DR. ANGELA MCGAUGHRAN (Orcid ID : 0000-0002-3429-8699)
DR. CERIDWEN FRASER (Orcid ID : 0000-0002-6918-8959)

Article type : Original Article

Corresponding author mail id:- ang.mcgaughran@gmail.com

Running title: Dispersal from Antarctic glacial refugia

Genome-wide SNP data reveal improved evidence for Antarctic glacial refugia and dispersal of terrestrial invertebrates

Angela McGaughran[1,2*], Aleks Terauds[3,4], Peter Convey[5], Ceridwen I. Fraser[4,6]

[1] Research School of Biology, Division of Ecology and Evolution, Australian National University, Canberra, ACT 2601, Australia

[2] Black Mountain Laboratories, Commonwealth Scientific and Industrial Research Organisation, Acton, ACT 2601, Australia

[3] Department of Energy and the Environment, Australian Antarctic Division, Kingston, Tasmania 7050, Australia.

[4] Fenner School of Environment and Society, College of Science, Australian National University, Canberra, ACT 2601, Australia

[5] British Antarctic Survey, NERC, High Cross, Madingley Road, Cambridge CB3 0ET, United Kingdom

[6] Department of Marine Science, University of Otago, PO Box 56, Dunedin 9054, New Zealand

[*] Corresponding author

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/MEC.15269

Abstract

Antarctica is isolated, surrounded by the Southern Ocean, and has experienced extreme environmental conditions for millions of years, including during recent Pleistocene glacial maxima. How Antarctic terrestrial species might have survived these glaciations has been a topic of intense interest, yet many questions remain unanswered, particularly for Antarctica's invertebrate fauna. We examine whether genetic data from a widespread group of terrestrial invertebrates, springtails (Collembola, Isotomidae) of the genus *Cryptopygus,* show evidence for long-term survival in glacial refugia along the Antarctic Peninsula. We use genome-wide SNP analyses (via genotyping-by-sequencing, GBS) and mitochondrial data to examine population diversity and differentiation across more than 20 sites spanning >950 km on the Peninsula, and from islands both close to the Peninsula and up to ~1,900 km away. Population structure analysis indicates the presence of strong local clusters of diversity, and we infer that patterns represent a complex interplay of isolation in local refugia coupled with occasional successful long-distance dispersal events. We identified wind and degree days as significant environmental drivers of genetic diversity, with windier and warmer sites hosting higher diversity. Thus, we infer that refugial areas along the Antarctic Peninsula have allowed populations of indigenous springtails to survive *in situ* throughout glacial periods. Despite the difficulties of dispersal in cold, desiccating conditions, *Cryptopygus* springtails on the Peninsula appear to have achieved multiple long-distance colonisation events, most likely through wind-related dispersal events.

Key words: Antarctic Peninsula, dispersal, mitochondrial, refugia, SNP, springtails

Introduction

Understanding the drivers of biodiversity changes in the past is crucial if we are to predict and manage biodiversity in the future (Perrings, Duraiappah, Larigauderie, & Mooney, 2011). We know a great deal about the extent, distribution, and stability of biological diversity globally, and that such diversity changes through time over both large and small temporal scales (Dornelas et al., 2013). These natural changes track environmental conditions and reflect dynamic ecological and evolutionary processes. However, anthropogenic activities are currently causing profound transformations in ecosystems and unprecedented losses of biological diversity (Magurran & Dornelas, 2010), and the ways in which species are coping with these changes are not often clear.

Antarctica has been no stranger to processes of environmental change. Since its separation from other Gondwanan continents, Antarctica has experienced a cooling climate accompanied by a decrease in biological diversity throughout the major epochs since the Eocene (~56 - 33.9 Ma) - this has involved a shift from angiosperm- to tundra- to alpine-dominated vegetation, with dry and frozen conditions persisting for the last ~14 Ma (Convey, 2017). More recently, Pliocene (5 - 2.6 Ma) and multiple Pleistocene (2.6 Ma - 10 Ka) glaciations, including the Last Glacial Maximum (LGM; ~22 - 17 Ka), saw thickened ice sheets that extended well beyond their current limits (Mackintosh et al., 2014).

Recent estimates suggest that today < 0.5% of Antarctica is ice-free (Brooks, Jabour, van den Hoff, & Bergstrom, 2019; Burton-Johnson, Black, Fretwell, & Kaluza-Gilbert, 2016) and the contemporary terrestrial biota is predominantly restricted to low altitude coastal areas (Convey, 2017; Convey et al., 2009), where seasonally ice- and snow-free habitats occur and liquid water is available. This biota is depauperate and often cryptic, but it is also unique and in many cases endemic to Antarctica, with the broader continental regions showing distinct signatures of biological diversity and often not sharing close relatives with lower latitude southern regions (Adams et al., 2006; Pugh & Convey, 2008; Terauds et al., 2012; Chown, Clarke, Fraser, Cary, Moon, et al., 2015). In addition, Antarctica is surrounded by the Southern Ocean and distant from other continents by a minimum of 1,000 km across the Drake Passage (and more generally up to 5,000 km). This, along with the strong oceanic Antarctic Circumpolar Current and atmospheric circulation patterns, means the continent should be predominantly isolated from an influx of dispersing colonists (Barnes, Hodgson, Convey, Allen, & Clarke, 2006; Clarke, Barnes, & Hodgson, 2005; Fraser, Nikula, Ruzzante, & Waters, 2012). Although recent research has

challenged this idea of biological isolation on contemporary timescales (e.g., Fraser et al., 2018), analysis of molecular genetic diversity in a range of terrestrial taxa has found that many extant Antarctic organisms show evidence of long-term *in situ* evolution and persistence (Bennett, Hogg, Adams, & Hebert, 2016; Chong, Pearce, & Convey, 2015; Convey et al., 2008, 2009; Iakovenko et al., 2015).

For many terrestrial species to have survived for millions of years in Antarctica, patches of ice-free habitat must have been available throughout Pleistocene and earlier ice ages. Recent work has identified the potential importance of geothermal refugia in contributing to such survival in some parts of Antarctica and has shown that diverse organisms still occur in Antarctic geothermal environments (Convey & Smith, 2006; Convey & Stevens, 2007; Fraser, Terauds, Smellie, Convey, & Chown, 2014; Fraser et al., 2012). In geothermal Antarctic areas today, soils can reach temperatures of more than 60°C (Convey & Smith, 2006; Soo, Wood, Grzymski, McDonald, & Cary, 2009) and soil and air in geothermal caves can also be remarkably warm (up to 25°C air temperature; Curtis & Kyle, 2011). Additional, non-geothermal refugia are assumed to have existed at inland locations, such as nunataks, and in the ice-free areas of the Dry Valleys (Victoria Land and the Transantarctic Mountains). Thus, Antarctic species are thought to have persisted during glacial maxima in refugial hubs and subsequently recolonised newly ice-free areas in periods of glacial retreat (Convey & Smith, 2006; Convey & Stevens, 2007; Fraser et al., 2012, 2014). Though postglacial recolonisation patterns and processes have enjoyed a great deal of focus in Northern Hemisphere taxa, comparatively less work has addressed recolonization of southern high latitudes (see Fraser et al., 2012).

The Antarctic terrestrial fauna is dominated by a limited number of terrestrial invertebrate groups, including arthropods (insects, springtails, and mites), nematodes, rotifers, and tardigrades (Chown & Convey, 2016; Convey, 2017). These species are generally characterised by a low dispersal ability (e.g., Convey, 1996; Coulson, Hodkinson, Webb, & Harrison, 2002; Hayward, 2004). Springtails in particular are presumed to disperse through the soil column in order to avoid exposure to desiccating conditions (McGaughran, Hogg, & Convey, 2011), and polar springtails tend to occur on the underside of rocks in moist ice-free areas such as snow patches, glacial foregrounds, and damp soils (Kennedy, 1993; Stevens & Hogg, 2002). Previous genetic work on springtails has consistently found evidence for long-term persistence, high genetic diversity, strong genetic structure, and fragmentation among populations (reviewed in McGaughran,

Stevens, Hogg, & Carapelli, 2011; see also Carapelli, Convey, Frati, Spinsanti, & Fanciulli, 2017; Collins, Hogg, Convey, Barnes, & McDonald, 2019).

This earlier molecular work included assessment of the phylogeographic structure of *Cryptopygus antarcticus antarcticus* Willem, a springtail that makes a useful model organism, as it is indigenous to, and abundant along, the Antarctic Peninsula. The Antarctic Peninsula is a long narrow mountain chain that supports ~120,000 km^2 of grounded ice and merges with the West Antarctic Ice Sheet at its most southern point (Barker & Camerlenghi, 2001; Smellie, McIntosh, & Esser, 2006). The area is characterised by extensive ice shelves, deep fjords, glaciers, and islands, and reaches elevations of up to 2,000 m above sea level (Smellie et al., 2006). This complex topography, combined with the extension of the Peninsula across ~10 degrees of latitude, sets up strong climatic gradients and makes the Antarctic Peninsula an ideal location to examine the ways in which species have persisted in Antarctica through glacial cycles.

In *C. a. antarcticus*, long-term persistence in Antarctica has been inferred previously (McGaughran et al., 2010; Stevens et al. 2005), with marked population structure detected, and evidence of early dispersal events, followed by limited contemporary contact (McGaughran et al., 2010). However, this former work was restricted to one (Stevens et al., 2005; n = 14) or two (McGaughran et al., 2010; n = 140 and 240) mitochondrial genes and modern, higher-resolution, genomic approaches could help to clarify differences and similarities among populations. Genomic data for Antarctic species are extremely scarce, perhaps partly due to the absence of 'model' or 'reference' data for these unique species, including *C. a. antarcticus*. However, genotyping by sequencing (GBS) is a useful genomic approach for exploring genetic patterns on a genome-wide scale in the absence of reference data (Elshire et al., 2011). With such methods, single nucleotide polymorphism (SNP) discovery occurs alongside genotyping, providing a rapid, high-throughput, and cost-effective tool to obtain genome-wide data for inferring evolutionary history.

Here, we used GBS data to examine Antarctic Peninsula *Cryptopygus* springtails and shed light on the long-term, *in situ*, survival of these Antarctic invertebrates. Specifically, we set out to determine whether single or multiple areas might have provided long-term refugia in the region, and whether movement among sites has been restricted to short, local dispersal or has included many long-distance events. Given the apparently limited active dispersal capacity of these springtails, we applied two model-based approaches to help elucidate the nature and drivers of inferred movement among populations. We also compared genetic signals from GBS and both

published and new mitochondrial data to examine how biogeographic inferences can change with the addition of higher-resolution genomic information.

