focussing on air temperatures at large.

Nevertheless, studies on microhabitat temperatures, so far, mostly focused on the maritime Antarctic, whereas the Continental zone received lesser attention (Davey et al., 1992; Moorhead et al., 2002; Quayle et al., 2002, 2003; Bokhorst et al., 2011; Convey et al., 2018; Obryk et al., 2020). For this reason, and in order to characterize temperatures in microhabitats inhabited by *C. terranovus* and other south polar invertebrates, a year-round record of temperatures was collected in terrestrial and shallow water microenvironments in Victoria Land and these values were compared with open-air temperatures. Overall, macroclimatic data from this region indicated that seasonal trends are in line with the ones proposed by Obryk et al. (2020), namely summer (November-February), autumn (March), winter (April-September) and spring (October) (Fig. 15). In agreement with previous works (Quayle et al., 2003; Convey et al., 2018), microclimatic data showed that microhabitat temperatures, though coupled to a large extent on a seasonal scale, differ considerably from air temperatures, and significant differences are also observed between pond and soil sites. Based on mean temperatures and daily excursions (Fig. 15) this latter outcome was particularly evident in summer, coinciding with a peak in the biological activity of these animals (Bokhorst et al., 2011). Similarly, the duration of the zero curtain period as well as freeze-thaw cycles indicated that warmer and more stable conditions were experienced in both microhabitats with respect to the open-air environment (Tab. 5).

Ponds were constantly warmer than soils, because of the higher heat capacity of water (Convey et al., 2018). This phenomenon can be appreciated in mean daily temperatures and in freeze-thaw cycles (Fig. 15, Tab. 4). Indeed, soil and pond temperatures increased in parallel at the beginning of the summer but collapsed more rapidly at the end of this period in soil sites, leading to a narrower thawing period in soils. Moreover, the cumulative degree days curve confirmed that freshwater environments were characterized by a larger heat accumulation effect compared to soils and, to a large extent, to the open-air (Fig. 18). In this respect, microhabitats appear to be protected and buffered from the stark temperature minima observed in the surface environment, especially during the good season. In full evidence, this may have a major outcome in terms of the length of the viable period for biological activity during summer, one of the major bottlenecks for the survival and development of polar species.

Along these lines, further studies may be warranted to (i) collect multi-year data records across biologically significant microhabitats to better define temperature patterns and variation experienced by these species over time and (ii) assess abiotic trends in other geographic sites where distinct Antarctic microfauna's species can be sampled.

Although, in general terms, Antarctica is a low-biodiversity region, a recent account listed over 1,142 species, with Collembola playing a key role in the soil ecology of this region (Leo et al., 2021; Phillips et al., 2022). In order to improve our knowledge on Collembola biodiversity in this area, and to complement the wealth of morphological data collected in the last century with molecular data, the complete mitochondrial genome sequence was obtained and characterized from two additional species of Antarctic springtails.

