**Population genetics of three sympatric springtail species (Hexapoda, Collembola) from the South Shetland Islands: evidence for a common biogeographic pattern**

**Antonio Carapelli1\*, Peter Convey2, Francesco Frati1,Giacomo Spinsanti1, Pietro Paolo Fanciulli1**

**1** Department of Life Sciences, University of Siena, Via Aldo Moro 2, 53100 Siena, Italy.

**2** British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK.

\*Corresponding author

Email: antonio.carapelli@unisi.it

Running title: Biogeography of South Shetlands springtails

**ABSTRACT**

Three sympatric springtail species, from the South Shetland Islands archipelago in the maritime Antarctic, are analyzed here in a common biogeographic and evolutionary framework. This study was designed to compare their population genetic structure using the same molecular marker. Haplotype data for the mitochondrial *cox1* gene have been obtained for seven populations of *Folsomotoma octooculata* and are compared with data obtained, in previous studies and the current one, for the sympatric species *Cryptopygus antarcticus antarcticus* and *Friesea grisea*. Molecular data obtained are compatible with the hypothesis that all species were present in the archipelago since well before the last glacial maximum (around 20,000 ybp), and that their early diversifications appear to be linked with known interglacial periods in the region. These springtails may have survived the last glacial cycle in local refugia, from which they dispersed subsequently to ice-free ground re-exposed during the current interglacial. The populations of the different species diversified at different times, although all of them within the Pleistocene epoch. We propose that the earliest diversification of haplotypes in all three springtails in this archipelago occurred xxxxxx, with subsequent spread of some haplotypes throughout the South Shetlands Islands.

ADDITIONAL KEYWORDS: Antarctica – Collembola­ – biogeography - evolutionary origin.

**INTRODUCTION**

The contemporary terrestrial biota of Antarctica is predominantly restricted to ice-free low altitude habitats of coastal regions, that collectively represent less than0.25% of the continental surface. Many taxa are unique to one of the two major geological elements of the Antarctic continent, in biogeographical terms often known as the maritime and continental Antarctic (Convey, 2013), which are broadly equivalent to West (including the Antarctic Peninsula) and East Antarctica (see inset to Fig. 1). The evolutionary history of this terrestrial biota, that is essentially limited to a few invertebrate groups, has been strongly influenced by multiple glacial events, the last of which occurred during the late Pleistocene. Ice-sheet reconstruction models suggest that terrestrial habitats within the continent were over-run by considerably thicker and more extensive ice sheets during the Last Glacial Maxima (LGM) (Convey et al., 2009) and, by implication. that previous terrestrial terrestrial diversity should have been completely wiped out and eventually replaced after subsequent ice retreat by colonists from refugial areas beyond the continent itself.

This simplified biogeographic pattern is challenged by geographic evidence. Antarctica is surrounded by the Southern Ocean and isolated from the other southern continents by 1-5,000 km. The combination of the powerful Antarctic Circumpolar Current and atmospheric circulations surrounding the continent effectively isolates it and has severely restricteds the southwards dispersal of colonists towards Antarctica both today and during Pleistocene and previous interglacials (Clarke *et al*., 2005; Barnes *et al*., 2006; Fraser *et al.,* 2012). Southward recolonization of the continent was and is particularly difficult for polar terrestrial invertebrates, given their typical lack of dispersal adaptations in their adversity-selected life history strategies (Convey, 1996), and the lack of intermediate ‘stepping stones’ that might reduce the distances required to be covered in dispersal events. Furthermore, recent and rapid increases in knowledge of the evolutionary history of Antarctic terrestrial invertebrates, obtained through molecular clock and phylogeographic studies, cannot be reconciled with a hypothesis of recent colonization of the continent (Convey et al., 2008, 2009; Fraser et al., 2012). Estimates of biogeographic and demographic expansion clearly support ancient diversification for most lineages, certainly predating the LGM and in many cases far longer. Molecular data supporting an ancient origin of Antarctic terrestrial biota, have been interpreted to suggest that at least some ice-free terrains compatible with the development of terrestrial ecosystems have persisted throughout successive ice ages.

Much of Antarctica’s present-day terrestrial invertebrate biodiversity is endemic to the Antarctic continent, and indeed to different distinct regions within in (Adams *et al*., 2006;Pugh & Convey, 2008; Terauds *et al*., 2012), in many cases not sharing close or recent relatives with other higher latitude regions of the other southern continents. Reconciliation of the contrasting inferences from biogeographical and geomorphological/glaciological modelling studies is yet to be achieved, although this is part of function of the different spatial scales currently involved in the description of terrestrial ecosystems and the development of glaciological models. In some, but not all, regions of the Antarctic the existence of geothermal refugia may have contributed to the survival of terrestrial biota through glacial periods (Convey & Smith, 2006; Fraser *et al.,* 2012; Fraser *et al.,* 2014), with subsequent recolonization of areas re-exposed during glacial retreat from these ‘local’ refugia.

The contemporary Antarctic terrestrial fauna comprises a limited number of invertebrates, with arthropods represented by mites and springtails, and only two species of ‘higher insects’ (Convey, 2013; Chown & Convey, 2016). Their distributions are generally limited to coastal regions where seasonally ice- and snow-free habitats are best developed and liquid water is available, including the best-developed vegetated habitats in Antarctica. A more limited number of species – again often endemic to particular regions – are found at inland ice-free locations such as nunataks, and the more extensive ice-free areas of the Dry Valleys of Victoria Land and in the Transantarctic Mountains (Adams et al., 2006; Stevens & Hogg 200X). Antarctic terrestrial invertebrates have low dispersal ability through their adversity selected life history strategies, as well as possible evolutionary selection against dispersal given their typically very isolated and ‘island-like’ distributions (Convey, 1996; see also discussion of dispersal in Chown & Convey, 2016).