Materials and Methods

Samples

Samples, consisting of moss, algae, and underlying soils, were collected by PC during British Antarctic Survey expeditions in 2008, 2011, and 2015, from 14 geographic locations along the Antarctic Peninsula, where populations ranged from inland to coastal sites and were separated by 6 - 791 km (Fig. 1; Table 1). Springtails and other micro-arthropods were extracted from these samples using a simple Tullgren extraction, and hand-sorted to separate the distinctive *C. a. antarcticus*, which is often the most common springtail species encountered (Convey & Smith, 1997). Additional samples from Saunders Island ('SAU' in the South Sandwich Islands, 1,892 km from the Peninsula tip; Fig. 1) were also included. All samples were stored directly in 100% ethanol, and later kept frozen, until DNA extraction. An additional 89 unique mitochondrial haplotype sequences were downloaded from GenBank (Accession Numbers: GQ215400-GQ215488) and information from McGaughran et al. (2010) was used to recreate their original 140 sequences (i.e., which had been collapsed into only unique haplotypes for GenBank submission); this also included the 14 sequences available from Stevens et al. (2005). These 140 samples were originally collected from an additional 14 locations across the Peninsula (Fig. 1; Table 1). A total of 204 (64 new) and 384 samples were used for mitochondrial and GBS analyses, respectively (Table 1).

Mitochondrial extractions, amplifications, and sequencing for 'new' samples

Total DNA was extracted from specimens using a DNEasy Blood and Tissue extraction kit in 96-well plate format (Qiagen, Valencia, CA, United States), and fragments of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene were amplified using the universal primers LCO1490 (5′-GGTCAACAAATCATAAAGATATTGGA-3′) and HCO2198 (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′) (Folmer, Black, Hoeh, Lutz, & Vrijenhoek, 1994). Amplifications for each specimen used a 25 μL reaction volume containing 2 μL of extracted DNA (unquantified), 12.5 μL of MyTaq HS Mix (Bioline, Alexandria, NSW, Australia), 1 μL of each 10 mM forward and reverse primer, and 8.5 μL of deionised water. PCRs were

carried out in an Eppendorf Mastercycler Nexus (Eppendorf, Hamburg, Germany) under the following conditions: 94°C for 2 min followed by five cycles of denaturation and polymerase amplification (94°C for 40 s, 45°C for 40 s, 1 min at 72°C) followed by 35 cycles of 94°C for 1 min, 51°C for 1.5 min and 1 min at 68°C, followed by 10 min at 72°C. Amplification success was assessed via gel electrophoresis and bands of between 500-1,000 bp were excised and cleaned using a PCR Purification kit (Qiagen), double-eluting with 20 µL of EB buffer. Sequencing used forward primers and was performed directly on a capillary ABI3730 Genetic Analyzer (Applied Biosystems Inc., Foster City, CA, United States) at the Biomolecular Resource Facility (John Curtin School of Medical Research, Australian National University).

GBS library preparation and sequencing

DNA was extracted for GBS from 384 samples following the method outlined above for mitochondrial DNA. Genotyping-by-sequencing library preparation followed the protocols of Elshire et al. (2011) with some modifications (Morris, Grabowski, & Borevitz, 2011): a uniquely barcoded PstI adapter (2.25 ng) was added to each sample before DNA digestion (incubation at 37°C for 2 h), which was performed using 4 U PstI-HF (New England Biolabs, Ipswich, MA, United States) in 1 X CutSmart Buffer. Combinatorial adapters were then ligated with T4 DNA ligase in 1 X ligation buffer (New England Biolabs), followed by incubation at 16°C for 90 min and 80°C for 30 min. Purification was performed using a Qiagen MinElute PCR Purification kit (Qiagen), with elution in 25 µl 1 X TE buffer. PCRs were carried out in 50 µl volumes, containing 10 µl purified DNA, 1 X MyTaq HS Master Mix (Bioline), and 1 µM each of PCR primers
5'AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC*T and
5'CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC*T (where * indicates phosphorothioation). PCRs were run in an Eppendorf Mastercycler Nexus (Eppendorf, Hamburg, Germany) under the following conditions: 72°C for 5 min, 95°C for 60 s, and 24 cycles of 95°C for 30 s, 65°C for 30 s, and 72°C for 30 s, with a final extension step at 72°C for 5 min. Concentrations for each sample were assessed using a LabChip GXII (Caliper Life Sciences, Hopkinton, MA, United States) and all samples were pooled equimolarly (20 ng DNA per sample). Size fractionation of the pooled library was achieved via electrophoresis on a 1.5% agarose gel, with a 300 bp size range (from 200 - 500 bp) selected for paired-end sequencing

(76 bp read lengths), which was carried out on one lane of an Illumina HiSeq 2500 (Illumina, San Diego, CA, United States) at the Biomolecular Resource Facility (John Curtin School of Medical Research, Australian National University).

Mitochondrial analysis

The 64 new mitochondrial DNA sequences were aligned, trimmed and manually checked for ambiguities using Geneious v6.1.8 (http://www.geneious.com/) and then aligned by eye to the 140 sequences from McGaughran et al. (2010) and trimmed to remove missing data. Genetic parameters, including the numbers of polymorphic sites and haplotypes, were obtained for each population using the program Arlequin v.3.5.2.2 (Excoffier & Lischer, 2010). The mean number of pairwise nucleotide differences (k), nucleotide diversity (π), theta (S), and their corresponding variances were calculated following equations from Nei (1987), as implemented in Arlequin. The demographic history of *C. a. antarcticus* was analysed using neutrality tests (Tajima's D test and Fu's FS test; Tajima, 1989; Fu, 1997) with 10,000 permutations in Arlequin. Analysis of molecular variance components (AMOVA; Excoffier, Smouse, & Quattro, 1992) was carried out with 10,000 permutations in Arlequin, to estimate genetic differentiation within and between populations. To determine whether any broader, regional structure existed among populations, we also performed hierarchical AMOVA tests using several definitions for 'region'. This included matching the 'northern', 'central', and 'southern' regions delineated in McGaughran et al. (2010), as well as testing structure that separated out populations collected over very small geographic scales. For example, we defined BYE, POT, HAR, HAN, and COP (all from the South Shetland Islands), ANC, LEO, JEN, MAC, ROT, BKIL and REP (all from/close to Adelaide Island), and SAU (South Sandwich Islands, ~1,900 km away from the Antarctic Peninsula tip) as regions, with all remaining Peninsula sites as a single region, or split into two regions roughly around the middle of the Peninsula (see Fig. 1, and population codes in Table 1). A haplotype network was generated in R v3.4.1 (R Core Team, 2019) using the package pegas v0.10 (Paradis, 2010), to depict phylogeographic relationships among haplotypes via the haplotNet function defaults (i.e., the haplotype network is built using an infinite sites model, pairwise deletion of missing data and the statistical parsimony algorithm of Templeton et al. 2010), and novel haplotypes were submitted to the GenBank database (accession numbers: MN473376-MN473393).

Some of our diversity estimates were found to be quite high relative to the others (e.g., nucleotide diversity values > 0.010; see Results). To investigate this further, we performed a

species delimitation analysis using Automatic Barcode Gap Discovery (ABGD; Puillandre et al. 2012) software at its web interface (bioinfo.mnhm.fr/abi/public/abgd/abgdweb.html; performed on 09 September 2019). We ran the analysis using a default value of relative gap width (X = 1.5), and three different models of nucleotide substitution ('simple', 'Jukes-Cantor – JC69', and 'Kimura – K80). All assignments for intraspecfic divergence (*P*), values between 0.001 and 0.100 were recorded, while other parameter values employed defaults. Our initial analysis identified two putative species groups, with 6/198 individuals falling into a second group separate from the rest. We then removed these six individuals and repeated the analysis as outlined above. To further investigate the potential for multiple species in the dataset, we used Molecular Evolutionary Genetics Analysis (MEGA) v10.0.5 (Kumar et al. 2018) to evaluate simple p-distances between the two groups identified by ABGB. The effects of possible cryptic species in our data were explored by running the diversity and differentiation analyses outlined above with the six individuals removed.

Finally, a suite of models were fitted to test the ability of a range of environmental parameters (Supporting Information S1) to explain the variation observed in several measures of genetic diversity (# sequences, # unique haplotypes, # polymorphic sites, mean # pairwise differences (k), nucleotide diversity (π) and theta (S)). Environmental parameters (source data used to derive these are provided in Supporting Information S1) were interpolated or downscaled to 1 km raster spatial layers (covering the extent of Antarctica). Corresponding values of each cell were then allocated to each genetic sampling site using a spatial overlay in the R package, raster v.2.0-12 (Hijmans & van Etten, 2012). For each response variable (i.e., genetic diversity metric), generalised linear models were fitted, starting with the saturated model (including all potential predictors). Using a stepwise model selection process (via the stepAIC function in the R package MASS v7.3-51.4 (Venables & Ripley, 2002), the best model was selected on the basis of the lowest AIC value. Gaussian, Poisson and negative Binomial distributions were fitted for each response variable and model fit was assessed using standard model diagnostic tools (QQ-plots, residuals *vs*. linear predictors, residual histograms).

GBS analysis pipeline

Raw Illumina data were processed *de novo* using ipyrad v0.7.28 (http://ipyrad.readthedocs.io). This software employs demultiplexing and various filtering and clustering steps to ultimately identify reads at each locus. To determine final filtering settings in ipyrad, we changed default

settings to explore the effect of several parameters on final output files, including filter_min_trim_len (60, which sets the minimum length a read can be trimmed down to during filtering; default: 35), min_samples_locus (defines the minimum number of samples that must have data at a given locus for it to be retained in the final dataset – we tried values of 2, 3, 4, 8, and 12; default: 4), max_low_qual_bases (10, which sets the upper limit on the number of Ns allowed in reads; default: 5), phred-Qscore-offset (23, which allows a 95% probability of a correct base call; default: 33), max_Ns_consens (10,10, which defines the maximum number of uncalled bases allowed in consensus sequences; default: 5,5). However, these filters had little effect in terms of removing individuals with lower quality data, or changing preliminary phylogenetic tree topologies (data not shown). Thus, a final ipyrad run was completed with only the following non-default settings: filter_adapters (2, where adapters are removed), filter_min_trim_len (60).

Vcftools v0.1.13 (Danecek et al., 2011) was then run with --missing-indv and -max-missing-count parameters to explore the effects of missing data on our SNP matrix. We finally decided to proceed with two datasets: (1) a less conservative dataset created by filtering individuals with more than 98% missing data: n = 162 individuals and 3,491 binary SNPs; and (2) a more conservative dataset created by applying the 98% missing data filter in (1) as well as setting missing genotype data across all individuals to 20%, and setting a minor allele frequency (MAF) cut-off of 5%: n = 162 individuals and 735 binary SNPs. We present the results for (2) in the main text and for (1) in the Supporting Information.