In this context, in order to standardize and facilitate data processing, an analytical pipeline was created and implemented as a freely available web server (EZmito: http://ezmito.unisi.it/) which can be used for mitochondrial genome data processing, visualization and analysis (Cucini et al., 2021a). The service is aimed at the general public of non-bioinformaticians and is characterized by a simple interface focussing on the analyses that are most commonly seen in the specialized literature. From February 2022, EZmito has accounted for more than 850 views from all continents (except Antarctica; Fig. 24). Following the document citations and users' feedbacks, EZcodon, a tool used to visualize codon usage in complete mitochondrial genomes, appears as the most used service within EZmito, followed by EZsplit, used for parsing of NCBI complete genome records with coding information, and EZpipe, capable of reformatting complete genome datasets in single gene datasets (Tab. 8).
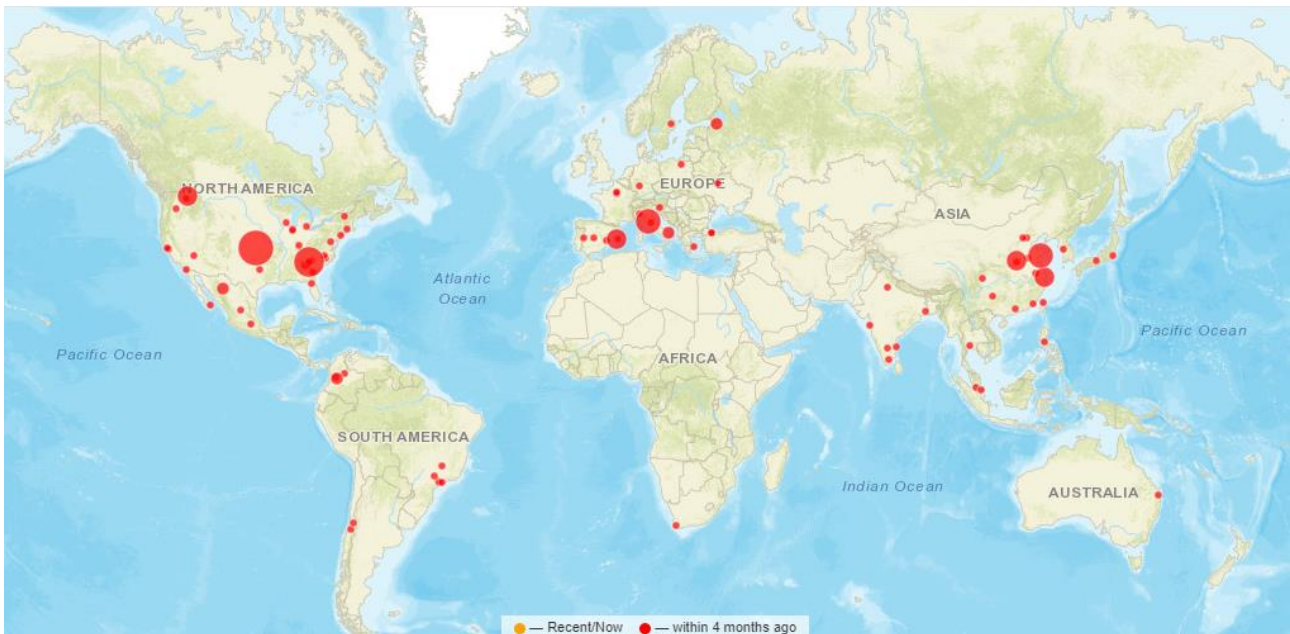
Figure 24. World-wide EZmito utilization from February 2022.

EZmito tool	Users
EZsplit	114
EZpipe	92
EZskew	82
EZcodon	577
EZmix	83

Table 8. EZmito tools' users at June 2022.

The EZmito web server was employed to study the mitochondrial features of the two new Antarctic genomes and to analyse these in the context of a wider dataset of 87 complete or semi complete mitochondrial genomes available from Collembola. Overall, the phylogenetic reconstruction recognized the four generally recognized orders (Symphypleona, Neelipleona, Entomobryomorpha and Poduromorpha) as monophyletic, with minor incongruences (such as the polyphyly of some taxa, see below). These latter incongruences were investigated in detail and resulted as not strongly supported, leading to the overall acceptance of relationships among major groups as described in Fig. 21.

Multiple studies have addressed the internal phylogeny of Collembola, focussing mostly on the monophyly of the four orders and on order level relationships, although with smaller datasets and/or a more limited taxon sampling, and different hypotheses were made for the phylogeny of Collembola (Fig. 3). For example, in Leo et al. (2019), Neelipleona emerged as the most basal taxon and Symphypleona appeared as the sister group of Entomobryomorpha. The basal position of Neelipleona found support also in Gao et al. (2008) based on a set of nuclear markers. The scenario supported by the present analysis, on the other hand, is generally concordant with the results of Sun et al. (2020), that similarly recovered Symphypleona as the basalmost taxon in the Collembola tree.

The polyphyly of Entomobryomorpha and the paraphyly of Hypogastruridae have been already discussed in other works, casting doubts on the real status of these taxa (Xiong et al., 2008; Leo et al., 2019; Sun et al., 2020). Furthermore, problematic families, such as Paronellidae and Entomobryidae, were observed to be polyphyletic in accordance with previous research (Sun et al., 2020).

The basal placement of Symphypleona, the clade composed of Poduromorpha *plus* Neelipleona, the close relationship between Entomobryidae and Isotomidae and polyphyly of the Entomobryomorpha caused by the displacement of Tomoceridae and Oncopoduridae, were also supported by a combination of nuclear markers (*18S* and *28S*) in a medium-large recent dataset (Yu et al., 2016).

Within the 87 analysed species, 10 were endemic of Antarctica and all these members belong to the Poduromorpha or Entomobryomorpha. *Cryptopygus* species (*C. antarcticus, C. a. travei, C. terranovus*), *Folsomotoma octooculata,* and the new *K. klovstadi* were placed in the Isotomidae clade whereas the rest were retrieved in three distinct families: Neanuridae (*Friesea propria, F. gretae* and *F. antarctica*), Tullbergidae (the newly characterised *T. mixta*) and Hypogastruridae (*Gomphiocephalus hodgsoni*). However, no specific correspondence between habitat and lineage was found, indicating that order and family radiations happened before the Gondwanaland separation and well before the Antarctica isolation (Leo et al., 2019). The emerging view is therefore that these species endured locally for a prolonged period, facing the cooling of the continent, in local refugia as happened during the Last Glacial Maximum (Carapelli et al., 2017).