The number of collembolan (springtail) lineages presently occurring in Antarctica is small in comparison with ecosystems at lower latitudes. Among the 30 described families of Collembolaonly four, all belonging to the suborder Arthropleona, have Antarctic representatives, with no records of taxa from the two remaining high-rank lineages, Neelipleona and Symphypleona. The distribution of the over 20 known contemporary Antarctic springtail species (Greenslade, 1995, 2010) is almost equally divided between the Maritime and Continental Antarctic. All these species, except that currently classified as *Friesea grisea* (but see Torricelli *et al.,* 2010a,b for evidence of deep molecular differentiation), inhabit one or the other, but not both, sides of an ancient biogeographic boundary known as the Gressitt Line, located at the base of the Antarctic Peninsula, and which separates not only the springtails but also representatives of most of the other major contemporary terrestrial invertebrate groups of Antarctica (Chown & Convey, 2007) (Fig. 1). Previous studies on Antarctic springtail populations, irrespective of their biogeographical location within the continent, have shown high levels of genetic diversity, strong genetic structure and fragmentation (Stevens *et al.,* 2006; McGaughran *et al.,* 2010, 2011). Most populations studied have unique haplotypes, with limited haplotype sharing between them. Genetic data also suggest that springtail populations have persisted in Antarctica in refugial sites which have remained ice-free, in some cases, for millions of years. Population genetic parameters are consistent with this view indicating differentiation and demographic expansion within a timescale that pre-date the last glacial maxima.

In the maritime Antarctic (encompassing the western side of the Antarctic Peninsula, its offshore islands, and the associated South Shetland, South Orkney and South Sandwich archipelagoes of the Scotia Arc) the five most common collembolan species, *Archisotoma brucei* (Isotomidae), *Cryptopygus antarcticus antarcticus* (Isotomidae), *Folsomotoma octooculata* (Isotomidae), *Friesea grisea* (Neanuridae) and *Tullbergia mixta* (Tullbergiidae), occur in vegetated or ornithogenic soils of coastal ice-free habitats, and display subtly different adaptations to cold and desiccation stress (Hayward *et al.,* 2004; Russell *et al.,* 2014). Among them, the three isotomid species appear to be more active soil-dwellers, at least during the austral summer. Although each of them has its own distinctive distribution pattern, most of the species known to occur in the maritime Antarctic are present in the South Shetland Islands archipelago, a group of islands (some of geologically recent volcanic origin) located north-west of the Antarctic Peninsula (Greenslade, 2010) (Fig. 1).

*Folsomotoma octooculata* is considered a native species of the South Shetland Islands (*sensu* Greenslade & Convey, 2012), and has congeneric counterparts in South America and Australia. Studies to date have primarily focused on its morphological features. Its taxonomic status has been updated several times in the recent past (Greenslade, 1986, 1995, 2010; Potapov, 2001)*.* Despite these taxonomic developments, a biogeographic study has never been attempted, unlike some other Antarctic springtails (Stevens *et al.,* 2006a; Hawes, Torricelli & Stevens,2010; McGaughran *et al.,* 2010). Here, in order to evaluate if different species inhabiting the same biogeographic region may have undergone similar evolutionary trajectories, we focused on three springtail species naturally resident in the South Shetlands Islands. We used a molecular framework to address the evolutionary history of *F. octooculata* in this archipelago, comparing these data with existing and new data for the Antarctic collembolans *Cryptopygus antarcticus antarcticus* and *Friesea grisea*. The primary aim of this study was to test which of the two competing hypotheses (recent colonization or dispersal from local glacial refugia) best explains the observed genetic structure.

**MATERIALS AND METHODS**

*Data collection*

Samples of *F. octooculata* (Willem, 1901) from the South Shetland Islands archipelago were collected from seven locations during a 2002-03 expedition involving a collaboration between the British Antarctic Survey (BAS) and the Italian National Antarctic Program (PNRA) (Fig. 1). The specimens were identified under a stereomicroscope, frozen and preserved at -80°C until their use for molecular analyses, as described in Appendix S1. The selected marker is the almost complete sequence of the mitochondrial gene encoding for the cytochrome *c* oxidase subunit I (*cox1*).

In order to obtain a direct comparison among species that inhabit the same Antarctic localities, a parallel genetic analysis was run for two additional collembolan species that live in sympatry with *F. octooculata*, with sequences of *C. a. antarcticus* and *F. grisea* also being analyzed for the same genetic marker. The *F. grisea* dataset reported by Torricelli *et al.* (2010a) was enlarged to include the same group of populations and the same number (10) of specimens obtained for *F. octooculata*, adding new sequences and new populations (PCK and RPN: see Table 1 for abbreviations). Methods used for DNA extraction, amplification and sequencing are as described in Torricelli *et al.* (2010a). Similarly, for *C. a. antarcticus* part of the data set analyzed for this species by McGaughran *et al.* (2010), limited to the South Shetland Islands localities, was augmented with two new sequences from PCK (obtained using the same methods as described by McGaughran *et al.*, 2010). The final dataset therefore included 10 specimens for each of the four compared populations. No samples were available for HAL and HPL, although all four individual islands under study were represented for this taxon.

*Data set assembly*

Ten specimens of *F. octooculata* from each of the seven sampling localities (Fig. 1, Table 1) were sequenced for 1533 nucleotides of *cox1* (between positions 1422-2954 of the *F. octooculata* mtDNA). The selected fragment includes all the codon positions of *cox1* (except for the last two triplets and the stop codon) and also includes the last six nucleotides at the 3’-end of *trnY* (the gene that foreruns the 5’-end of *cox1*, along the J-strand of mtDNA).

The 70 sequences of *F. octooculata* were manually aligned with MacClade 4.08 (Maddison & Maddison, 2005), resulting in a 1533-bp matrix, with no indels. The 70 sequences of *F. grisea* and the 40 of *C. a. antarcticus* were also aligned, resulting in matrices of 478- and 618-bp, respectively.

Frequencies of haplotypes for all species (Table 1) were obtained using the online tool DNAcollapser (<http://users-birc.au.dk/biopv/php/fabox/software>), and used for the network clade analysis using TCS 1.21 (Clement, Posada & Crandall, 2000), with the connection limit set to 95% (Fig. 1A-C). Haplotype nomenclature for *C. a. antarcticus* and *F. grisea* follows that used in McGaughran *et al.* (2010) and Torricelli *et al.* (2010a), respectively.