We used the R packages vcfR v1.8.0 (Knaus & Grünwald, 2016), adegenet v2.1.1 (Jombart, 2008), and poppr v.2.8.1 (Kamvar, Tabima, & Grünwald, 2014) to examine patterns of genetic differentiation in Antarctic Peninsula springtails. We calculated Nei's distance (Nei, 1987) and Weir and Cockerham's F_{ST} (Weir & Cockerham, 1984) among populations. Mantel tests were performed in R to test for association between geographic and genetic distances for all populations, and for all populations excluding the non-Peninsula site, Saunders Island (SAU). We used the distance in kilometres between sites as the geographic distance, calculated by inputting GPS coordinate information for each location to Geographic Distance Matrix Generator v1.2.3 (Ersts, n.d.), and both Nei's and F_{ST} genetic distances. Mantel tests were executed with 10,000 permutations. PCA analysis was performed in R to examine differentiation based on orthogonal transformation of the SNP data. fastSTRUCTURE v1.0 (Raj, Stephens, & Pritchard, 2014; Skotte, Korneliussen, & Albrechtsen, 2013) was used to identify admixture proportions among populations using SNP calls. We tested 1–10 genetic clusters by specifying the K parameter and

ran five replicates for each K-value, using a simple prior. Results for each K-value were visualised using the distruct.py script associated with the fastSTRUCTURE program. Finally, SpaceMix v.0.13 (Bradburd et al. 2016) was used to perform a spatially-explicit analysis of gene flow between populations – producing a 'geogenetic' map, in which the distances between samples reflect genetic rather than geographic distances, and inferring potential long-distance admixture events. In the underlying model, SpaceMix estimates allele frequency covariances associated with geography and non-local gene flow can be identified as significantly strong covariance over long distances. In our SpaceMix analyses, we removed individuals with more than 50% missing data– this resulted in a final dataset for analysis that included 65 individuals (with representatives from all 14 sites) and 735 SNP loci. To run the software, we performed 10 initial fast replications of 10 generations and a final 'long' run of 1e+07 generations, with a sampling frequency of 1,000. In all runs, we used "source and target" as the model option and "sample.covariance" as the data-type.

To further investigate the variation in diversity patterns with respect to environmental drivers, we fitted models as outlined above for mitochondrial data. In the case of the GBS data, we used genetic differentiation as a measure of the genetic differences among 13 populations (with SAU and DEC excluded to match the sites for which we had environmental data extracted), using both Nei's and F_{ST} matrices, as calculated above.

Results

Mitochondrial data

Our mitochondrial dataset consisted of 204 samples from 27 populations (all except Deception Island; Table 1) and the final alignment was 546 bp in length. Measures of genetic diversity varied among populations. For example, nucleotide diversity ranged from 0.000±0.000 to 0.043±0.023 (Table 2). Population neutrality indices (Tajima's D, Fu's FS) were often significantly low and negative, although patterns differed for each statistic. Specifically, five populations showed low, negative Tajima's D values and four of these were located towards the southern part of the Peninsula (Table 2). Meanwhile 16 of the 27 populations showed low and negative Fu's FS values and these populations were distributed all along the Peninsula and surrounding islands. Low neutrality indices are often taken to indicate an excess of low frequency

polymorphisms (Tajima's D) or number of alleles (Fu's FS), as would be expected from recent population expansions or from genetic hitchhiking.

From a total of 204 sequences, we obtained 94 unique haplotypes (Fig. 2). This included 76 of the 89 haplotypes identified in McGaughran et al. (2010), as well as 18 novel haplotypes. In the current study, we used the same software (Arlequin) to estimate haplotype diversity, and the same statistical parsimony algorithm (Templeton et al., 1992) to build a haplotype network; haplotypic differences between the former and current study are due to different sequence alignment lengths (618 bp in McGaughran et al. 2010 and 546 bp here).

Overall, twenty-one haplotypes were shared among individuals, and six of these instances occurred exclusively within populations: BYE, ECL, PET, SPE, MAC (x2), and ROT. The remaining 15 haplotypes were shared among individuals from different populations. Nine of these inter-population haplotype-sharing events involved populations that were in close geographic proximity. For example, two haplotypes were shared between COP and HAR (neighbouring islands ~25 km apart), and several were shared among close central or southern Peninsula locations, including ECL/PAR, ECL/PRO, ECL/PAR/PTL, ECL/TER, and PAR/PTL (x3). The six remaining instances of inter-population haplotype sharing occurred over larger geographic distances (Fig. 2). This included haplotype sharing between the furthest north and furthest south populations (COP/JEN, BYE/ANC/MAC/ROT, HAN/ALA/SAU, DET/BYE), and between the central and southern populations (PRO/ALA/DET/EVP/KIL/REP/ROT, BER/ECL/GRA/PRO/TER/JEN/KIL/REP, KIL/LEO). The occurrence of haplotype sharing across very large geographic distances implies retained shared ancestral haplotypes or extremely long-distance dispersal events across regions of the Antarctic Peninsula and Scotia Arc, including to/from the South Sandwich Islands (SAU).

Despite the occurrence of shared haplotypes across vast geographic distances, an examination of genetic distances (F_{ST}) indicated that populations were highly differentiated (Table 2, Supporting Information S1). Nonetheless, in an exact test of population differentiation (where we test for non-random distribution of haplotypes into population samples under a null hypothesis of panmixia), we found no evidence for 'global' differentiation/non-panmixia ($P = 1.000$; 30,000 Markov steps). Analysis of molecular variance (AMOVA) confirmed this pattern. We performed AMOVA tests for a single panmictic population and in a hierarchical fashion using several 'regional' definitions. In the former test, 57.42% and 42.58% of genetic variation was apportioned among and within populations, respectively ($P < 0.001$). In the latter tests, we tried several

definitions of 'region' (see Materials and Methods) and found that the highest degree of variation was apportioned to regions when we defined three regions, corresponding to South Shetland Island populations (BYE, POT, HAR, HAN, and COP), SAU, and all other populations (15.62%; P = 0.050). In our other tests, the variation apportioned to regions was 13.19% (P = 0.017) when we defined regions according to McGaughran et al. (2010), 0% when we defined three regions (Adelaide Island populations - ANC, LEO, JEN, MAC, ROT, KIL and REP; SAU; rest) and 7.29% (P = 0.099) when we defined four regions (Adelaide Island, South Shetland Islands, SAU, rest). In each of these tests, the proportion of variation attributed among populations was 5-15% higher than that attributed within populations as for the non-hierarchical test. Thus, while some genetic signal can be detected among greater geographic regions, the vast majority (~85 to >90%) of signal comes from variation shared among and within populations, indicating little support for strong regional/hierarchical structure in the mitochondrial data.

Species delimitation analysis in ABGD identified two groups in the data (Table 3). Upon further investigation, we confirmed that this was due to six of the ten samples from Byers Peninsula (BYE) forming their own group in the analysis. A subsequent ABGD analysis with these individuals removed confirmed that their absence resulted in delimitation of a single species group. The two putative species groups were 8.14% divergent (p-distance). With these six samples removed from our diversity analyses, estimates decreased for all measures (k = 6.333(3.801) vs. 23.444(11.286); π = 0.012(0.008) vs. 0.043(0.023); S = 6.545(3.859) vs. 16.967(7.175)). Differentiation analyses were not significantly changed either way (e.g., ANOVA percentages within 1-2%) and these BYE samples were not involved in long-distance haplotype sharing events (Fig. 2).

In our analysis of the environmental drivers of mitochondrial genetic diversity, models which explained a significant proportion of the genetic variation could only be fitted for three of the seven response variables (Supporting Information S1). In the best model for *number of haplotypes* (R^2 = 0.90, F = 4.7, P = 0.018), there was a positive relationship between *wind* and *degree days*, suggesting that warmer areas with higher winds had more unique haplotypes compared to colder sites with less wind. A similar result was observed when we standardised haplotypes by the number of sequences. The best model for this metric only included wind and was the best fit of all models (as indicated by R^2 = 0.30, F = 10.3, P = 0.004).

GBS data

From 384 starting individuals, our SNP calling pipeline recovered 735 binary SNPs from 162 individuals after conservative filtering (all results for a less conservative dataset of 3,491 binary SNPs are presented in the Supporting Information S3 – note that we find overall agreement between the two datasets, with slight differences between them not changing our ultimate conclusions). Graphics showing the sequencing depth per sample and the location of missing data in the SNP matrix (59.38% missing data for the more conservative dataset, 72.51% for the less conservative) are available in Supporting Information S3.

An examination of genetic distances indicated a lack of differentiation at the scale of population moving from north to south down the Peninsula (Table 4, Supporting Information S2, S3). For example, ALA showed low differentiation (F_{ST}=0.086) to the neighbouring (~127 km) ANC population, but higher differentiation to other nearby sites (F_{ST}=0.284 and F_{ST}=0.206 to DET:207 km and EVP:296 km, respectively), and lower differentiation to SPE (F_{ST}=0.131), which is further away (~616 km) (Supporting Information S2). Similarly, GRA had a low genetic distance to neighbouring (~56 km) PTL (Nei's D=0.064) and distant (~435 km) LEO (D=0.087), but a higher genetic distance to neighbouring (~120 km) BER (D=0.127) (see Supporting Information S2).

The lack of concordance between genetic and geographic distance was confirmed with Mantel tests. Although we picked up a signal for isolation by distance when we used Nei's distance as our genetic distance measure and included SAU in our analysis (R = 0.680; $P < 0.001$ and R = 0.448; $P = 0.002$ for our more and less conservative datasets, respectively), this signal was not present in either dataset when SAU was not considered (R = -0.086; $P = 0.824$ and R = -0.086; $P = 0.828$). Results were also non-significant when we used F_{ST} as our distance measure (with SAU: R = 0.254; P =0.068, R = 0.248; $P = 0.112$, without SAU: R = -0.039; P =0.623, R = -0.036; $P = 0.600$, more and less conservative datasets, respectively).

A PCA confirmed the lack of strong regional structure in the GBS dataset (Fig. 3, Supporting Information S3), with the first two principal components (PCs) together accounting for just under 17% of the total variance in the data. After colouring populations by the broader northern, central, and southern regions, it is clear that individuals from different regions mostly overlap throughout the PC space (Fig. 3, Supporting Information S3).

The fastSTRUCTURE algorithm indicated that values of K between 4-6 and 2-5 maximised the marginal likelihood for the more and less conservative datasets, respectively. However, in the case of higher values of K (> 3), for both the conservative and less conservative

datasets, the membership probabilities of individuals in the additional clusters was generally very low (< 0.0003; Supporting Information S3), thus we present the results for K = 2 and 3 in Figure 4. For K = 2, we saw a separation of the more centrally-located Antarctic Peninsula populations from all other populations for the more conservative dataset (Fig. 4), and a separation of the more northern Antarctic Peninsula populations from the remainder for the less conservative dataset (Supporting Information S3). Results for K-values of 3 and 4 were consistent for both conservative and less conservative datasets: for K = 3, we saw separation of the northern, central, and southern Peninsula populations, while K = 4 further divided the central Peninsula populations to separate SPE (Fig. 4). K = 5 further divided the previous result to separate out the more southern population, DET, and K=6 separated out additional northerly-located individuals from COP (Supporting Information S3). Thus, with the fastSTRUCTURE analysis, we detect an interesting degree of regional genetic structure that corresponds to broad northern, central, and southern Antarctic Peninsula areas (i.e., K = 3) and agrees with the patterns of mitochondrial haplotype sharing described above (Fig. 4). However, the admixture we see across all clusters is consistent with our above analyses, indicating an overall pattern of gene flow that links geographically-distant populations across the Antarctic Peninsula.

