Zoologica Scripta

WILEY | KUNGL. VETENSKAPS-AKADEMIEN

ORIGINAL ARTICLE OPEN ACCESS

Genome Wide Analyses Corroborate Cryptic Speciation in *Anurida maritima* (Collembola, Poduromorpha)

Martijn Timmermans

Department of Natural Sciences, Middlesex University, London, UK

Correspondence: Martijn Timmermans (m.timmermans@mdx.ac.uk)

Received: 6 January 2025 | **Revised:** 1 March 2025 | **Accepted:** 29 March 2025

Funding: The work was supported by funding from the Faculty of Science and Technology, Middlesex University.

ABSTRACT

Cryptic diversity is ubiquitous within Collembola (springtails). Numerous studies have reported substantial genetic divergence within morphological species. One potentially involved morphospecies is *Anurida maritima* (Guérin-Méneville, 1836). *A. maritima* is confined to the intertidal zone, where it often occurs in high densities. A recent study reported two distinct mitochondrial lineages from North-western Europe. One of the two mitochondrial lineages was found to be associated with *Anurida bisetosa* (Bagnall, 1949), a species that was synonymised with *A. maritima* in 1953. Here, genome-wide analyses are used to show that the observed mitochondrial divergence extends to the nuclear genome. Phylogenetic analysis of four natural populations using data on ~1500 single copy orthologous genes supported a separation of *A. bisetosa* sp. dub. and *A. maritima*. Genetic analyses using a Pool-seq approach corroborated the genome-wide differentiation. The four populations under investigation came from a relatively narrow geographical range (United Kingdom, The Netherlands). *A. maritima* has a Holarctic distribution and, hence, it is quite possible that the group includes further highly diverged genetic lineages or cryptic species. This study offers another compelling example of genetic divergence not necessarily leading to noticeable morphological change.

1 | Introduction

Collembola (springtails) are a globally distributed Class of hexapods. They are among the most abundant terrestrial arthropods worldwide and inhabit a wide range of environments (Hopkin 1997), including polar deserts (Carapelli et al. 2020), forest soils (Cicconardi et al. 2010), coastal sands (Thibaud and Christian 1997) and caves (Kováč et al. 2023). Their diets are wide ranging, and the group includes specialist microbivores, herbivores, scavengers, and carnivores (Malcicka et al. 2017). Despite their global distribution and remarkable ecological adaptations, they exhibit only moderate morphospecies diversity (Cicconardi et al. 2013). Currently, ~9500 species are recognised (Bellinger et al. 1996), but the true number is doubtlessly significantly higher. Estimates are likely too low because of incomplete sampling and unrecognised genetic diversity and divergence (Bellini et al. 2023; Katz et al. 2015; Rusek 1998). High levels of intraspecific genetic diversity have been reported for many species, and cryptic speciation is expected to be widespread. At least one molecular study has estimated that there may be up to be 500,000 extant Collembola species (Cicconardi et al. 2013). Even in the absence of observable morphological differences, genetic separation is often ancient and began millions of years ago (Mya). For example, Porco et al. (2012) focussed on six widespread and morphologically robust springtail species and, using phylogenetic methods, revealed 20 independently evolving lineages that originated 8.52–17.88 Mya (Neogene). Characterising such hidden diversity is critical for understanding species richness, (global) biodiversity patterns, and preventing taxonomic confusion (Hending 2025).

The collembolan *Anurida maritima* (Guérin-Méneville, 1836) is a littoral species that forages on decaying matter in the upper intertidal zone (Imms 1906) (Figure 1). It has