Analyses of demographic history of the *F. octooculata* populations, as well as those of *C. a. antarcticus* and *F. grisea*, were performed with Arlequin 3.11 (Excoffier, Laval & Schneider, 2005), through mismatch distribution (MMD) analysis (16,000 iterations), and using the distribution of observed and simulated pairwise differences among haplotypes within each population (Table 3). The time of expansion (*t*) for each population was calculated using the population demographic parameter tau (*τ*), and applying the formula *t* = *τ*/2μ, where μ is the mutation rate per locus per generation (Rogers & Harpending, 1992), and assuming a divergence rate of 1.5-2.3% Myr-1 (Brower, 1994) and a generation time of 3 y (McGaughran, Hogg & Stevens*,* 2008). Parametric bootstrapping was then used to estimate signatures of demographic expansion and the time at which it occurred, using the population demographic parameters tau (*τ*) and theta (Θ; with Θ0 at pre- and Θ1 at post-expansion population sizes, respectively). Tau measures the time of the expansion in units of mutational time (with confidence interval = 95%).

Sum of squared deviations (*SSD*) (Rogers & Harpending, 1992) between observed and expected mismatch patterns, as well as Raggedness (*R*) index, were used to test the model of demographic expansion, assessing the fit of the observed distribution with population expansions chosen as the null hypothesis (Harpending, 1994) (Table 3, Appendix S2). Mismatch analysis describes the distribution of the observed number of differences between pairs of haplotypes, compared with that expected under a sudden expansion model (Schneider & Excoffier, 1999). In populations at demographic equilibrium the distribution in the plot is usually multimodal, whereas in those that have recently expanded the distribution is unimodal (Rogers & Harpending, 1992).

Genetic divergence among haplotypes for all species under study, using absolute and pairwise distance methods, was assessed with PAUP\* (version 4b10-x86-macosx) (Swofford, 2003), and FST genetic distances were calculated using Arlequin 3.11 (Excoffier *et al.,* 2005). Matrices of geographical distances among samples (obtained with the program Geographic Distance Matrix Generator; version 1.2.3) and FST distances were compared (through 16,000 permutations) using Arlequin 3.11 (Excoffier *et al.,* 2005), to assess the significance of the correlated values, as implemented in the Mantel test (Mantel, 1967). The same program was also used to estimate haplotype (*h*) and nucleotide (*π*) diversity indices (Nei 1987) and to run “neutrality tests” among populations, applying Tajima’s *D* (Tajima, 1989) and Fu’s *FS* (Fu, 1997) parameters, with the significance of the values evaluated over 16,000 permutations. Tajima’s *D* statistics compare the mean pairwise differences and the number of segregating sites in a sample. Negative values are expected for population size expansion and/or purifying selection, whereas positive values are interpreted as signatures of balancing selection and population contraction. Fu’s test is applied to detect departure from population equilibrium (i.e. population expansion). A negative value of *FS* is evidence for an excess number of alleles, as would be expected from a recent population expansion, whereas a positive value (deficiency of alleles) is expected after a recent population bottleneck. AMOVA analysis was performed with Arlequin (Excoffier *et al.,* 2005) to produce estimates of variance between haplotypes at different hierarchical levels. The seven *F. octooculata* populations were therefore clustered into: a) four groups, each corresponding to the sampled island; b) two groups, associating haplotypes obtained from King George I. and Nelson I. (KN group) and from Livingston I. and Robert I. (LR group). In addition, for all species the components of differentiation were calculated with no structure enforced. Hierarchical clustering of haplotypes was performed using BAPS 6.0 to assess dependence between unlinked markers under the Bayesian model of clustering method (Cheng *et al.,* 2013). This latter analysis was also applied to the *C. a. antarcticus* and *F. grisea* datasets.

*Phylogenetic analysis*

Haplotype sequences of *cox1* of the three springtail species (Tables 1 and 2) and of the outgroup species *Onychiurus orientalis* (Cook, Yue & Akam, 2005) were manually aligned (resulting in a matrix of 1554 bp) and used for the phylogenetic analysis. Aligned nucleotides were partitioned in three groups according to their codon position (1st, 2nd and 3rd), and examined for the best partitioning strategy and evolutionary model for each partition, as implemented in PartitionFinder 1.0.1 (Lanfear *et al.,* 2012). The resulting partitioning scheme and evolutionary models (1st= GTR+Γ; 2nd= HKY+I; 3rd= HKY+ Γ) were applied in a bayesian analysis using MrBayes 3.2.1 (Ronquist & Huelsenbeck, 2003). Two parallel runs, each consisting of four chains, were run for 5 million generations, sampling every 1000th generation and removing 20% as burnin, upon stationarity of log-likelihood values. The final consensus tree was used to define the genetic relationships among haplotypes and to visualize the different patterns obtained with the clustering analysis (Fig. 2).

*Abbreviations*

AMOVA, Analysis of molecular variance; Γ, Gamma; GTR, General Time Reversible; HKY, Hasegawa Kishino Yano; *h*, haplotype diversity; I, Invariant; LGM, Last Glacial Maximum; Myr, million years ago; *π*, nucleotide diversity; PCR, Polymerase Chain Reaction; ybp, years before present.