To examine long-distance dispersal more closely, we employed spatially explicit analyses with SpaceMix software. For each sample, the software identifies the geogenetic position of the sample itself, as well as of the proposed source of admixture/gene flow into that sample. The software identified three main clusters, corresponding to the northern, central, and southern Antarctic Peninsula (Fig. 5) and an overall pattern of inferred long-distance dispersal that is consistent with the PCA and fastSTRUCTURE results (Supporting Information S3). In particular, from the 65 individuals tested (including at least one individual from all 14 populations with GBS data; see Materials and Methods), 37 had a geogenetic position within the Antarctic Peninsula region (i.e., northern, southern, central) of its collection, but a geogenetic position of proposed source of gene flow that was outside that region (Fig. 5; Supporting Information S3). The majority (29) of these cases involved gene flow from neighbouring Peninsula regions (e.g., between southern and central; Supporting Information S3), however, eight of these individuals had an admixture/gene flow source that spanned the Peninsula (southern and northern regions; Fig. 5, Supporting Information S3).

In our analysis of the environmental drivers, models which explained a significant proportion of the genetic variation could only be fitted for F_{ST} (Supporting Information S1). In

this model *rugosity*, *solar radiation*, and *wind* were strongly significant positive predictors of F_{ST}, while *melt* was a significant negative predictor ($R^2 = 0.68$, $F = 7.4$, $P = 0.009$).

Discussion

Broadly, our results indicate that *C. a. antarcticus* springtails form discrete local population genetic clusters along the Antarctic Peninsula that are connected through dispersal events, some of which span vast geographic distances within the sampled distribution (including to/from Saunders Island, ~1,900 km distant). GBS data were consistent with the mitochondrial patterns, indicating that populations were locally differentiated, but also finding support for a degree of broad, regional structure and indicating genetic links between geographically-distant populations. In particular, our fastSTRUCTURE and SpaceMix plots indicated the presence of three genetic clusters that corresponded to a broad delineation of populations located in southern, central, and northern areas of the Antarctic Peninsula distribution, as well as a high degree of admixture within clusters and gene flow within and between regions that supports long-distance population connectivity.

Our results give support for some of the patterns found in previous mitochondrial analyses of *C. a. antarcticus* (McGaughran et al., 2010), but there were some intriguing discrepancies. With 140 individuals spanning populations along the Antarctic Peninsula, the COI data from the previous work showed strong, significant population structure and a pattern the authors attributed to demographic expansion within extremely isolated populations (McGaughran et al., 2010). A large historical refuge was indicated for the northern Peninsula, with smaller refugia inferred along the central and southern Peninsula regions. This inference was based largely on slightly higher genetic diversity in the north, as well as the central placement of northern haplotypes linking to central and southern locations in a haplotype network. McGaughran et al. (2010) also found very limited evidence for haplotype sharing, with only ~11% of haplotypes found in more than one population, and the majority of these shared within the larger geographic regions *vs.* trans-regionally. In the current study, we found evidence for nearly twice as much haplotype sharing (~21%) from a total of 204 samples and just under one third of this sharing occurred across the Antarctic Peninsula regions identified in McGaughran et al. (2010). Our current research thus demonstrates a higher degree of trans-regional movement in small terrestrial invertebrates along the Antarctic Peninsula. This reflects and highlights the stochastic nature of sampling in genetic

studies and the importance of considering research limitations caused by sample sizes and sampling effort.

Our GBS data provided enhanced analytical resolution, allowing us to gain further insights into the evolutionary history of Antarctic Peninsula springtails than could be achieved from mitochondrial data alone. Whereas previous work inferred strong and significant population structure and a pattern of increasing isolation among populations with increased geographic distance (McGaughran et al., 2010), the GBS data showed greater evidence for genetic connectivity among geographically-distant populations. For example, the PCA showed clear overlap of samples from different geographic regions throughout the PC space, and the fastSTRUCTURE analyses indicated a high degree of admixture among clusters. Meanwhile, the SpaceMix analysis found strong evidence for frequent local and occasional long-distance dispersal, with the geogenetic positions of the source of gene flow coming from outside the geographic region in which individuals were located ~57% of the time. Thus, use of GBS data, which confers the advantage of a more representative sample of the entire genome, allows for a finer inference of evolutionary history (e.g., Excoffier, Dupanloup, Huerta-Sánchez, Sousa, & Foll, 2013; Fraser, McGaughran, Chuah, & Waters, 2016; Yang et al., 2017).

The high level of diversity observed within some sites, combined with genotypic differences among sites, suggest that multiple glacial refugia were available for terrestrial taxa along the Antarctic Peninsula to allow species persistence through time. These refugia may have been many and small, or few and large, but because numerous dispersal events are inferred to have occurred among these sites (and into newly-available ice-free regions), we cannot conclusively say which scenario is more likely. Other terrestrial Antarctic taxa show similar patterns of *in situ* survival through glacial periods, including mites (McGaughran, Hogg, & Stevens, 2008; van Vuuren, Lee, Convey, & Chown, 2018), nematodes (Courtright et al., 2000; McGaughran et al., 2008), mosses (Biersma et al., 2018; Hills, Stevens, & Gemmill, 2010; Pisa et al., 2014), and rotifers (Iakovenko et al., 2015), as do maritime Antarctic springtails (e.g., Carapelli et al., 2017).

In addition, such high diversity may indicate the presence of cryptic species. Indeed, our species delimitation analysis provided evidence for a possible cryptic species unique to Byers Peninsula (BYE; Livingston Island). We obtained ten specimens from this location and six of them together formed a separate species group in the delimitation analysis. These individuals also fell out as distinct in the mitochondrial haplotype network and are responsible for BYE having the highest measured nucleotide diversity among the populations for which we obtained mitochondrial

data. While there is precedent for sympatric cryptic lineages in Antarctic springtails (e.g., Fanciulli et al. 2001; Torricelli et al. 2010b), the output from our delimitation analysis should be considered a starting point rather than a firm conclusion, as we currently have no morphological, ecological, or geological data to support these findings and we did not have genomic data from the BYE site.

Surprisingly, despite challenging environmental conditions, Antarctic Peninsula springtails – which seem intrinsically poorly suited to dispersal due to their small size and extremely limited mobility – do not appear to have experienced geographic limitations to local and long-distance dispersal, nor have these dispersal events appeared to follow consistent routes. Our environmental analysis supports this finding, indicating an overall pattern of increased genetic diversity and differentiation in areas of high wind for both the mitochondrial and GBS data. The logical explanation for this is that high wind facilitates dispersal. Our GBS-based environmental models also show that rugosity and solar radiation are positive predictors of F_{ST}; this, along with the positive relationship between diversity and degree days inferred from the mitochondrial data, could indicate that diversity in springtails in this region is positively related to energy (temperature, wind, and solar radiation) and environmental heterogeneity. Likewise, in other research, energy availability has been inferred to be a primary covariate of species richness in Antarctic springtails (Baird, Janion-Scheepers, Stevens, Leihy, & Chown, 2019).

While the linear models we employed provide some preliminary insights into the environmental drivers of genetic diversity, there are limitations in the underlying data that mean the results should be interpreted with some caution. For example, the resolution of the environmental drivers (1 km) is unlikely to capture the finer scale microclimate or topography that is experienced by terrestrial invertebrates – further insights would likely be obtained with additional sampling sites and more site-specific environmental data. However, these analyses do provide preliminary empirical support for the role of wind in increasing Antarctic genetic diversity.

Although this is the first study, to our knowledge, to show correlations between wind strength and genetic diversity in Antarctica, previous studies have found evidence (e.g., presence of springtails in pitfall traps) for wind dispersal of soil invertebrate species in Antarctica areas, including Ross Island, the Antarctic Peninsula, and the Dry Valleys (McGaughran et al., 2011; Nkem et al., 2005; Hawes, Worland, Bale, & Convey, 2008; Vega, Convey, Hughes, & Olalla-Tárraga, 2019). Recent modelling of potential dispersal timescales for moss propagules (Biersma

et al., 2018) has indicated that considerable aerial distances (e.g., across the Drake Passage) can be travelled in as little as 24 h, while older ecophysiological studies of desiccation survival in Antarctic springtails including *C. a. antarcticus* suggest that survival in a humid air column at low temperature is plausible on such timescales (Worland & Convey, 2008; Worland, Leinaas, & Chown, 2006). Natural dispersal between sub-Antarctic islands has also been catalogued in a range of invertebrate and plant taxa (Fraser, Nikula, Spencer, & Waters, 2009; Greenslade, Farrow, & Smith, 1999; Mortimer et al., 2011; Nikula, Fraser, Spencer, & Waters, 2010) and springtails have been shown to successfully float on salt-water (Hawes et al., 2008; McGaughran et al., 2011). Thus, as well as aerial dispersal, in springtails and other terrestrial fauna, dispersal might occur through hitchhiking on floating material (e.g., buoyant macroalgae or driftwood) – which can be exacerbated by wind around Antarctica (Fraser et al. 2018) – or with migrating animals (e.g., seabirds) (Barnes et al., 2006; Gillespie et al., 2012).

While we note that the genetic links among populations that we have identified here could have resulted from the retainment of shared ancestral haplotypes rather than gene flow among close and distant populations, we would have expected such a scenario to have resulted in a stronger signal of regional population structure, as is seen in other endemic Antarctic springtails (Torricelli et al. 2010a; McGaughran, Stevens, Hogg, & Carapelli, 2011; Carapelli, Convey, Frati, Spinsanti, & Fanciulli, 2017). Elsewhere in Antarctica, and in other springtail species, previous single- or multi-locus genetic studies have indicated survival in long-term refugia and some movement among populations (see review: McGaughran et al., 2011). For example, in Victoria Land (continental Antarctica) springtails have a distribution that is largely constrained by distinct geographic breaks (Adams et al., 2006; Sinclair & Stevens, 2006; Collins et al. 2019). Barriers such as glacial ice tongues separate refugial relict populations that have recently expanded and stochastic colonisation events have resulted in genetic links between geographically-isolated populations (Bennett et al., 2016; McGaughran et al., 2011). In contrast, along the Antarctic Peninsula, there are no major single glacial barriers to north-south dispersal (McGaughran et al. 2010; Torricelli et al. 2010a), and no apparent diversity trends that correspond to a latitudinal gradient. Instead, the ecophysiological restrictions on movement of taxa among localised areas with differing amounts of ice-free habitat is most likely to underpin broad spatial patterns. Demographic hypothesis testing in an Approximate Bayesian Computation (ABC) framework could allow further exploration of this hypothesis.

After decades of ecophysiological research, we now have a good understanding of how some animals survive in the extreme Antarctic environment on a day-to-day basis. For example, the terrestrial microfauna can tolerate daily temperature fluctuations > 30°C and freeze/thaw events using physiological mechanisms such as those reliant on antifreeze proteins and supercooling (Block, Smith, & Kennedy, 2009; Convey, 1996; Everatt, Convey, Bale, Worland, & Hayward, 2015; Peck, Convey, & Barnes, 2006). Survival over longer time-spans is less well-characterised, though molecular and biogeographic data are enabling us to test hypotheses about the evolutionary history of different species. Next generation sequencing data will allow us to answer some of the remaining questions about the evolutionary history of Antarctic organisms, and this study is one of the first to carry out genomic analyses for an Antarctic species. Genome-wide SNP markers allowed us to deduce patterns of population genetic structure not previously detectable with single mitochondrial markers. While GBS data have some inherent problems, including being subject to missing data through allelic dropout (Arnold, Corbett-Detig, Hartl, & Bomblies, 2013; Gautier et al., 2013), this reduced representation sequencing method is revolutionising population genetic research. In our case, GBS data allowed us to infer dispersal events that were not as apparent using more traditional approaches.