Gene orders, i.e. the order and orientation of genes along the molecule, are an additional character of primary phylogenetic interest, as they are generally lineage-specific, little affected by convergence, and as such can provide solid synapomorphic traits within given taxonomic groups. Overall, the Pancrustacea model was confirmed as the ancestral gene order for the group, as well as the most common. Second most common is the *Tetrodontophora* model, that is observed in all members of the Onychiuridae family studied thus far. This gene order model can be hypothesized to be a true synapomorphy of the family and provides a substantial support for the monophyly of this latter. Among the newly analysed genomes, an autapomorphy characterised the taxon *Pseudachorutes*, with *P. palminensis* as the only representative species (Fig. 4). Species belonging to Entomobryomorpha showed the most uniform in terms of genome arrangement, with 49 out of 51 species showing the Pancrustacea model. Only in three species, reanalysed in terms of gene order based on primary data from Godeiro et al. (2020), new gene orders were observed: *Trogolaphysa*, *Seira* sp. *3* and *Seira parabiensis* (Fig. 20). While their status as new genome arrangements are obviously plausible, some caution is nevertheless required, as congeneric species show the classical Pancrustacea gene order. At variance, a noticeable "hotspot" of genomic rearrangements was observed in Symphypleona, where 5 out of 9 species are characterized by a different gene order compared to the presumed ancestral state (see Nardi et al., 2020 for a detailed account).

The newly determined genomes displayed two different genome arrangements: *K. klovstadi* was characterised by a classical Pancrustacea model whereas *T. mixta* appeared to lack the *trnC* gene, at variance with its congeneric *T. bisetosa* which showed the complete Pancrustacea gene order. The absence of a *tRNA* is not an overly unusual feature, as it is

shared with other species such as *Podura aquatica* and the *Trogolaphysa* model, that are missing *trnY* and *trnS*1, respectively. Nevertheless, the possibility must be considered that this may also be an outcome of an annotation bias if *trnC* of *T. mixta* had evolved a largely different conformation compared to the typical cloverleaf structure of tRNAs used as a model for tRNA gene identification (see section 2.4.1).

Considering all the gene orders analysed for Collembola, it is evident that modifications were not randomly distributed along the molecule. In particular, two mitogenomic regions appears as "hotspots" for gene order changes: the *A+T rich-nad2* region, where most translocations were observed in Symphypleona, and the *trnA-nad1* region.

The availability of the largest dataset of complete or semi complete mitochondrial genomes in Collembola thus far allowed for a revision of some genome wide trends previously discussed in the literature based on smaller datasets, including AT bias and codon position biases. Since the mitochondrial genome replication is an asynchronous and asymmetrical process, where one hemi-helix (N) remains in a single strand conformation for a longer period respect to the other, Reactive Oxygen Species (ROSs) induced mutations are uneven between strands (Hassanin et al., 2005). Such mutational biases were already studied in Collembola (Carapelli et al., 2014; Carapelli et al., 2019). The A+T bias was confirmed in Collembola, although to a lesser extent than Hexapoda at large, where some highly derived taxa (e.g. Hymenoptera) display exceptionally high levels of A+T biase (Fig. 22). Concerning codon position biases, different levels of dispersion were assessed: first and second codon positions were relatively more uniform with respect to two- and four-fold degenerate third codon positions. This is in line with the expectations bases on the degeneration of the genetic code, as third codon position in degenerate sites are more liberal in terms of the possible mutations to be retained because mutations rarely affect protein sequences and are not wiped off by purifying selection. Outliers were observed at all levels (except for third codon two-fold site) that may indicate a different selection pressure in some individual species, without a clear taxonomic or ecological pattern (Carapelli et al., 2019; Fig. 23).

5. Conclusions

This thesis draws from multiple different areas of research, from transcriptomics to phylogenetics, through bioinformatics and basic climatology, with the aim of adding one additional piece to the puzzle of Antarctic biodiversity and our knowledge about how Antarctic species may be affected in a context of global warming.