**RESULTS**

*Haplotype composition and diversity indices in* F. octooculata

Screening of the 70 sequences for *F. octooculata* resulted in 17 haplotypes (Table 1) that differed in a total of 22 variable sites (Appendix S2). Most haplotypes occurred at low frequency: 12 were represented by only one individual and 15 were unique to a single population, while only two (A and J) were found in more than one site (Table 1). Haplotype A was present at all locations, while J occurred only in Livingston and Robert Islands (the two southernmost islands of the four investigated) (Fig. 1A). Most of the low frequency haplotypes differed from A or J by a single nucleotide substitution. Haplotype P was the most divergent, with 4 and 5 nucleotide differences compared with J and A, respectively (Appendix S2). The largest number of haplotypes (6) was observed in the Nelson I. population (HPN), whereas that of King George I. (PCK) only hosted individuals with haplotype A. Collectively, the two Nelson Island populations (HPN and RPN) also had the greatest number (8) of the 15 unique haplotypes identified. Nucleotide substitutions among the 70 examined sequences occurred in all three codon positions of *cox1*. The single nucleotide substitutions that differentiate haplotypes C and D from A (at aligned positions 1010 and 70, respectively) led to an amino acid change (Appendix S2).

Haplotype diversity values were high in four populations (DPL, HPL, HPN and RPN) and lower for the remaining three (CPR, HAL and PCK) (Table 1). Conversely, nucleotide diversity was low (*π*<0.0008) for all populations (Table 1).

*Haplotype composition and diversity indices in* F.grisea *and* C. a. antarcticus

In *F. grisea*, 30 new individuals were sequenced, resulting in 3 P7 haplotypes from DPL, and 27 P3 haplotypes from HAL (3), HPN (3), HPL (1), PCK (10) and RPN (10).

In *C. a. antarcticus* the two new individuals sequenced from PCK had haplotypes H20 and H31.

Overall, seven haplotypes were present (Table 1) in *F. grisea*, with two being very common (P3=55 and P7=10) and five unique (Fig. 1B). P3 was found at all sampled sites with the exception of DPL, whereas the remaining haplotypes were each unique to one location. The DPL population, from the southern-most Livingston Island, contained only P7, and is the only one where the most frequent haplotype was not represented, suggesting genetic divergence from the most common genetic pool. In both *F. octooculata* and *F. grisea*, at the Devils Point site (Livingston Island) the overall most frequent haplotype was absent or present at a very low frequency (Tables 1 & 2; Fig. 1A-B). In all three populations of *F. grisea* from Livingston Island five haplotypes were present, including the most frequent overall (P3) and the most abundant of the six unique haplotypes (P7) (Table 1). In *F. grisea*, *h* values were lower than in *F. octooculata*, while values of *π* were similar between the two species, except for those obtained from the southern-most HAL site (Table 1), where the *π* value was the highest amongst all sampling sites, due to the highly divergent haplotype P6 (9 to 11 substitutions). A large number of haplotypes (28) was observed for *C. a. antarcticus*, with the number of haplotypes per population ranging between 5 and 9. Only two haplotypes (H15 and H20) were shared between populations of the “central” Nelson and Robert Islands (both between CPR and HPN), and these were represented by a limited number (four and two, respectively) of sequences (Table 1). The haplotype network identifies three clusters (H29+H30, H6-H8+H10, all the remaining haplotypes, respectively; Fig. 1C). Consequently, high values of *h* and *π* (Table 1) were found in these populations.

*Genetic distances*

The matrix of pairwise genetic distances between haplotypes highlighted substantial uniformity within populations of *F. octooculata*, which differed on average by 0.24% and at maximum by 0.52%. The latter corresponds to 8 changes, and is observed when haplotype P is compared with L, N and Q. Conversely, the lowest estimate (0.06% = one substitution) was obtained when the following haplotypes were compared: A *vs* B-K, O and M; J *vs* Q, N and L. Values of *p*-distance were considerably higher when haplotypes of *F. octooculata* were compared with the other collembolan species *C. antarcticus antarcticus* (specimens sampled in Killingbeck I., Antarctic Peninsula: S 67°C 32’; W 68° 07’) (Carapelli *et al.,* 2008) and *O. villosa* (Carapelli *et al.,* 2007)*,* ranging from 20% to 21%. The genetic distances calculated between the seven *F. grisea* haplotypes ranged from 0.21% to 2.30% (corresponding to 1 and 11 nucleotide changes, respectively), giving an average 0.91% divergence. Most of the variability was generated when P6 was compared with the other haplotypes. If P6 is excluded from the comparison, the average *p*-distance value drops to 0.40%.

In *C. a. antarcticus* ranges of nucleotide substitutions (from 1 to 51) and distance estimates among haplotypes were substantially larger than in the other two springtail species, with the highest value (8.25%) observed when H18 was compared with H25. Average *p*-distances calculated between haplotypes in the present data set (2.6%) were similar to that obtained (2.5%) in a previous analysis based on 14 *cox1* sequences from three Antarctic Peninsula sites (Stevens *et al.,* 2003). However, it should be noted that evaluation of genetic distances calculated for the three different species, although based on the same mitochondrial gene, are not completely comparable, due to the different size of the analyzed fragments. The proportion of nucleotide substitutions (*p*-distance) estimated in *F. octooculata* is based on a longer fragment of *cox1* than that used in *C. a. antarcticus* and *F. grisea*. An overall 59% of the genetic variability of the *cox1* dataset observed in *F. octooculata* (13 out of 22 nucleotide changes) was distributed in the aligned fragment shared between this species and *C. a. antarcticus*.

*Mantel test*

The Mantel test showed a significant correlation between genetic and geographic distances in the *F. octooculata* and *F. grisea* populations studied (*r* = 0.487, p = 0.019 and *r* = 0.434, p = 0.012, respectively), indicating that neighboring populations were genetically more similar than expected by chance. In contrast, in *C. a. antarcticus* the Mantel test rejected the hypothesis of correlation between genetic and geographical data between the four populations (*r* = 0.603, p = 0.074).

*Analysis of mismatch distributions*

Mismatch analysis (MMD) provided a bimodal distribution of substitution frequencies detected between haplotypes for the complete set of *F. octooculata* populations, as well as for each of the four collected on Livingston and Robert Islands, suggesting demographic equilibrium and constant population size, with no contraction or expansion; estimates of times of expansion support a pre-LGM expansion only for DPL and HPL (Table 3). For the two populations sampled in Nelson Island (HPN and RPN), calculations of MMD resulted in an unimodal distribution (Table 3), which is a feature of populations that have undergone a recent demographic expansion (Rogers & Harpending, 1992).