One of the great challenges of carrying out genomic work on Antarctic invertebrates is that individuals are generally extremely small and often will not yield the quantities of DNA required by most genomic library preparation protocols. Indeed, this limitation affected our capacity to include more species in our analyses (*C. a. antarcticus*, a relatively large-bodied taxon, was the only Antarctic springtail species from which we could extract workable quantities of high molecular weight DNA), and is probably at the root of the relatively low read recovery and thus high levels of missing data in our SNP alignments. Future attempts to carry out phylogeographic genomic analyses of Antarctic invertebrate taxa could consider pooling population samples or amplifying DNA extracts prior to library preparation to overcome these hurdles. The improved insights provided by genomic analyses over traditional Sanger sequencing methods, shown here, highlight the importance of finding ways to carry out genomic research on Antarctic species to resolve long-standing evolutionary and biogeographic questions.

Acknowledgements

We wish to thank Payal Patel, Amanda Padovan, and Rachel Rathjen for laboratory assistance, four anonymous reviewers for helping us to improve the manuscript, Gideon Bradburd for assistance with the SpaceMix analyses, and the following funding sources: Australian Research Council (ARC) Discover Early Career Researcher Award (DECRA) DE160100685 to AM, Australian Antarctic Science Program (AAS 4296) funding to AT, ARC Future Fellowship FT170100281 and Rutherford Discovery Fellowship RDF-UOO1803 funding to CIF, National Environment Research Council (NERC) core funding to the British Antarctic Survey (BAS) 'Biodiversity, Evolution and Adaptation' Team for PC, and BAS logistic support to PC in making collections.

References

Adams, B. J., Bardgett, R. D., Ayres, E., Wall, D. H., Aislabie, J., Bamforth, S., … Stevens, M. I. (2006). Diversity and distribution of Victoria Land biota. *Soil Biology & Biochemistry*, 38(10), 3003–3018.

Arnold, B., Corbett-Detig, R. B., Hartl, D., & Bomblies, K. (2013). RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, 22(11), 3179–3190.

Baird, H. P., Janion-Scheepers, C., Stevens, M. I., Leihy, R. I., & Chown, S. L. (2019). The ecological biogeography of indigenous and introduced Antarctic springtails. *Journal of Biogeography*, https://doi.org/10.1111/jbi.13639.

Barker, P. F., & Camerlenghi, A. (2001). Glacial History of the Antarctic Peninsula from Pacific Margin Sediments. In P. F. Barker, A. Camerlenghi, G. D. Acton, A. T. S. Ramsay (Eds.), *Antarctic glacial history and sea-level change,* College Station, Texas, Ocean Drilling Program.

Barnes, D. K. A., Hodgson, D. A., Convey, P., Allen, C. S., & Clarke, A. (2006). Incursion and excursion of Antarctic biota: past, present and future. *Global Ecology and Biogeography: A Journal of Macroecology*, 15(2), 121–142.

Bennett, K. R., Hogg, I. D., Adams, B. J., & Hebert, P. D. N. (2016). High levels of intraspecific genetic divergences revealed for Antarctic springtails: evidence for small-scale isolation during Pleistocene glaciation. *Biological Journal of the Linnean Society*, 119(1), 166–178.

Biersma, E. M., Jackson, J. A., Bracegirdle, T. J., Griffiths, H., Linse, K., & Convey, P. (2018). Low genetic variation between South American and Antarctic populations of the bank-forming moss *Chorisodontium aciphyllum* (Dicranaceae). *Polar Biology*, 41(4), 599–610.

Biersma, E. M., Jackson, J. A., Stech, M., Griffiths, H., Linse, K., & Convey, P. (2018). Molecular data suggest long-term in situ Antarctic persistence within Antarctica's most speciose plant genus,

Schistidium. *Frontiers in Ecology and Evolution*, https://doi.org/10.3389/fevo.2018.00077.

Block, W., Smith, R. I. L., & Kennedy, A. D. (2009). Strategies of survival and resource exploitation in the Antarctic fellfield ecosystem. *Biological Reviews of the Cambridge Philosophical Society*, 84(3), 449–484.

Bradburd, G. S., Ralph, P. L., & Coop G. M. (2016). A spatial framework for understanding population structure and admixture. *PLoS Genetics*, 12(1): e1005703.

Brooks, S. T., Jabour, J., van den Hoff, J., & Bergstrom, D. M. (2019). Our footprint on Antarctica competes with nature for rare ice-free land. *Nature Sustainability*, 2, 185-190.

Burton-Johnson, A., Black, M., Fretwell, P. T., & Kaluza-Gilbert, J. (2016). An automated methodology for differentiating rock from snow, clouds and sea in Antarctica from Landsat 8 imagery: a new rock outcrop map and area estimation for the entire Antarctic continent. *The Cryosphere*, 10(4), 1665–1677.

Carapelli, A., Convey, P., Frati, F., Spinsanti, G., & Fanciulli, P. P. (2017). Population genetics of three sympatric springtail species (Hexapoda: Collembola) from the South Shetland Islands: evidence for a common biogeographic pattern. *Biological Journal of the Linnean Society*, 120(4), 788–803.

Chong, C.-W., Pearce, D. A., & Convey, P. (2015). Emerging spatial patterns in Antarctic prokaryotes. *Frontiers in Microbiology*, 6, 1058.

Chown, S. L., Clarke, A., Fraser, C. I., Cary, S. C., Moon, K. L., & McGeoch, M. A. (2015). The changing form of Antarctic biodiversity. *Nature*, 522 (7557), 431.

Chown, S. L., & Convey, P. (2016). Antarctic Entomology. *Annual Review of Entomology*, 61, 119–137.

Clarke, A., Barnes, D., & Hodgson, D. (2005). How isolated is Antarctica? *Trends in Ecology & Evolution*, 20(1), 1–3.

Collins, G. E., Hogg, I. D., Convey, P., Barnes, A. D., & McDonald, I. R. (2019). Spatial and temporal scales matter when assessing the species and genetic diversity of springtails (Collembola) in Antarctica. *Frontiers in Ecology and Evolution*, https://doi.org/10.3389/fevo.2019.00076.

Convey, P. (1996). The influence of environmental characteristics on life history attributes of Antarctic terrestrial biota. *Biological Reviews of the Cambridge Philosophical Society*, 71(2), 191–225.

Convey, P. (2017). Antarctic Ecosystems. In *Reference Module in Life Sciences*, Elsevier.

Convey, P., Gibson, J. A. E., Hillenbrand, C.-D., Hodgson, D. A., Pugh, P. J. A., Smellie, J. L., & Stevens, M. I. (2008). Antarctic terrestrial life – challenging the history of the frozen continent? *Biological Reviews of the Cambridge Philosophical Society*, 83(2), 103–117.

Convey, P., & Smith, R. I. L. (1997). The terrestrial arthropod fauna and its habitats in northern Marguerite Bay and Alexander Island, maritime Antarctic. *Antarctic Science*, 9, 12-26.

Convey, P., & Smith, R. I. L. (2006). Geothermal bryophyte habitats in the South Sandwich Islands, maritime Antarctic. *Journal of Vegetation Science*, 17(4), 529–538.

Convey, P., & Stevens, M. I. (2007). Ecology. Antarctic biodiversity. *Science*, 317(5846), 1877–1878.

Convey, P., Stevens, M. I., Hodgson, D. A., Smellie, J. L., Hillenbrand, C.-D., Barnes, D. K. A., … Craig Cary, S. (2009). Exploring biological constraints on the glacial history of Antarctica. *Quaternary Science Reviews*, 28(27-28), 3035–3048.

Coulson, S. J., Hodkinson, I. D., Webb, N. R., & Harrison, J. A. (2002). Survival of terrestrial soil-dwelling arthropods on and in seawater: implications for trans-oceanic dispersal. *Functional Ecology*, 16(3), 353–356.

Courtright, E. M., Wall, D. H., Virginia, R. A., Frisse, L. M., Vida, J. T., & Thomas, W. K. (2000). Nuclear and mitochondrial DNA sequence diversity in the Antarctic nematode *Scottnema lindsayae*. *Journal of Nematology*, 32(2), 143–153.

Curtis, A., & Kyle, P. (2011). Geothermal point sources identified in a fumarolic ice cave on Erebus volcano, Antarctica using fiber optic distributed temperature sensing. *Geophysical Research Letters*, 38(16). https://doi.org/10.1029/2011gl048272

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , 27(15), 2156–2158.

Dornelas, M., Magurran, A. E., Buckland, S. T., Chao, A., Chazdon, R. L., Colwell, R. K., … Vellend, M. (2013). Quantifying temporal change in biodiversity: challenges and opportunities. *Proceedings of the Royal Society B: Biological Sciences*, 280(1750), 20121931.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, 6(5), e19379.

Ersts, P. J. (n.d.). Geographic Distance Matrix Generator (version 1.2.3). Retrieved January 9, 2019, from http://biodiversityinformatics.amnh.org/open_source/gdmg/

Everatt, M. J., Convey, P., Bale, J. S., Roger Worland, M., & Hayward, S. A. L. (2015). Responses of invertebrates to temperature and water stress: A polar perspective. *Journal of Thermal Biology*, 54, 118–132.

Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M. (2013). Robust demographic inference from genomic and SNP data. *PLoS Genetics*, 9(10), e1003905.

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479–491.

Fanciulli, P. P., Summa, D., Dallai, R., & Frati, R. (2001). High levels of genetic variability and population

differentiation in *Gressittacantha terranova* (Collembola, Hexapoda) from Victoria Land, Antarctic. *Antarctic Science*, 13, 246-254.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294–299.

Fraser, C. I., McGaughran, A., Chuah, A., & Waters, J. M. (2016). The importance of replicating genomic analyses to verify phylogenetic signal for recently evolved lineages. *Molecular Ecology*, 25(15), 3683–3695.

Fraser, C. I., Morrison, A. K., Hogg, A. M., Macaya, E. C., van Sebille, E., Ryan, P. G., … Waters, J. M. (2018). Antarctica's ecological isolation will be broken by storm-driven dispersal and warming. *Nature Climate Change*, 8(8), 704–708.

Fraser, C. I., Nikula, R., Ruzzante, D. E., & Waters, J. M. (2012). Poleward bound: biological impacts of Southern Hemisphere glaciation. *Trends in Ecology & Evolution*, 27(8), 462–471.

Fraser, C. I., Nikula, R., Spencer, H. G., & Waters, J. M. (2009). Kelp genes reveal effects of subantarctic sea ice during the Last Glacial Maximum. *Proceedings of the National Academy of Sciences of the United States of America*, 106(9), 3249–3253.

Fraser, C. I., Terauds, A., Smellie, J., Convey, P., & Chown, S. L. (2014). Geothermal activity helps life survive glacial cycles. *Proceedings of the National Academy of Sciences of the United States of America*, 111(15), 5634–5639.