Victoria Land microhabitat temperatures were observed to diverge strongly with respect to air temperatures as monitored by the AWS system, and the abiotic conditions to which micro-invertebrates are exposed were described in some detail. The response of *C. terranovus* to an experimentally induced heat stress was studied in terms of gene expression, supporting a high degree of transcriptional plasticity in this species and identifying some pathways that appear to be involved in the response. The characterization of two new mitochondrial genomes from Antarctic species increased the representativity of these data with respect to Antarctic biodiversity, including one new family. Finally, our basic standard workflow for the analysis of complete mitochondrial genomes was implemented as a feely available web server to foster the application of these methods through an easy interface.

With this thesis I wish to underline the importance of a timely effort by the scientific community, at large, to study Antarctic diversity and ecology as a prerequisite for a focussed reaction to the challenges that are being posed by global warming as far as the Great South.

Acknowledgements

References

Bahrndorff S., Lauritzen J.M.S., Sørensen M.H., Noer N.K. & Kristensen T.N. 2021. Responses of terrestrial polar arthropods to high and increasing temperatures. J. Exp. Biol. **224**: jeb230797.

Bergstrom, D.M., Convey, P., & Huiskes, A.H.L. 2006. Trends in Antarctic terrestrial and limnetic ecosystems: Antarctica as a global indicator. Springer, Dordrecht. 369 pp.

Bernt M., Braband A., Schierwater B. & Stadler P.F. 2013a. Genetic aspects of mitochondrial genome evolution. Mol. Phylogenet. Evol. **69**: 328–338.

Bernt M., Donath A., Jühling F., Externbrink F., Florentz C., Fritzsch G., Pütz J., Middendorf M. & Stadler P.F. 2013b. MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. **69**: 313–319.

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 Evol. Biol. **18**: 143.

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 Biol. Biochem. **40**: 1547–1556.

Bokhorst S., Huiskes A., Convey P., Sinclair B.J., Lebouvier M., Van de Vijver B. & Wall D.H. 2011. Microclimate impacts of passive warming methods in Antarctica: implications for climate change studies. Polar Biol. **34**: 1421–1435.

Boore J.L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. **27**: 1767–1780.

Boore J.L. & Brown W.M. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. Curr. Opin. Genet. Dev. **8**: 668–674.

Brunetti C., Siepel H., Fanciulli P.P., Nardi F., Convey P. & Carapelli A. 2021. Two New Species of the Mite Genus *Stereotydeus* Berlese, 1901 (Prostigmata: Penthalodidae) from Victoria Land, and a Key for Identification of Antarctic and Sub-Antarctic Species. Taxonomy. **1**: 116–141.

Bushmanova E., Antipov D., Lapidus A. & Prjibelski A.D. 2019. rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. GigaScience. **8**: giz100.

Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K. & Madden T.L. 2009. BLAST+: architecture and applications. BMC Bioinformatics. **10**: 421.

Carapelli A., Convey P., Nardi F. & Frati F. 2014. The mitochondrial genome of the antarctic springtail *Folsomotoma octooculata* (Hexapoda; Collembola), and an update on the phylogeny of collembolan lineages based on mitogenomic data. Entomologia. .

Carapelli A., Fanciulli 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 Biol. **42**: 715–732.

Carapelli A., Leo C. & Frati F. 2017. High levels of genetic structuring in the Antarctic springtail *Cryptopygus terranovus*. Antarct. Sci. **29**: 311–323.

Carapelli A., Liò P., Nardi F., van der Wath E. & Frati F. 2007. Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. BMC Evol. Biol. **7**: S8.

Castresana J. 2000. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol. Biol. Evol. **17**: 540–552.

Cesari M., McInnes S.J., Bertolani R., Rebecchi L. & Guidetti R. 2016. Genetic diversity and biogeography of the south polar water bear *Acutuncus antarcticus* (Eutardigrada : Hypsibiidae) – evidence that it is a truly pan-Antarctic species. Invertebr. Syst. **30**: 635.

Cicconardi F., Borges P.A.V., Strasberg D., Oromí P., López H., Pérez-Delgado A.J., Casquet J., Caujapé-Castells J., Fernández-Palacios J.M., Thébaud C. & Emerson B.C. 2017. MtDNA metagenomics reveals large-scale invasion of belowground arthropod communities by introduced species. Mol. Ecol. **26**: 3104–3115.

Clarke A. & Beaumont J.C. 2020. An extreme marine environment: a 14-month record of temperature in a polar tidepool. Polar Biol. **43**: 2021–2030.

Clem K.R., Fogt R.L., Turner J., Lintner B.R., Marshall G.J., Miller J.R. & Renwick J.A. 2020. Record warming at the South Pole during the past three decades. Nat. Clim. Change. **10**: 762–770.