Mismatch analysis also provided bimodal or multimodal distributions of substitution frequencies for *C. a. antarcticus*, with an estimated time of expansion for DPL that dates to the mid-Pleistocene. In *F. grisea* populations (Table 3), unimodal distribution of substitution frequencies would imply demographic equilibrium, with older times of expansion again being indicated in the Livingston Island population. Collectively, p-values for *R* (raggedness) and *SSD* (sum of squared deviations) indexes were not significant, suggesting failure to reject the null hypothesis of demographic expansion of populations.

*Amova and haplotype clustering in* F. octooculata

Analysis of molecular variance, performed with *F. octooculata* populations either not grouped or grouped according to the island of origin, suggested that the largest total variation was observed within the populations and secondarily among populations within groups (Table 3). When populations were divided in the two groups KN and LR, total variance was still mostly attributable to the intra-population level (Table 3) and haplotype J represents the molecular signature of the genetic dissimilarities between the KN and LR groups. Bayesian analysis of population structure revealed a nested genetic population subdivision into three clusters (CFO 1-3), with log-marginal likelihood of optimal partition of -197.0363. CFO 1 was represented by haplotypes A-I, K, M and O; CFO 2 by J, L, N and Q; CFO 3 by P. This subdivision exactly corresponds to the three major branches of the haplotype network (Fig. 1) and was also represented in the branching patterns of Fig. 2. Among the four investigated islands, only on Livingston Island were populations with haplotypes of all three clusters identified.

*Amova and haplotype clustering in* F. grisea *and* C. a. antarcticus

AMOVA analysis for *F. griesea*, whose populations were tentatively clustered in a similar way to *F. octooculata*, was carried out in order to evaluate if their genetic structure was connected with the geographical distances among islands, and generated negative values of variance components suggesting absence of genetic structure. However, when the components of differentiation were calculated with no structure enforced, most of the variation identified was attributable to the among populations (rather than within populations) comparison (Table 3). In *C. a. antarcticus*, the more limited number of samples (one for each of the four islands under study) prevented association of the populations into groups and the percentage of variation was similar in both among and within population comparisons (Table 3). Bayesian clustering of haplotypes in *F. grisea* also led to the identification of three clusters: CFG 1 (P2 and P7), CFG 2 (P1, P3, P4 and P5) and CFG 3 (P6) (log-marginal likelihood of optimal partition: -86.0226) (Fig. 2). King George and Robert Islands had only haplotypes belonging to CFG 2, and Nelson I. to clusters CFG 1 and CFG 2, while all three clusters were represented in Livingston Island. Clustering of haplotypes for *C. a. antarcticus* resulted in four clusters: CCA 1 (H21-H28), CCA 2 (H6-H8 and H10), CCA 3 (H11, H13-H20, H31, H32, H38 and H39) and CCA 4 (H29 and H30) (log-marginal likelihood of optimal partition: -736.9849). CPR, HPN and PCK had only haplotypes of CCA clusters 3, 2 and 1, respectively, whereas at Devils Point (Livingston Island) three (CCA 2-4) of the four clusters were represented (Fig. 2). Comparison between the clustering analysis and the haplotype network, in *C. a. antarcticus,* suggested that CCA 2 and CCA 4 are grouped in separate subnetworks with respect to the remaining haplotypes (Fig. 1C).

*Molecular clock and timing of population expansion*

Dating the time of the earliest and most recent diversification between haplotypes of *F. ocotooculata*, assuming a divergence rate of 1.5-2.3% Myr-1 (Brower, 1994), led to date range between 104,347 and 160,000 ybp, suggesting an Upper-Middle Pleistocene, pre-LGM (Last Glacial Maximum), differentiation within the species in the South Shetland Islands. The application of the same rate for the seven *F. grisea* haplotypes, and the use of average divergence values for the populations (0.91%) generated a more ancient differentiation (between 395,652 and 606,666 ybp) than in *F. octooculata*, corresponding to the Middle Pleistocene. The average genetic distance between P6 (the most basal haplotype of the branching pattern in Fig. 2) and the remaining haplotypes was 2.2%, leading to an estimated time of earlier divergence between the CFG 1 and CFG 2 clusters of 0.9 to 1.5 million ybp, within the Lower Pleistocene. However, when the most divergent haplotype (P6) is excluded, the estimated diversification time, calculated from the average *p*-distance value among the remaining haplotypes (0.4% between CFG 1-2) reduced to between 173,913 and 266,667 ybp. In *C. a. antarcticus* the average level of genetic divergence (2.6%) led to Lower Pleistocene dates (1.1–1.7 million ybp), suggesting a 10-times older diversification than in *F. octooculata*.

The calculated time of demographic expansion was very similar in *F. octooculata* and in *F. grisea* for their HPN populations (ranging from 7,432 to 12,343 ybp), whereas in HAL and HPL from Livingston Island, *F. grisea* provided evidence of a more ancient expansion date (from 44,413 to 69,735 ybp). In each possible comparison, the populations of *C. a. antarcticus* appear to have expanded much earlier than the other two species studied. Inter-specific comparisons identified the populations of Devils Point specifically, and Livingston Island generally, as showing the earliest demographic expansions (Table 3).

*Demographic analysis*

All *F. octooculata* populations, except for DPL and HPL, had negative values of Tajima’s *D* coefficient, suggesting that the expected average level of variation among haplotypes (i.e. the number of segregating sites) is higher than that observed (Appendix S3). Apart from the mono-haplotypic population of PCK, all *D* values were greater than 0, implying departure from neutrality of nucleotide substitutions. These data collectively provide support to the hypothesis that recent expansion of populations may have occurred, although the low number of specimens investigated for each locality suggests caution in drawing conclusions. Similar results were obtained for *F. grisea*, with negative values observed for all populations (except those represented by a single haplotype), whereas in *C. a. antarcticus* PCK was the only population with a negative *D* value (Appendix S3). Fu’s test of neutrality also suggested recent expansion of the HPN and RPN populations in *F. octooculata*, and for CPR, HPN and PCK in *C. a. antarcticus*. HPN and HPL had negative Fu’s values in *F. grisea.* However, none of these values were statistically significant (at p<0.05) apart from the Fu’s parameter calculated for HPN in *F. octooculata*, for CPR in *C. a. antarcticus* and for HPN in *F. grisea*.