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915–925.

Gautier, M., Gharbi, K., Cezard, T., Foucaud, J., Kerdelhué, C., Pudlo, P., … Estoup, A. (2013). The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, 22(11), 3165–3178.

Gillespie, R. G., Baldwin, B. G., Waters, J. M., Fraser, C. I., Nikula, R., & Roderick, G. K. (2012). Long-distance dispersal: a framework for hypothesis testing. *Trends in Ecology & Evolution*, 27(1), 47–56.

Greenslade, P., Farrow, R. A., & Smith, J. M. B. (1999). Long distance migration of insects to a subantarctic island. *Journal of Biogeography*, 26(6), 1161–1167.

Hawes, T. C., Worland, M. R., Bale, J. S., & Convey, P. (2008). Rafting in Antarctic Collembola. *Journal of Zoology*, 274(1), 44-50.

Hayward, S. (2004). Habitat moisture availability and the local distribution of the Antarctic Collembola *Cryptopygus antarcticus* and *Friesea grisea*. *Soil Biology & Biochemistry*, 36(6), 927–934.

Hijmans, R. J., & van Etten, J. (2012). Raster: Geographic analysis and modelling with raster data. R package version 2.0-12. http://CRAN.R-project.org/package=raster.

Hills, S. F. K., Stevens, M. I., & Gemmill, C. E. C. (2010). Molecular support for Pleistocene persistence of

the continental Antarctic moss *Bryum argenteum*. *Antarctic Science*, 22(06), 721–726.

Iakovenko, N. S., Smykla, J., Convey, P., Kašparová, E., Kozeretska, I. A., Trokhymets, V., … Janko, K. (2015). Antarctic bdelloid rotifers: diversity, endemism and evolution. *Hydrobiologia*, 761(1), 5–43.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, 2, e281.

Kennedy, A. D. (1993). Water as a limiting factor in the Antarctic terrestrial environment: A biogeographical synthesis. *Arctic and Alpine Research*, 25(4), 308.

Knaus, B. J., & Grunwald, N. J. (2016). VcfR: an R package to manipulate and visualize VCF format data. *Molecular Ecology Res*ources*,* 17(1), 44-53.

Kumar, S., Stecher G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA 10.0.5: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547-1549.

Mackintosh, A. N., Verleyen, E., O'Brien, P. E., White, D. A., Selwyn Jones, R., McKay, R., … Masse, G. (2014). Retreat history of the East Antarctic Ice Sheet since the Last Glacial Maximum. *Quaternary Science Reviews*, 100, 10–30.

Magurran, A. E., & Dornelas, M. (2010). Biological diversity in a changing world. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 365(1558), 3593–3597.

McGaughran, A., Hogg, I. D., & Convey, P. (2011). Extended ecophysiological analysis of *Gomphiocephalus hodgsoni* (Collembola): flexibility in life history strategy and population response. *Polar Biology*, 34(11), 1713–1725.

McGaughran, A., Hogg, I. D., & Stevens, M. I. (2008). Patterns of population genetic structure for springtails and mites in southern Victoria Land, Antarctica. *Molecular Phylogenetics and Evolution*, 46(2), 606–618.

McGaughran, A., Stevens, M. I., Hogg, I. D., & Carapelli, A. (2011). Extreme glacial legacies: A synthesis of the Antarctic springtail phylogeographic record. *Insects*, 2(2), 62–82.

McGaughran, A., Torricelli, G., Carapelli, A., Frati, F., Stevens, M. I., Convey, P., & Hogg, I. D. (2010). Contrasting phylogeographical patterns for springtails reflect different evolutionary histories between the Antarctic Peninsula and continental Antarctica. *Journal of Biogeography*, 37(1), 103–119.

Morris, G. P., Grabowski, P. P., & Borevitz, J. O. (2011). Genomic diversity in switchgrass (*Panicum virgatum*): from the continental scale to a dune landscape. *Molecular Ecology*, 20(23), 4938–4952.

Mortimer, E., Jansen van Vuuren, B., Lee, J. E., Marshall, D. J., Convey, P., & Chown, S. L. (2011). Mite dispersal among the Southern Ocean Islands and Antarctica before the Last Glacial Maximum. *Proceedings of the Royal Society B: Biological Sciences*, 278(1709), 1247–1255.

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press.

Nguyen L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, 32(1), 268-274.

Nikula, R., Fraser, C. I., Spencer, H. G., & Waters, J. M. (2010). Circumpolar dispersal by rafting in two subantarctic kelp-dwelling crustaceans. *Marine Ecology Progress Series*, 405, 221–230.

Nkem, J. N., Wall, D. H., Virginia, R. A., Barrett, J. E., Broos, E. J., Porazinska, D. L., & Adams, B. J. (2005). Wind dispersal of soil invertebrates in the McMurdo Dry Valleys, Antarctica. *Polar Biology*, 29(4), 346–352.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419–420.

Peck, L. S., Convey, P., & Barnes, D. K. A. (2006). Environmental constraints on life histories in Antarctic ecosystems: tempos, timings and predictability. *Biological Reviews of the Cambridge Philosophical Society*, 81(1), 75–109.

Perrings, C., Duraiappah, A., Larigauderie, A., & Mooney, H. (2011). Ecology. The biodiversity and ecosystem services science-policy interface. *Science*, 331(6021), 1139–1140.

Pisa, S., Biersma, E. M., Convey, P., Patiño, J., Vanderpoorten, A., Werner, O., & Ros, R. M. (2014). The cosmopolitan moss *Bryum argenteum* in Antarctica: recent colonisation or in situ survival? *Polar Biology*, 37(10), 1469–1477.

Pugh, P. J. A., & Convey, P. (2008). Surviving out in the cold: Antarctic endemic invertebrates and their refugia. *Journal of Biogeography*, 35(12), 2176–2186.

Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, Automatic Barcode Gap Discovery for primary species delimitation. *Molecular Ecology*, 21(8), 1864-1877.

Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573–589.

R Core Team. (2019). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/

Sinclair, B. J., & Stevens, M. I. (2006). Terrestrial microarthropods of Victoria Land and Queen Maud Mountains, Antarctica: Implications of climate change. *Soil Biology & Biochemistry*, 38(10), 3158–3170.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, 195(3), 693–702.

Smellie, J. L., McIntosh, W. C., & Esser, R. (2006). Eruptive environment of volcanism on Brabant Island: Evidence for thin wet-based ice in northern Antarctic Peninsula during the Late Quaternary. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *231*(1-2), 233–252.

Soo, R. M., Wood, S. A., Grzymski, J. J., McDonald, I. R., & Cary, S. C. (2009). Microbial biodiversity of

thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica. *Environmental Microbiology*, 11(3), 715–728.

Stevens, M. I., Greenslade, P., Hogg, I. D., & Sunnucks, P. (2005). Southern Hemisphere springtails: could any have survived glaciation of Antarctica? *Molecular Biology and Evolution*, 23(5), 874-882.

Stevens, M. I., & Hogg, I. D. (2002). Expanded distributional records of Collembola and Acari in southern Victoria Land, Antarctica. *Pedobiologia*, 46(5), 485–495.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.

Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic association with haplotypes inferred from restriction site endonuclease mapping and DNA sequence data. (III) Cladogram estimation. *Genetics*, 132, 619-633.

Terauds, A., Chown, S. L., Morgan, F., Peat, H. J., Watts, D. J., Keys, H., … Bergstrom, D. M. (2012). Conservation biogeography of the Antarctic. *Diversity and Distributions*, 18(7), 726–741.

Torricelli, G., Frati, F., Convey, P., Telford, M., & Carapelli, A. (2010). Population structure of *Friesea grisea* (Collembola, Neanuridae) in the Antarctic Peninsula and Victoria Land: evidence for local genetic differentiation of pre-Pleistocene origin. *Antarctic Science*, 22(06), 757–765.

Torricelli, G., Carapelli, A., Convey, P., Nardi, F., Boore, J.L. & Frati, F. (2010). High divergence across the whole mitochondrial genome in the 'pan-Antarctic' springtail *Friesea grisea*: evidence for cryptic species? *Gene*, 449, 30–40.

van Vuuren, B. J., Lee, J. E., Convey, P., & Chown, S. L. (2018). Conservation implications of spatial genetic structure in two species of oribatid mites from the Antarctic Peninsula and the Scotia Arc. *Antarctic Science*, 30(02), 105–114.

Vega, G. C., Convey, P., Hughes, K. A., & Olalla-Tárraga. (2019). Humans and wind, shaping Antarctic soil arthropod biodiversity. *Insect Conservation and Diversity,* https://doi.org/10.1111/icad.12375.

Venables, W. N., & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

Weir, B. S., & Clark Cockerham, C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution,* 38(6), 1358–1370.

Worland, M. R., & Convey, P. (2008). The significance of the moult cycle to cold tolerance in the Antarctic collembolan *Cryptopygus antarcticus*. *Journal of Insect Physiology*, 54(8), 1281–1285.

Worland, M. R., Leinaas, H. P., & Chown, S. L. (2006). Supercooling point frequency distributions in Collembola are affected by moulting. *Functional Ecology*, 20(2), 323–329.

Yang, B., Cui, L., Perez-Enciso, M., Traspov, A., Crooijmans, R. P. M. A., Zinovieva, N., … Megens, H.-J. (2017). Genome-wide SNP data unveils the globalization of domesticated pigs. *Genetics, Selection, Evolution: GSE*, 49(1), 71.

Data Accessibility

Mitochondrial DNA sequences: Genbank accessions: MN473376-MN473393. We provide the full SNP data sets in VCF format as Supporting Information S4 (more conservative) and S5 (less conservative). The source data used to derive the environmental parameters in our spatial modelling is provided in Supporting Information S1 and can be made available upon request to aleks.terauds@gmail.com.

Author contributions

AM and CF conceived the ideas; CF and PC collected samples; AM performed the analyses; AT performed the environmental driver analysis; AM wrote the manuscript, with input from AT, PC and CF.

Tables

Table 1. Sampling information. Sampling information for all *Cryptopygus antarcticus antarcticus* samples from the Antarctic Peninsula used in this study. The table lists populations, population codes referred to in the text and Figure 1, and the number of samples in each population for both mitochondrial and genomic datasets.