Condamine F.L., Guinot G., Benton M.J. & Currie P.J. 2021. Dinosaur biodiversity declined well before the asteroid impact, influenced by ecological and environmental pressures. Nat. Commun. **12**: 3833.

Convey P. 2011. Antarctic terrestrial biodiversity in a changing world. Polar Biol. **34**: 1629–1641.

Convey P. 2013. Antarctic Ecosystems. pp. 179–188. Encyclopedia of Biodiversity, Elsevier,

Convey P. 2017. Antarctic Ecosystems. Reference Module in Life Sciences, Elsevier.

Convey P., Abbandonato H., Bergan F., Beumer L.T., Biersma E.M., Bråthen V.S., D'Imperio L., Jensen C.K., Nilsen S., Paquin K., Stenkewitz U., Svoen M.E., Winkler J., Müller E. & Coulson S.J. 2015. Survival of rapidly fluctuating natural low winter temperatures by High Arctic soil invertebrates. J. Therm. Biol. **54**: 111–117.

Convey P., Chown S.L., Clarke A., Barnes D.K.A., Bokhorst S., Cummings V., Ducklow H.W., Frati F., Green T.G.A., Gordon S., Griffiths H.J., Howard-Williams C., Huiskes A.H.L., Laybourn-Parry J., Lyons W.B., McMinn A., Morley S.A., Peck L.S., Quesada A., Robinson S.A., Schiaparelli S. & Wall D.H. 2014. The spatial structure of Antarctic biodiversity. Ecol. Monogr. **84**: 203–244.

Convey P., Coulson S.J., Worland M.R. & Sjöblom A. 2018. The importance of understanding annual and shorter-term temperature patterns and variation in the surface levels of polar soils for terrestrial biota. Polar Biol. **41**: 1587–1605.

Convey P. & Peck L.S. 2019. Antarctic environmental change and biological responses. Sci. Adv. **5**: eaaz0888.

Cowan D.A., Khan N., Pointing S.B. & Cary S.C. 2010. Diverse hypolithic refuge communities in the McMurdo Dry Valleys. Antarct. Sci. **22**: 714–720.

Cucini C., Carapelli A., Brunetti C., Molero-Baltanás R., Gaju-Ricart M. & Nardi F. 2021b. Characterization of the complete mitochondrial genome of *Neoasterolepisma foreli* (Insecta: Zygentoma: Lepismatidae) and the phylogeny of basal Ectognatha. Mitochondrial DNA Part B. **6**: 119–121.

Cucini C., Fanciulli P.P., Frati F., Convey P., Nardi F. & Carapelli A. 2020. Re-Evaluating the Internal Phylogenetic Relationships of Collembola by Means of Mitogenome Data. Genes. **12**: 44.

Cucini C., Leo C., Iannotti N., Boschi S., Brunetti C., Pons J., Fanciulli P.P., Frati F., Carapelli A. & Nardi F. 2021a. EZmito: a simple and fast tool for multiple mitogenome analyses. Mitochondrial DNA Part B. **6**: 1101–1109.

Cucini C., Nardi F., Magnoni L., Rebecchi L., Guidetti R., Convey P. & Carapelli A. 2022. Microhabitats, macro-differences: a survey of temperature records in Victoria Land terrestrial and freshwater environments. Antarct. Sci. 1–10.

Dartnall H. 2017. The freshwater fauna of the South Polar Region: A 140-year review. Pap. Proc. R. Soc. Tasman. **151**: 19–57.

Davey M.C., Pickup J. & Block W. 1992. Temperature variation and its biological significance in fellfield habitats on a maritime Antarctic island. Antarct. Sci. **4**: 383–388.

Delsuc F., Phillips M.J. & Penny D. 2003. Comment on "Hexapod Origins: Monophyletic or Paraphyletic?" Science. **301**: 1482–1482.

Dierckxsens N., Mardulyn P. & Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. **45**: e18.

Dong J., Zhang F. & Wang X. 2020. Complete mitochondrial genome of *Pseudachorutes palmiensis* (Collembola: Neanuridae). Mitochondrial DNA Part B. **5**: 394–395.

Elnitsky M.A., Hayward S.A.L., Rinehart J.P., Denlinger D.L. & Lee R.E. 2008. Cryoprotective dehydration and the resistance to inoculative freezing in the Antarctic midge, *Belgica antarctica*. J. Exp. Biol. **211**: 524–530.

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*. J. Therm. Biol. **38**: 264–271.

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

Fu L., Niu B., Zhu Z., Wu S. & Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. **28**: 3150–3152.