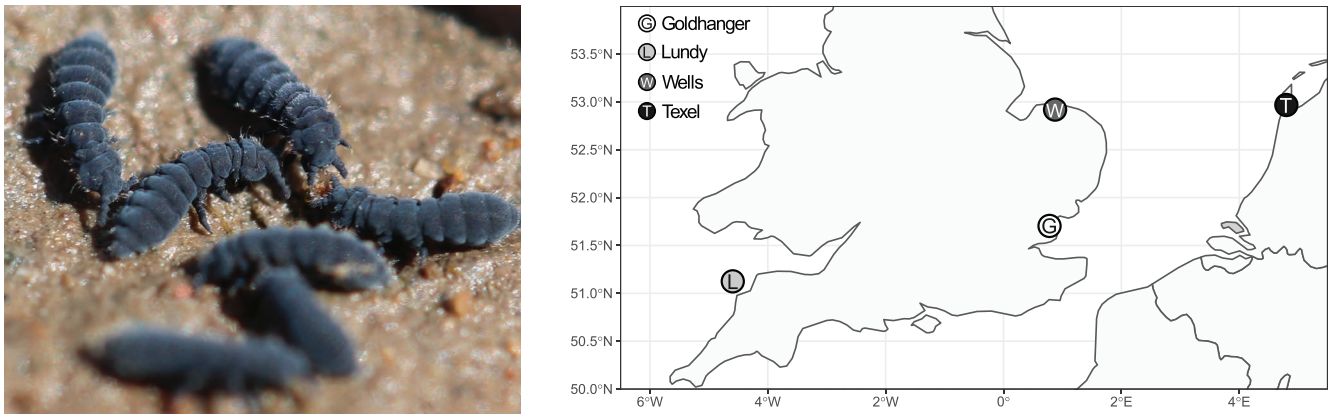

FIGURE 1 | Left: *Anurida maritima* (Guérin-Méneville, 1836). Photo credit: Stephen Kett. Right: The four locations, Goldhanger, Texel, Wells-next-the-Sea, Lundy.

a Holarctic distribution (Imms 1906; Joosse 1966) and can be extremely abundant in places (Dexter 1943). The species potentially has a sibling species, *Anurida bisetosa*, which was first described by Bagnall in 1949 (Bagnall 1949). *A. bisetosa* was considered a junior synonym of *A. maritima* by Goto and Delamare Deboutteville in 1953 (Goto and Delamare Deboutteville 1953) due to the defining morphological characters being variable and considered uninformative. With the use of a different set of taxonomic characters, Arbea (2001) was able to confidently differentiate *A. bisetosa* from *A. maritima* and hence suggested the reinstatement of the former. This reinstatement, however, is currently not generally accepted (Bellinger et al. 1996; Hopkin 2007).

Recently a study that analysed full mitochondrial genome sequences revealed two deeply diverged *A. maritima* lineages in north-western Europe (Timmermans et al. 2022). Interestingly, one of the two mitochondrial lineages was found to be linked to samples that bore the morphological characters of *A. bisetosa* (as described by Arbea 2001; see also Timmermans et al. 2022). In this study, genome-wide analyses are used to investigate the genomic extent of the genetic divergence. More specifically, publicly available RNA-seq data are supplemented by whole genome shotgun sequencing data from pooled individuals from three natural populations to obtain information on ~1500 single copy orthologous genes. The datasets are analysed using phylogenetic methods and a Pool-seq approach. These analyses provide valuable insights into population diversity and divergence, further clarifying the taxonomic validity of *A. bisetosa*.

2 | Methods

Anurida bisetosa is currently not accepted as a valid species and will be referred to as 'species dubia' (sp. dub.), 'doubtful species' (Sigovini et al. 2016).

2.1 | Sample Collection and Publicly Available Data

For this study, animals were collected on Lundy, Bristol Channel, UK (51°09′45.8″N 4°39′19.2″W) on the 17th of June 2023 and near Goldhanger, Maldon, UK (51°44′25.9″N 0°45′54.1″ E) on the 19th of July 2023 (Figure 1). Animals from both locations have previously been morphologically identified as *Anurida bisetosa* sp. dub. (J. Arbea, personal communication; see Timmermans et al. 2022 for more detail). Animals were transported to the laboratory and kept in small plastic containers on moist filter paper until DNA extraction and sequencing. *A. bisetosa* sp. dub. samples were supplemented by publicly available sequence data on two populations of *A. maritima* (Figure 1): (1) A whole genome shotgun sequence dataset that originated from 30 pooled specimens from Wells-next-the-Sea, Norfolk, UK (Timmermans et al. 2023; NCBI SRA Accession number: SRR23074001; Oxford Nanopore Technology); (2) A transcriptome dataset that originated from 100 pooled specimens from Texel, North-Holland, The Netherlands (Misof et al. 2014; NCBI SRA Accession number: SRR921564; Assembly Accession number GAUE00000000; Illumina technology).