Population	**Code**	**Date Collected**	**Number of samples**			**Latitude (S)**	**Longitude (W)**
			Mitochondria	*GBS pre-filtering*	*GBS post-filtering*		
Bernardo O'Higgins	BOH	Jan 2015	2	12	3	63°19'15.00"	57°53'55.00"
Coppermine Peninsula (Robert Island)	COP	Jan 2015	15*	30	16	62°22'45.63"	59°41'53.03"
Deception Island	DEC	Jan 2015	-	4	3	63°0'34.58"	60°35'58.66"
Byers Peninsula (Livingston Island)	BYE		10*			62º40'	61º13'
Hannah Point (Livingston Island)	HAN	Jan 2015	2	12	6	62°39'16.00"	60°36'48.00"
Harmony Point (Nelson Island	HAR		10*			62º19'	59º10'
Paulet Island	PAU		1*			63°35'	55°46'

Potter Cove (King George Island)	POT		8*			62°14'	58°39'
Berthelot Island	BER	Feb 2008	5	30	9	65°19'42.17"	64°8'35.80"
East Coast Lemaire	ECL		12*			64°49'	62°57'
Grand Island	GRA	Feb 2008	5	34	21	64°24'24.00"	62°48'36.00"
Paradise Harbour	PAR		12*			64°51'	62°54'
Petermann Island	PET		5*			65°10'	64°10'
Port Lockroy (Gouider Island)	PTL	Feb 2008	15*	30	11	64°49'3.87"	63°29'59.64"
Prospect Point (Graham Land)	PRO		11*			66°01'	65°49'
Spert Island	SPE	Feb 2008	5	30	16	63°50'42.00"	60°56'54.00"
Terrada Point	TER		10*			64°23'	62°14'
Alamode Island	ALA	Jan 2011	5	22	9	68°42'56.46"	67°32'2.65"
Anchorage	ANC	Jan 2015	5	30	16	67°36'1.94"	68°12'34.09"
Detaille Island	DET	Feb 2008	5	30	15	66°53'7.23"	66°41'4.44"W

Evensen Point	EVP	Feb 2008	6	30	14	66° 8'42.00"	65°43'24.00
Jenny Island	JEN	Feb 2008	5	30	2	67°43'54.07"	68°21'53.30"
Killingbeck Island	KIL		10*			67°34'	68°25'
Leonie Island	LEO	Jan 2015	5	30	17	67°35'59.52"	68°20'59.08"
Mackay Point (Adelaide Island)	MAC		11*			67°32'	68°05'
Reptile Ridge (Adelaide Island)	REP		10*			67°33'	68°98'
Rothera (Adelaide Island)	ROT		10*			67°34'	68°08'
Saunders Island (South Sandwich Islands)	SAU	Jan 1997	4	30	4	57°48'30.26"	26°25'4.28"
Total			**204**	**384**	**162**		

*Includes, or solely consists of, data from McGaughran et al. 2010

Table 2. Diversity statistics. Genetic diversity statistics for all *Cryptopygus antarcticus antarcticus* populations from the Antarctic Peninsula used in this study. Statistics were obtained from mitochondrial (cytochrome *c* oxidase subunit I) sequence data and are presented for each population in rough geographical order from the north of the Peninsula to the south. We were unable to obtain mitochondrial sequence data for the DEC location. For the Tajima's D and Fu's FS tests, S = significant (*P* between 0.05 and 0.00000) HS = highly significant ($P \leq 0.00000$). Numbers in parentheses represent standard deviations for those statistics. See Table 1 and Figure 1 for further population information.

Population code	**No. of sequences**	**No. of unique haplotypes**	**No. of polymorphic sites**	**Mean no. pairwise differences (k)**	**Nucleotide diversity (π)**	**Theta (S)**	**Tajima's D**	**Fu's FS**	**Mean F_{ST}**
SAU	4	1	16	8.000 (4.715)	0.015 (0.010)	8.727 (5.034)	-0.849^{S}	0.069	0.304 (0.220)
PAU	1	1	0	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000	-	0.733 (0.258)
BOH	2	2	10	10.000 (7.416)	0.018 (0.019)	10.000 (7.416)	0.000	2.303	0.507 (0.278)
POT	8	8	15	5.429 (2.924)	0.010 (0.006)	5.785 (2.822)	-0.315	-3.675^{S}	0.694 (0.104)
HAR	10	7	24	10.800 (5.374)	0.020 (0.011)	8.483 (3.759)	1.297	-3.264^{S}	0.335 (0.160)
HAN	2	0	1	1.000 (1.000)	0.002 (0.003)	1.000 (1.000)	0.000	0.000	0.589 (0.317)
COP	15	6	8	2.276 (1.322)	0.004 (0.003)	2.460 (1.202)	-0.276	-18.971^{HS}	0.686 (0.147)
BYE	10	3	48	23.444 (11.286)	0.043 (0.023)	16.967 (7.175)	1.891	-1.501	0.486 (0.094)
SPE	5	1	0	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000	>>>	0.770 (0.165)
GRA	5	0	1	0.600 (0.562)	0.001 (0.001)	0.480 (0.480)	1.225	-6.274^{HS}	0.514 (0.345)

TER	10	3	10	3.244 (1.825)	0.006 (0.004)	3.535 (1.750)	-0.364	-7.907^{HS}	0.446 (0.314)
PTL	15	6	6	3.171 (1.7370	0.006 (0.004)	1.845 (0.967)	2.516	-15.826^{HS}	0.520 (0.250)
ECL	12	2	11	2.394 (1.398)	0.004 (0.003)	3.643 (1.722)	-1.426	-12.895^{HS}	0.519 (0.284)
BER	5	0	0	0.300 (0.366)	0.001 (0.001)	0.000 (0.000)	0.000	-8.564^{HS}	0.578 (0.312)
PAR	12	4	10	2.970 (1.669)	0.005 (0.003)	3.311 (1.593)	-0.425	-11.383^{HS}	0.547 (0.245)
PRO	11	1	20	6.509 (3.335)	0.012 (0.007)	6.828 (3.018)	-0.212	-5.820^{S}	0.502 (0.265)
PET	5	3	15	8.600 (4.795)	0.016 (0.010)	7.200 (3.925)	1.422	-0.542	0.377 (0.217)
EVP	6	0	0	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000	>>>	0.791 (0.244)
DET	5	0	0	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000	>>>	0.720 (0.290)
ANC	5	1	15	4.600 (2.713)	0.009 (0.006)	7.200 (3.925)	-2.640^{HS}	-1.481	0.462 (0.216)
LEO	5	4	12	4.100 (2.452)	0.008 (0.005)	5.760 (3.206)	-2.084^{HS}	-1.674	0.418 (0.304)
JEN	5	2	9	2.500 (1.610)	0.005 (0.003)	4.320 (2.484)	-2.699^{HS}	-2.596^{S}	0.356 (0.326)
KIL	10	3	24	11.556 (5.728)	0.021 (0.012)	8.484 (3.759)	1.720	-3.079^{S}	0.307 (0.198)
MAC	11	6	11	4.182 (2.250)	0.008 (0.008)	3.756 (1.801)	0.488	-7915^{S}	0.449 (0.259)
REP	10	1	21	4.778 (2.613)	0.009 (0.005)	6.363 (2.901)	-1.065	-5.991^{S}	0.414 (0.275)

ROT	10	4	18	4.911 (2.800)	0.009 (0.005)	7.070 (3.188)	-1.169	-5.668[S]	0.456 (0.191)
ALA	5	1	5	1.400 (1.019)	0.003 (0.003)	2.400 (1.513)	-2.810[HS]	-3.901[HS]	0.638 (0.279)
Total	**204**								

Table 3. Species delimitation analysis. Number of groups identified for a given prior intraspecific divergence (*P*) in Automatic Barcode Gap Discovery (ABGD) analysis, based on different substitution models (Simple, JC = Jukes Cantor JC69, K2P = Kimura K80). In each case, the analysis converges (agreement between initial and recursive partitions) at a final species group size of 2 (with six divergent samples from Byers Peninsula, BYE, included) or 1 (with the divergent BYE samples removed). See Materials and Methods for further information.

Prior intraspecific divergence (*P*)	Substitution model – with BYE			Subsitution model – without BYE		
	Simple	JC	K2P	Simple	JC	K2P
0.0010	3	33	33	2	31	31
0.0017	3	33	33	2	31	31
0.0028	2	2	2	1	1	1
0.0046	2	2	2			
0.0077	2	2	2			
0.0129	2	2	2			
0.0215	2	2	2			

Table 4. Genetic differentiation among populations. Genetic differentiation among *Cryptopygus antarcticus antarcticus* populations from the Antarctic Peninsula. Measures of genetic differentiation are shown for genomic data, and values are presented for each population in rough geographical order from the north of the Peninsula to the south. Numbers in parentheses represent standard deviations. See Table 1 and Figure 1 for further population information and Table 2 for mitochondrial F_{ST} measures.

	GBS	
Population	**Mean Nei's D**	**Mean F_{ST}**
SAU	0.245 (0.04671)	0.222 (0.148)
BOH	0.194 (0.06030)	0.196 (0.161)
HAN	0.185 (0.05074)	0.195 (0.124)
COP	0.147 (0.03817)	0.148 (0.075)
DEC	0.167 (0.03990)	0.089 (0.081)
SPE	0.147 (0.05814)	0.296 (0.093)
GRA	0.120 (0.03648)	0.080 (0.062)
PTL	0.162 (0.04673)	0.179 (0.088)
BER	0.166 (0.04788)	0.178 (0.075)
EVP	0.134 (0.04048)	0.179 (0.064)
DET	0.188 (0.05218)	0.192 (0.068)
ANC	0.137 (0.05011)	0.147 (0.084)
LEO	0.138 (0.03912)	0.093 (0.058)
JEN	0.186 (0.06029)	0.151 (0.121)
ALA	0.186 (0.07147)	0.279 (0.132_

Figures Legends

Figure 1. Sampling locations. Map of sampling locations for all *Cryptopygus antarcticus antarcticus* populations from the Antarctic Peninsula used in this study. Populations new to the current study are indicated with red dots and white text; those from McGaughran et al. (2010) are indicated with black dots and light yellow text.

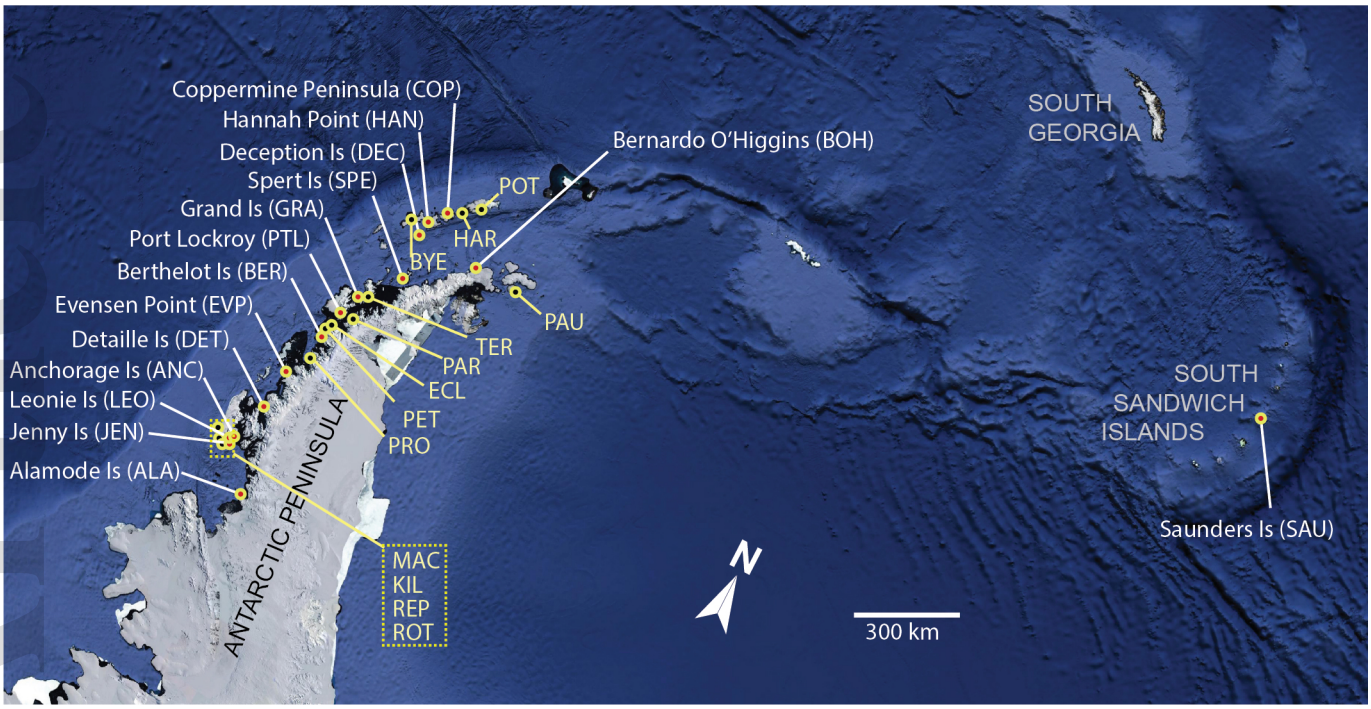
Figure 2. Haplotype network showing haplotypes shared among, or unique to, regions. Genetic relationships among *Cryptopygus antarcticus antarcticus* samples and populations from the Antarctic Peninsula based on mitochondrial data. Population codes (see Table 1, Figure 1) are indicated to the left in each panel. Each haplotype is represented as a coloured circle whose size is proportional to its frequency, while the length of mutational steps between haplotypes (indicated by the line that connects each haplotype) are not to scale. Divergent samples from Byers Peninsula (BYE) are indicated by a black box.