**DISCUSSION**

The onset of the most recent deglaciation in the South Shetland Islands was around 11,000 ybp, with the process continuing until 8,400-6,000 ybp (Ingólfsson *et al.,* 1998). More recent re-advance of ice took place in some areas of the archipelago around 5,000 ybp. This recent glaciological history, although with less severe outcomes with respect to that which impacted the invertebrate biota of the Northern Hemisphere (Bergstrom & Chown, 1999), would have permitted the establishment at local scale of ecosystems capable of hosting the invertebrate life typical of contemporary terrestrial habitats around 5-7,000 ybp (e.g. Hodgson & Convey, 2005). Despite this, the presence of “relict” springtail (and many other) species that must have survived multiple glacial cycles in ice-free refugia in Antarctica is now generally accepted (e.g. Stevens *et al.,* 2003, 2006b; McGaughran *et al.,* 2011), overturning the pre-existing paradigm that the vast majority of the continent’s contemporary terrestrial biota must consist of recent post-LGM dispersers (Convey & Stevens, 2007; Convey *et al.,* 2008).

In the “refugial” scenario, surviving invertebrate taxa would recolonize habitats made available through glacial retreat, effectively now existing in populations isolated by natural barriers. The genetic parameters of the three species studied here, and the molecular estimate of differentiation times generated from them, support the idea that the diversification of their haplotypes in the South Shetland Islands started within the Pleistocene but well before the LGM. However, our data also suggest that separate differentiation events occurred during different time intervals within this geological period. Thus, the higher numbers of total and intra-population haplotypes observed in *C. a. antarcticus* in comparison with its sympatric counterparts suggest that the species has an evolutionary history in the archipelago that can be traced back at least to the Lower Pleistocene. In contrast, more recent diversification events in the Upper-Middle Pleistocene are indicated for the *F. grisea* and *F. octooculata* populations examined.

The high number of haplotypes and low nucleotide diversity observed in the latter species also suggest a recent demographic expansion of its populations (possibly connected with a demographic bottleneck), with many of the haplotypes obtained at low frequency or from single individuals. Demographic analyses collectively provide support to the hypothesis that recent expansion of populations may have occurred, although the low number of specimens investigated for each locality means that caution in drawing conclusions is required.

The estimate obtained here of the timing of diversification within *F. octooculata* (104-160,000 ybp) is consistent with the Valdivian interglacial period recorded from southern South America (115-130,000 ybp) (Astorga & Pino, 2011; NEEM Community Members 2013). Our data indicate that *F. grisea* diversified in the archipelago earlier (396-607,000 ybp) than *F. octooculata*. While a wider date range, this encompasses at least two recorded interglacial periods in southern South America, the most recent being the Hoxnian (374-424,000 ybp) (Lisiecki & Raymo, 2005). The 1.1-1.7 million ybp divergence estimate for *C. a. antarcticus* is also a wide age range. However, given the pre-Pastonian glacial period extended from 0.8-1.3 million ybp, this estimate is consistent with divergence within the preceding Bramertonian interglacial stage of the Pleistocene (1.3-1.55 million ybp) (see Gibbard & van Kolfschoten, 2004 for an overview of glacial/interglacial periods).

The genetic diversity of the South Shetland Islands populations of springtails recorded here is likely to represent only a limited fraction of their total variability. Indeed, previous studies performed on haplotype data for populations examined within a larger geographical context have demonstrated much greater levels of intraspecific differentiation (Stevens *et al.,* 2006a; Torricelli *et al.,* 2010a). In *C. a. antarcticus*, the high number of haplotypes found at low frequency and the almost complete absence of linkage between populations suggest recent population expansion (McGaughran *et al.,* 2011), although genetic divergence parameters and MMD analysis date the differentiation of haplotype lineages (clusters) and demographic expansion to deeper in the past. In *F. grisea*, the amount of genetic divergence observed in the South Shetland Islands populations is negligible in comparison with that calculated between populations inhabiting the entire range of distribution of the species along the Antarctic Peninsula (0.91% *vs* 2.7%) and even more so if related to samples obtained from Victoria Land, in the continental Antarctica (14.4-17.2% divergence between maritime and continental Antarctic haplotypes) (Torricelli *et al.,* 2010a). It should be noted, however, that taxonomic analysis is under way to investigate whether West and East Antarctic populations of *F. grisea* belong to the same species. AMOVA data support the view that different patterns of genetic differentiation are observed in the three species. The component of differentiation is almost equivalent among and within populations of *C. a. antarcticus* grouped according to their island of origin. opposite contrasting pattern is observed in the other two species, with most of the genetic diversity attributable to within population comparisons in *F. octooculata* and to among populations within groups in *F. grisea.*

In *F. octooculata* the large number of haplotypes represented by unique sequences, the high values of *h*, the low estimates of *π* and the results of the MMD analysis all suggest recent demographic expansion, at least for the two populations from Nelson Island for which MMD calculations suggest a demographic expansion subsequent to post-LGM glacial retreat. Among the remaining populations, DPL and HPL (two out of the three from Livingston Island), display higher *π* values and include haplotypes belonging to all the groups identified by the cluster analysis.