Gao F., Chen C., Arab D.A., Du Z., He Y. & Ho S.Y.W. 2019. EasyCodeML: A visual tool for analysis of selection using CodeML. Ecol. Evol. **9**: 3891–3898.

Gao Y., Bu Y. & Luan Y.-X. 2008. Phylogenetic Relationships of Basal Hexapods Reconstructed from Nearly Complete *18S* and *28S* rRNA Gene Sequences. Zoolog. Sci. **25**: 1139–1145.

Godeiro N.N., Pacheco G., Liu S., Gioia Cipola N., Berbel-Filho W.M., Zhang F., Gilbert M.T.P. & Bellini B.C. 2020. Phylogeny of Neotropical Seirinae (Collembola, Entomobryidae) based on mitochondrial genomes. Zool. Scr. **49**: 329–339.

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. Biol. Res. **51**: 8.

Goto S.G., Philip B.N., Teets N.M., Kawarasaki Y., Lee R.E. & Denlinger D.L. 2011. Functional characterization of an aquaporin in the Antarctic midge *Belgica antarctica*. J. Insect Physiol. **57**: 1106–1114.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N. & Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. **29**: 644–652.

Haese C.A.D. 2002. Were the first springtails semi-aquatic? A phylogenetic approach by means of 28S rDNA and optimization alignment. Proc. R. Soc. Lond. B Biol. Sci. **269**: 1143–1151.

Hassanin A., Léger N. & Deutsch J. 2005. Evidence for Multiple Reversals of Asymmetric Mutational Constraints during the Evolution of the Mitochondrial Genome of Metazoa, and Consequences for Phylogenetic Inferences. Syst. Biol. **54**: 277–298.

Hrbáček F., Cannone N., Kňažková M., Malfasi F., Convey P. & Guglielmin M. 2020. Effect of climate and moss vegetation on ground surface temperature and the active layer among different biogeographical regions in Antarctica. CATENA. **190**: 104562.

Hopkin, S.P. 1997. Biology of the springtails (Insecta: Collembola). OUP Oxford.

Hughes K.A., Convey P. & Turner J. 2021. Developing resilience to climate change impacts in Antarctica: An evaluation of Antarctic Treaty System protected area policy. Environ. Sci. Policy. **124**: 12–22.

International Governmental Panel on Climate Change (IPCC). 2022. Climate Change 2022: Impacts, Adaptation and Vulnerability. IPCC Special Report 6th Assestment.

Kaspari M. & Valone T.J. 2002. On ectotherm abundance in a seasonal environment—studies of a desert ant assemblage. Ecology. **83**: 2991–2996.

Kishino H. & Hasegawa M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J. Mol. Evol. **29**: 170–179.

Kishino H., Miyata T. & Hasegawa M. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. **31**: 151–160.

Kristensen N.P. 1998. The groundplan and basal diversification of the hexapods. pp. 281–293. In: Fortey R.A. & Thomas R.H. (eds), Arthropod Relationships, Springer Netherlands, Dordrecht.

Lanfear R., Frandsen P.B., Wright A.M., Senfeld T. & Calcott B. 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol. Biol. Evol. msw260.

Lee B.D. 2018. Python Implementation of Codon Adaptation Index. J. Open Source Softw. **3**: 905.

Lee J.R., Raymond B., Bracegirdle T.J., Chadès I., Fuller R.A., Shaw J.D. & Terauds A. 2017. Climate change drives expansion of Antarctic ice-free habitat. Nature. **547**: 49–54.

Lembrechts J. et al. Global maps of soil temperature. 2022. Global Change Biology. 28: 3110-3144

Leo C., Carapelli A., Cicconardi F., Frati F. & Nardi F. 2019. Mitochondrial Genome Diversity in Collembola: Phylogeny, Dating and Gene Order. Diversity. **11**: 169.

Leo C., Nardi F., Cucini C., Frati F., Convey P., Weedon J.T., Roelofs D. & Carapelli A. 2021. Evidence for strong environmental control on bacterial microbiomes of Antarctic springtails. Sci. Rep. **11**: 2973.

Li D., Liu C.-M., Luo R., Sadakane K. & Lam T.-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. **31**: 1674–1676.

Luan Y., Mallatt J.M., Xie R., Yang Y. & Yin W. 2005. The Phylogenetic Positions of Three Basal-Hexapod Groups (Protura, Diplura, and Collembola) Based on Ribosomal RNA Gene Sequences. Mol. Biol. Evol. **22**: 1579–1592.

MacManes M.D. 2018. The Oyster River Protocol: a multi-assembler and kmer approach for de novo transcriptome assembly. PeerJ. **6**: e5428.