Figure 3. Principal Components Analysis supports gene flow among regions. Orthogonal transformation of genomic data from Antarctic Peninsula springtails (*Cryptopygus antarcticus antarcticus*), to show genetic relationships among individuals and geographic regions on the Antarctic Peninsula ("NOR" = northern, "CEN" = central, "STH" = southern; see Table 1, Figure 1). The percentage of variation explained by each principal component is indicated on the axes.

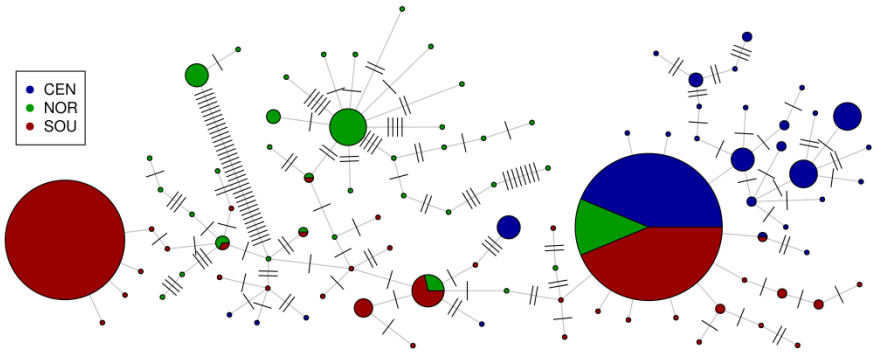
Figure 4. fastSTRUCTURE analysis indicates some regional differences, as well as admixture among distant sites. Visualisation of population structure of Antarctic Peninsula springtails (*Cryptopygus antarcticus antarcticus*) using fastSTRUCTURE software. Results are presented for K-values of 2 and 3, with colours in each panel representing the assignment of individuals to each genetic cluster, and each horizontal bar of the plot representing an individual sample. Cases where a single bar has > 1 colour indicate admixture of that individual (i.e., the sharing of genetic structure across more than one genetic group/cluster). Population locations and the broader 'northern', 'central', and 'southern' Antarctic Peninsula regions referred to in the text are indicated on the map on the right, which also identifies the frequency of mitochondrial haplotypes in each population. Haplotype pie charts are coloured according to the sharing of haplotypes among regions (white = unshared haplotypes unique to each population, shades of green, blue, and

red indicated sharing within/among northern, central, and southern Peninsula regions, respectively). See Supporting Information S3 for plots for additional K-values.

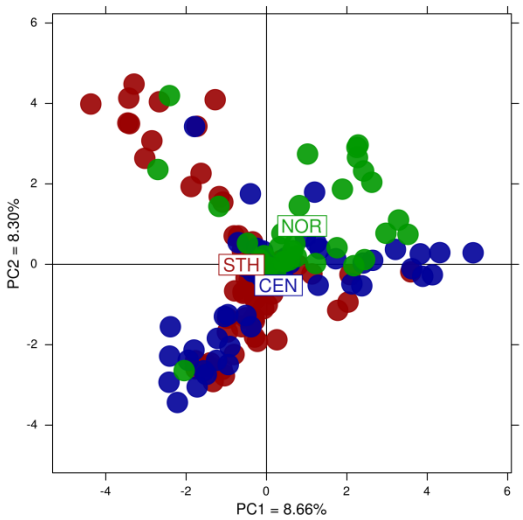
Figure 5. SpaceMix analyses support long-distance dispersal. Overall, there are three main genetic clusters (as indicated by shades of similar colour) and an example of long distance dispersal is shown – LEO.17 in bold type reflects that samples' geogenetic position (95% confidence interval is indicated with a solid red ellipse), LEO.17 in italics reflects the geogenetic position of the proposed source of admixture into LEO.17 (95% confidence indicated with the dashed black ellipse). Axes represent degrees of Southern latitude (vertical) and Western longitude (horizontal). See Supporting Information S3.5 for examples of local and neighbouring dispersal.



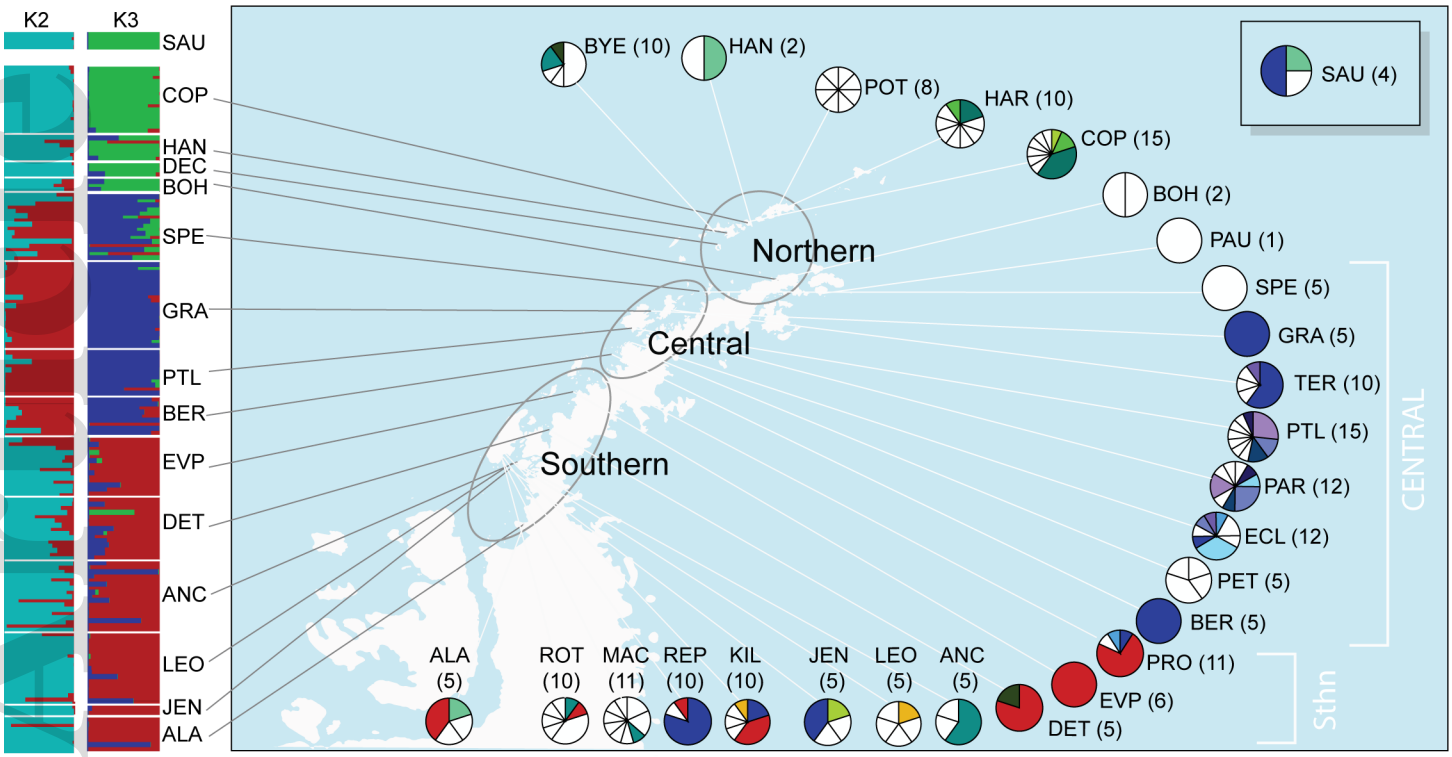
mec_15269_f1.tif



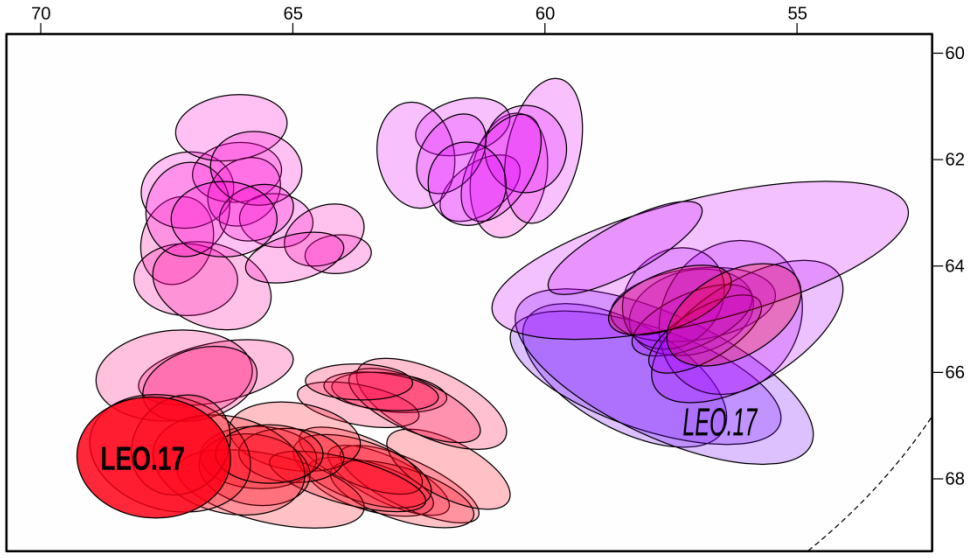
mec_15269_f2.tif



mec_15269_f3.tif



mec_15269_f4.tif



mec_15269_f5.tif