2.2 | DNA Extraction, Sequencing and Draft Genome Assembly

DNA extractions of *Anurida bisetosa* sp. dub. from Lundy and Goldhanger were performed on 15 pooled specimens per sample using the NEB Monarch DNA extraction kit following the manufacturer supplied protocol for insects. DNA quality and quantity were assessed using an Agilent 2200 TapeStation system (Agilent, Waldbronn, Germany). DNA sequencing libraries were constructed using the Rapid Sequencing Kit (SQK-RAD004; Oxford Nanopore, Oxford, UK), which were then sequenced on a MinION and a R9.4.1 flow cell. GPU Base-calling was performed using ont-guppy (guppy_basecaller version 5.0.7 + 2332e8d) and NanoPlot 1.43.0 (De Coster and Rademakers 2023) was used to summarise the sequencing data.

To estimate genome sizes, the data were analysed using kmerfreq and GCE (Liu et al. 2013; Wang et al. 2020). For these analyses, the raw sequences were first corrected and trimmed using CANU 2.0 (Koren et al. 2017) as described by Timmermans et al. (2023) with the difference that only reads with a length longer than 4000 were included. Kmerfreq used default settings (i.e., kmer size 17) and GCE was run in the heterozygous mode (i.e., using options: -c 75 -H 1). The analyses were also

performed on the *A. maritima* data from Wells-next-the-Sea from Timmermans et al. (2023) (Accession number: SRR23074001).

The three trimmed and corrected ONT datasets were subsequently assembled using flye version 2.9.3-b1797 (Kolmogorov et al. 2019) with the following settings: —asm-coverage 40—genome-size 100 m. Completeness of the draft genomes was assessed using BUSCO v5.2.2 (Simão et al. 2015) with the arthropoda_odb10 gene set. Finally, the lengths of the draft genomes were compared to all Collembola genomes publicly available via NCBI Datasets (last accessed: 22/10/2024; 167 genomes) and to the Collembola genome size estimates reported by the MetaInvert project (Collins et al. 2023; 77 species).

2.3 | Mitochondrial Haplotypes

To determine which mitochondrial lineages (i.e., *A. maritima* or *A. bisetosa* sp. dub.) were present in each of the four populations (Wells-next-the-Sea, Texel, Lundy, Goldhanger), the raw sequencing reads were mapped simultaneously onto two publicly available partial cytochrome c oxidase subunit I (COX1) reference gene sequences from Wells-next-the-Sea (*A. maritima*) and Lundy (*A. bisetosa* sp. dub.) (Barcode region; Genbank accession numbers: MT434145 and MT434146). This used minimap2 (Li 2018) (setting: -ax map-ont) for the ONT datasets, and bwa mem (0.7.18-r1243-dirty) (Li and Durbin 2009) for the Illumina RNA-seq dataset. Mapped reads were trimmed back to the COX1 reference sequences, extracted using Geneious Prime Version 2024, and exported in fastq format. The ONT fastq files were filtered with NanoFilt (De Coster et al. 2018) removing reads with a minimum average read quality score below 12. Consensus sequences were generated using amplicon_sorter (Vierstraete and Braeckman 2022) using reads with a length of 640 bp or longer. This was not possible for the Illumina-based Texel fastq file. Instead, the reads were assembled using SPAdes v3.13.1 (Prjibelski et al. 2020) using the 'meta' flag for metagenomic sample. The obtained consensus sequences were compared with *Anurida* COX1 barcodes available in BOLD v4 (Ratnasingham et al. 2024) (accessed: 9 December 2024). For this, the sequences were aligned using MAFFT (Katoh and Standley 2013) and a phylogenetic tree was constructed using the web version of IQ-TREE (Trifinopoulos et al. 2016), using ModelFinder (Kalyaanamoorthy et al. 2017) to select an appropriate model and ultrafast bootstrap (1000 replicates) to assess branch support.

2.4 | Single Copy Orthologous Gene Set and Phylogenetic Analyses

To conduct phylogenomic analyses, a single-copy orthologous gene (SOG) set was created based on the Collembola_obd1 BUSCO dataset of Sun et al. (2020) (Available from: https://figshare.com/articles/dataset/BUSCO_dataset_designed_for_Collembola/10269947). This set is Collembola specific and consists of 1997 single-copy orthologous genes. To generate a SOG set for *A. maritima*, it was compared to the assembled Illumina transcriptome dataset of Misof et al. (2014) (22,076 transcripts; Accession number: GAUE00000000). Comparisons between the two datasets were performed using TBLASTN (Altschul et al. 1997), keeping the first good hit for each only (-max_target_seqs 1 -evalue 10e-10). For each retrieved *A. maritima* transcriptome sequence, the longest Open Reading Frame (ORF) was obtained using TransDecoder (version 5.7.0) (https://github.com/TransDecoder/TransDecoder), keeping the single best ORF (—single_best_only) and without refining the start site (—no_refine_starts).