McLoughlin S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. Aust. J. Bot. **49**: 271.

Miller M.A., Pfeiffer W. & Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE), 1–8.

Misof B., et al. 2014. Phylogenomics resolves the timing and pattern of insect evolution. Science. **346**: 763–767.

Moorhead D.L., Wall D.H., Virginia R.A. & Parsons A.N. 2002. Distribution and life-cycle of Scottnema lindsayae (Nematoda) in Antarctic soils: a modeling analysis of temperature responses. Polar Biol. **25**: 118–125.

Morley S.A., Peck L.S., Sunday J.M., Heiser S. & Bates A.E. 2019. Physiological acclimation and persistence of ectothermic species under extreme heat events. Glob. Ecol. Biogeogr. **28**: 1018–1037.

Nardi F., Cucini C., Leo C., Frati F., Fanciulli P.P. & Carapelli A. 2020. The complete mitochondrial genome of the springtail *Allacma fusca* , the internal phylogenetic relationships and gene order of Symphypleona. Mitochondrial DNA Part B. **5**: 3103–3105.

Nardi F., Spinsanti G., Boore J.L., Carapelli A., Dallai R. & Frati F. 2003. Hexapod Origins: Monophyletic or Paraphyletic?. Science. **299**: 1887–1889.

Obryk M.K., Doran P.T., Fountain A.G., Myers M. & McKay C.P. 2020. Climate From the McMurdo Dry Valleys, Antarctica, 1986–2017: Surface Air Temperature Trends and Redefined Summer Season. J. Geophys. Res. Atmospheres. **125**: .

Patro R., Duggal G., Love M.I., Irizarry R.A. & Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods. **14**: 417–419.

Peck L.S., Convey P. & Barnes D.K.A. 2005. Environmental constraints on life histories in Antarctic ecosystems: tempos, timings and predictability. Biol. Rev. **81**: 75.

Phillips L.M., Leihy R.I. & Chown S.L. 2022. Improving species-based area protection in Antarctica. Conserv. Biol. .

Polvani L.M., Banerjee A., Chemke R., Doddridge E.W., Ferreira D., Gnanadesikan A., Holland M.A., Kostov Y., Marshall J., Seviour W.J.M., Solomon S. & Waugh D.W. 2021. Interannual SAM Modulation of Antarctic Sea Ice Extent Does Not Account for Its Long-Term Trends, Pointing to a Limited Role for Ozone Depletion. Geophys. Res. Lett. **48**: .

Quayle W.C., Convey P., Peck L.S., Ellis-Evans C.J., Butler H.G. & Peat H.J. 2003. Ecological responses of maritime Antarctic lakes to regional climate change. Antarct. Res. Ser. **79**: 159–170.

Quayle W.C., Peck L.S., Peat H., Ellis-Evans J.C. & Harrigan P.R. 2002. Extreme Responses to Climate Change in Antarctic Lakes. Science. **295**: 645–645.

Rambaut A., Drummond A.J., Xie D., Baele G. & Suchard M.A. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. **67**: 901–904.

Reed A.J. & Thatje S. 2015. Long-term acclimation and potential scope for thermal resilience in Southern Ocean bivalves. Mar. Biol. **162**: 2217–2224.

Rinehart J.P., Hayward S.A.L., 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. Proc. Natl. Acad. Sci. **103**: 14223–14227.

Robertson G., Schein J., Chiu R., Corbett R., Field M., Jackman S.D., Mungall K., Lee S., Okada H.M., Qian J.Q., Griffith M., Raymond A., Thiessen N., Cezard T., Butterfield Y.S., Newsome R., Chan S.K., She R., Varhol R., Kamoh B., Prabhu A.-L., Tam A., Zhao Y., Moore R.A., Hirst M., Marra M.A., Jones S.J.M., Hoodless P.A. & Birol I. 2010. De novo assembly and analysis of RNA-seq data. Nat. Methods. **7**: 909–912.

Robinson J.T. 2011. Integrative genomics viewer. C O Rresp O N N Ce. **29**: 3.

Robinson S.A., Klekociuk A.R., King D.H., Pizarro Rojas M., Zúñiga G.E. & Bergstrom D.M. 2020. The 2019/2020 summer of Antarctic heatwaves. Glob. Change Biol. **26**: 3178–3180.

Ronquist F., Teslenko M., van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu L., Suchard M.A. & Huelsenbeck J.P. 2012. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. **61**: 539–542.

Schneider C., Cruaud C. & D'Haese C.A. Unexpected diversity in Neelipleona revealed by molecular phylogeny approach (Hexapoda, Collembola). 16.