The pattern of distribution for the haplotypes is most likely dependent on the initial distribution of A over the entire range of suitable environments, followed by differentiation of J in a more restricted area. The abundance of the J haplotype in the southern part of the sampled area now equals or exceeds that of any other haplotype apart from A, although J is the most frequent haplotype only at Devil’s Point (Livingston Island) (Table 1). The two populations from Nelson Island (HPN and RPN) have the largest number of low-frequency haplotypes derived from A through a single nucleotide substitution (Fig. 1; Appendix S2), and are also the only populations with unimodal distributions of mismatch parameters (Table 3) (implying that they have undergone a recent demographic expansion). Southern populations of *F. octooculata* (i.e. those with either A and J haplotypes; Table 1) are candidate locations that may have ancestrally colonized (or inter-glacially re-colonized) the South Shetland Islands. Specimens with the most frequent haplotypes, A and J (coexisting in Livingston and Robert Islands), may have then dispersed northwards. Only A has (so far) successfully established on Nelson and King George Islands, where it has locally differentiated into several low-frequency haplotypes in the most recently colonised sites (those of Nelson Island).

These factors are consistent with the first haplotype diversification processes taking place on Livingston Island, originating from refugia where the species persisted throughout the LGM period. This process may have initially involved haplotypes from cluster 1 (where the most frequent haplotype A is present), which have locally differentiated to generate clusters 2 and 3 in Livingston and Robert Islands. Cluster 1 is the only cluster whose members are also present on Nelson and King George Islands, supporting a south-to-north route of dispersal. Similarly, in the other two species studied, most (if not all) of the detected haplotype clusters were present in the Livingston Island populations. In addition, the branching pattern obtained in the haplotype phylogenetic tree highlights that CFG 3 and CCA 4 are the most basal clusters (and therefore the ancestral lineages of the remaining groups) for *F. grisea* and *C. a. antarcticus*, respectively, and both groups are unique to Livingston Island. These data again suggest that an earlier diversification occurred on this island, likely initiating from local glacial refugia, with subsequent colonization of the other locations in the archipelago.

The genetic structure of these three springtail species in the South Shetland Islands is characterized by subdivision of the haplotypes into several groups, the demarcation of which is usually restricted to a small number of nucleotide changes. These data highlight slow rates of molecular differentiation in *F. octooculata*, and are consistent with a recent but pre-LGM evolutionary origin for this species within the South Shetland Islands archipelago. In a wider context, source populations over the different timescales required for all three springtail species considered here are likely to have been elsewhere in the maritime Antarctic region, given their wider contemporary distributions, but at present-time no appropriate molecular data are available from across these distributions to allow such an assessment to be made.

**ACKNOWLEDGMENTS**

This work was funded by the Italian Program of Research in Antarctica (PNRA), and the Italian MIUR (PRIN). The logistic support of the British Antarctic Survey (BAS) and HMS *Endurance* was crucial for the collection of the material. Partial support was also provided by the University of Siena. We thank three anonymous reviewers for their constructive and helpful comments. We also thank Dr. Alessandro Manenti for his assistance in the collection of data. PC is supported by NERC core funding to the BAS ‘Biodiversity, Evolution and Adaptation’ Team. This paper also contributes to the Scientific Committee on Antarctic Research ‘State of the Antarctic Ecosystem’ (AntEco) international research programme.

**REFERENCES**

**Ashworth AC, Kuschel G. 2003.** Fossil weevils (Coleoptera: Curculionidae) from latitude 85° S Antarctica. *Palaeogeography, Palaeoclimatology, Palaeoecology* **191:** 191-202.

**Astorga G, Pino M. 2011.** Fossil leaves from the last interglacial in Central-Southern Chile: inferences regarding the vegetation and paleoclimate. *Geologica Acta* **9:** 45-54.

**Bergstrom DM, Chown SL. 1999.** Life at the front: history, ecology and change on southern ocean islands. *Trends in Ecology and Evolution* **14:** 472-477.

**Brower AVZ. 1994.** Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America* **91:** 6491–6495.

**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 Evolutionary Biology* **7:** S8.

**Carapelli A, Comandi S, Convey P, Nardi F, Frati F. 2008.** The complete mitochondrial genome of the Antarctic springtail *Cryptopygus antarcticus* (Hexapoda: Collembola). *BMC Genomics* **9:** 315.

**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* **2:** 190.

**Cheng L, Connor TR, Sirén J, Aanensen DM, Corander J. 2013.** Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Molecular Biology and Evolution* **30:** 1224-1228.

**Chown SL, Convey P. 2007.** Spatial and temporal variability across life's hierarchies in the terrestrial Antarctic. *Philosophical Transactions of the Royal Society B* **362:** 2307–2331.

**Clement M, Posada D, Crandall KA. 2000.** TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657-1659.

**Convey P, Stevens MI. 2007.** Antarctic Biodiversity. *Science* **317:** 1877–1878.

**Convey P, Gibson JAE, Hillenbrand C-D, Hodgson DA, Pugh PJA, Smellie JL, Stevens MI. 2008.** Antarctic terrestrial life - challenging the history of the frozen continent? *Biological reviews of the Cambridge Philosophical Society* **83:** 103–117.

**Convey P, Stevens MI, Hodgson DA, Smellie JL, Hillenbrand C-D, Barnes DKA, Clarke A, Pugh PJA, Linse K, Cary SC. 2009.** Exploring biological constraints on the glacial history of Antarctica. *Quaternary Science Reviews* **28:** 3035–3048.

**Cook CE, Yue Q, Akam M. 2005.** Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. *Proceedings of the Royal Society B* **272:** 1295–1304.

**Enderlein G. 1903.** Die Insekten und Arachnoiden der Kerguelen. *Valvidia* **3:** 199-249.

**Excoffier L, Laval G, Schneider S. 2005.** Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1:** 47–50.

**Fraser CI, Nikula R, Ruzzante DE, Waters JM. 2012.** Poleward bound: biological impacts of Southern Hemisphere glaciation. *Trends in Ecology and Evolution* **27:** 462-471.

**Fraser CI, Terauds A, Smellie J, Convey P, Chown L. 2014.** Geothermal activity helps life survive glacial cycles. *PNAS* **111:** 5634-5639.

**Fu YX. 1997.** Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147:** 915-925.

**Gibbard P, van Kolfschoten T. 2004.** The Pleistocene and Holocene Epochs. In: Gradstein FM, Ogg JG, Smith A. Gilbert, eds. *A Geologic Time Scale 2004.* Cambridge: Cambridge University Press, 441-453.

**Greenslade P. 1986.** Additions to the Collembola of Heard Island. *Records of the South Australian Museum* **19:** 91–96.

**Greenslade P. 1995.** Collembola from the Scotia Arc and Antarctic Peninsula including description of two new species and notes on biogeography. *Polskie Pismo Entomologiczne* **64:** 305–319.

**Greenslade P. 2010.** South Shetlands Collembola fauna revisited. *Antarctic Science* **22:** 233–242.

**Greenslade P, Convey P. 2012.** Exotic Collembola on subantarctic island: pathways, origin and biology. *Biological Invasions* **14:** 405-417.

**Harpending HC. 1994.** Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* **66:** 591-600.

**Hawes TC, Torricelli G, Stevens MI. 2010.** Haplotype diversity in the Antarctic springtail *Gressittacantha terranova* at fine spatial scales - a Holocene twist to a Pliocene tale. *Antarctic Science* **22:** 766-773.

**Hayward SAL, Worland MR, Convey P, Bale JS. 2004.** Effects of moisture on the local distribution of the Antarctic Collembola *Cryptopygus antarcticus* and *Friesea grisea*. *Soil Biology and Biochemistry* **36:** 927-934.

**Hodgson DA, Convey P. 2005.** A 7000 year record of the oribatid mite communities on a maritime-Antarctic island: responses to climate change. *Arctic Antarctic and Alpine Research* **37:** 239-245.

**Ingólfsson Ó, Hjort C, Berkman P, Björck S, Colhoun E, Goodwin ID, Hall B, Hirakawa K, Melles M, Möller P, Prentice M. 1998.** Antarctic glacial history since the Last Glacial Maximum: an overview of the record on land. *Antarctic Science* **10:** 326-344.

**Lanfear R, Calcott B, Ho SYW, Guindon S. 2012.** Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29:** 1695-1701.

**LeMasurier WE. 1990.** *Volcanoes of the Antarctic Plate and Southern Oceans*. Thomson JW. American Geophysical Union, Vol 48: Washinghton, DC, 1-487.

**Lewis A, Marchant D, Ashworth A, Hedenäs L, Hemming S, Johnsong J, Leng M, Newton A, Raine J, Willenbring J, Williams M, Wolfem A. 2008.** Mid-Miocene cooling and the extinction of tundra in continental Antarctica. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 1–5.

**Lisiecki LE, Raymo ME. 2005.** A Pliocene-Pleistocene stack of 57 globally distributed benthic d18O records. *Paleoceanography* **20:** 1-17.

**Maddison DR, Maddison WP. 2005.** MacClade 4: Analysis of phylogeny and character evolution. Version 4.08a*.* Available at: http://macclade.org.

**Mantel N. 1967.** The detection of disease clustering and a generalized regression approach. *Cancer Research* **27:** 209–220.

**McGaughran A, Hogg ID, Stevens MI. 2008.** Patterns of population structure for springtails and mites in southern Victoria Land, Antarctica. *Molecular Phylogenetics and Evolution* **46:** 606–618.

**McGaughran A, Redding GP, Stevens MI, Convey P. 2010.** Patterns of temporal and spatial metabolic rate variation in an Antarctic springtail. *Journal of Insect Physiology* **56:** 57–64.

**McGaughran A, Stevens MI, Hogg ID, Carapelli A. 2011.** Extreme glacial legacies: A synthesis of the Antarctic springtail phylogeographic record. *Insects* **2:** 62-82.

**NEEM community members (2013).** Eemian interglacial reconstructed from a Greenland folded ice core. *Nature* **493:** 489–94.

**Nei M. 1987.** *Molecular Evolutionary Genetics*. Columbia University Press: New York.

**Potapov M. 2001.** Synopses on Palaearctic Collembola vol. 3: Isotomidae. *Abhandlungen und Berichte des Naturkundemuseums Görlitz* **73:** 1-603.

**Ronquist F, Huelsenbeck JP. 2003.** MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572-1574.

**Rogers A, Harpending H. 1992.** Population growth makes waves in the distribution of pairwise differences. *Molecular Biology and Evolution* **9:** 552-569.

**Russell DJ, Hohlberg K, Potapov M, Bruckner A, Otte V, Christian A. 2014.** Native terrestrial invertebrate fauna from the northern Antarctic Peninsula: new records, state of current knowledge and ecological preferences – Summary of a German federal study. *Soil Organisms* **86:** 1-58.

**Schneider S, Excoffier L. 1999.** Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics* **152:** 1079–1089.

**Stevens MI, Hogg ID. 2003.** Long-term isolation and recent range expansion from glacial refugia revealed for the endemic springtail *Gomphiocephalus hodgsoni* from Victoria Land, Antarctica. *Molecular Ecology* **12:** 2357–2369.

**Stevens MI, Greenslade P, Hogg ID, Sunnucks P. 2006a.** Examining Southern Hemisphere springtails: could any have survived glaciation of Antarctica? *Molecular Biology and Evolution* **23:** 822–874.

**Stevens MI, Fjellberg A, Greenslade P, Hogg ID, Sunnucks P. 2006b.** Redescription of the Antarctic springtail *Desoria klovstadi* using morphological and molecular evidence. *Polar Biology* **29:** 820–830.

**Swofford DL. 2003.** PAUP\* Phylogenetic analysis using parsimony (\* and other methods), Version 4. Sinauer Associates, Sunderland, MA.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105:** 437-460.

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

**Torricelli G, Frati F, Convey P, Telford M, Carapelli A. 2010a.** 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:** 757–765.

**Torricelli G, Carapelli A, Convey P, Nardi F, Boore JL, Frati F. 2010b.** High divergence across the whole mitochondrial genome in the ‘pan-Antarctic’ springtail *Friesea grisea*: evidence for cryptic species? *Gene* **449:** 30–40.

**Willem V. 1901.** Les Collemboles recueillis par l’expédition Antarctique belge. *Annales de la Société Entomologique de Belgique* **45:** 260-262.