Seppey M., Manni M. & Zdobnov E.M. 2019. BUSCO: Assessing Genome Assembly and Annotation Completeness. pp. 227–245. In: Kollmar M. (eds), Gene Prediction, Springer New York, New York, NY.

Shimodaira H. & Hasegawa M. 1999. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. Mol. Biol. Evol. **16**: 1114–1116.

Solomon S. 2019. The discovery of the Antarctic ozone hole. Nature. **75**: 46-47

Strimmer K. & Rambaut A. 2002. Inferring confidence sets of possibly misspecified gene trees. Proc. R. Soc. Lond. B Biol. Sci. **269**: 137–142.

Suchard M.A., Lemey P., Baele G., Ayres D.L., Drummond A.J. & Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. **4**: .

Sun X., Yu D., Xie Z., Dong J., Ding Y., Yao H. & Greenslade P. 2020. Phylomitogenomic analyses on collembolan higher taxa with enhanced taxon sampling and discussion on method selection. PLOS ONE. **15**: e0230827.

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 Res. gkv711.

Teets N.M. & Denlinger D.L. 2014. Surviving in a frozen desert: environmental stress physiology of terrestrial Antarctic arthropods. J. Exp. Biol. **217**: 84–93.

Teets N.M., Peyton J.T., Colinet H., Renault D., Kelley J.L., Kawarasaki Y., Lee R.E. & Denlinger D.L. 2012. Gene expression changes governing extreme dehydration tolerance in an Antarctic insect. Proc. Natl. Acad. Sci. **109**: 20744–20749.

Terauds A. & Lee J.R. 2016. Antarctic biogeography revisited: updating the Antarctic Conservation Biogeographic Regions. Divers. Distrib. **22**: 836–840.

Thompson D.W.J., Solomon S., Kushner P.J., England M.H., Grise K.M. & Karoly D.J. 2011. Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. Nat. Geosci. **4**: 741–749.

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*. Mar. Genomics. **3**: 35–44.

Timmermans M., Roelofs D., Mariën J. & van Straalen N. 2008. Revealing pancrustacean relationships: Phylogenetic analysis of ribosomal protein genes places Collembola (springtails) in a monophyletic Hexapoda and reinforces the discrepancy between mitochondrial and nuclear DNA markers. BMC Evol. Biol. **8**: 83.

Timmermans M.J.T.N., Roelofs D., Nota B., Ylstra B. & Holmstrup M. 2009. Sugar sweet springtails: on the transcriptional response of *Folsomia candida* (Collembola) to desiccation stress. Insect Mol. Biol. **18**: 737–746.

Turner J., Barrand N.E., Bracegirdle T.J., Convey P., Hodgson D.A., Jarvis M., Jenkins A., Marshall G., Meredith M.P., Roscoe H., Shanklin J., French J., Goosse H., Guglielmin M., Gutt J., Jacobs S., Kennicutt M.C., Masson-Delmotte V., Mayewski P., Navarro F., Robinson S., Scambos T., Sparrow M., Summerhayes C., Speer K. & Klepikov A. 2014. Antarctic climate change and the environment: an update. Polar Rec. **50**: 237–259.

Turner J., Lu H., White I., King J.C., Phillips T., Hosking J.S., Bracegirdle T.J., Marshall G.J., Mulvaney R. & Deb P. 2016. Absence of 21st century warming on Antarctic Peninsula consistent with natural variability. Nature. **535**: 411–415.

Turner J., Marshall G.J., Clem K., Colwell S., Phillips T. & Lu H. 2020. Antarctic temperature variability and change from station data. Int. J. Climatol. **40**: 2986–3007.

Wauchope H.S., Shaw J.D. & Terauds A. 2019. A snapshot of biodiversity protection in Antarctica. Nat. Commun. **10**: 946.

Wernersson R. 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. **31**: 3537–3539.

Woods H.A., Dillon M.E. & Pincebourde S. 2015. The roles of microclimatic diversity and of behavior in mediating the responses of ectotherms to climate change. J. Therm. Biol. **54**: 86–97.

Xiong Y., Gao Y., Yin W. & Luan Y. 2008. Molecular phylogeny of Collembola inferred from ribosomal RNA genes. Mol. Phylogenet. Evol. **49**: 728–735.

Yu D., Zhang F., Stevens M.I., Yan Q., Liu M. & Hu F. 2016. New insight into the systematics of Tomoceridae (Hexapoda, Collembola) by integrating molecular and morphological evidence. Zool. Scr. **45**: 286–299.