To generate a data matrix for phylogenetic analyses, each ORF of the *A. maritima* SOG set was used as 'bait' to extract homologous sequences from the three flye genome assemblies (i.e., Wells-next-the-Sea, Goldhanger, Lundy; see above) and two publicly available *Anurida* genome assemblies (*Anurida granaria* Accession number: GCA_034694825 and *Anurida granulata* Accession number: GCA_034699565). The latter two were included to serve as outgroups. Extractions were performed using the Alibaseq sequence extraction software (Knyshov et al. 2021). The transcriptome baits (Texel) were subsequently combined with the extracted sequences of the five taxa and aligned using MAFFT v7.490 (Katoh et al. 2009) using the 'auto' setting. The 'remove-alignment-gaps' utility of the MEME suite (Bailey et al. 2015) was used to remove any alignment columns that contained gaps in the Texel 'in-frame' ORF bait sequence. The individual gene alignments were then concatenated using AMAS (Borowiec 2016) and used for phylogenetic analysis. Phylogenetic analysis was performed using IQ-TREE version 2.2.6 (Nguyen et al. 2015) and involved model selection (ModelFinder) (Kalyaanamoorthy et al. 2017) and assessment of branch supports with ultrafast bootstrap approximation and SH-like approximate likelihood ratio test (Guindon et al. 2010), each with 1000 replicates. Individual gene alignments that contained data on all six taxa were also used for the construction of gene trees. Trees were constructed using IQ-TREE version 2.2.6 (Nguyen et al. 2015) (GTR + I + G model) and a majority-rule consensus tree was generated. The individual gene trees were also plotted using the densiTree function of the R package phangorn (Schliep 2011). This latter step was repeated using ultrametric trees obtained using the chronos function of the R package ape (Paradis and Schliep 2019).

2.5 | Pool-Seq Analyses

To take within-sample allelic variation into account, Pool-seq analyses were performed on the three raw ONT datasets (Wells-next-the-Sea, Goldhanger, Lundy). The Texel dataset was excluded because differential gene expression among genes and individuals within samples can affect relative allele frequencies, and hence transcriptome datasets are not suitable for Pool-seq analyses. ONT reads were mapped onto each of the ORFs of the *A. maritima* SOG set using minimap2 (Li 2018) (setting: -ax map-ont). Clair3 v1.0.4 (Zheng et al. 2022) was used for variant calling using the ont_guppy5 model. A minimum SNP allele frequency of 10% (—snp_min_af = 0.10) was used, and Indel variation was ignored (—call_snp_only). No phasing was performed (—no_phasing_for_fa) and calls (including reference calls; —print_ref_calls) were outputted in gVCF format (—gvcf). The three gVCF files were merged using GLnexus (Lin et al. 2018) into a single BCF file, which was converted to VCF using bcftools (Danecek et al. 2021). At this stage, 3 ALT alleles added by GLnexus for sites that could not be unified ('MONOALLELIC')

were removed. To obtain a VCF file for every single gene, this merged VCF file was split using the python bioinfokit toolkit (splitvcf) (Bedre 2022). The individual files were then parsed through the R package poolfstat (Gautier et al. 2022; Hivert et al. 2018) to obtain estimates of population pairwise Fst and within-population heterozygosities over all SNPs using the compute.fstats function. Values were visualised using density plots with ggplot2 (Wickham and Sievert 2016) in R 4.3.3 (R Core Team 2013). In addition, the full dataset (merged VCF) was analysed using the compute.pairwiseFST function to obtain an overall estimate of population pairwise Fst, with standard errors estimated using block-jackknife (nsnp.per.bjack.block = 10).

3 | Results

3.1 | Genome Assembly

Oxford Nanopore sequencing data were generated for natural populations from Lundy and Goldhanger. For Lundy, 2,806,008 reads with a mean length of 4269 bp and Phred quality of 12.4 were obtained. For Goldhanger, 881,590 reads with a mean length of 4945 bp and Phred quality of 10.8 were obtained. Reads of both samples, and that of a previously obtained sample from Wells-next-the-Sea (Timmermans et al. 2023), were corrected and trimmed and used to obtain genome size estimates, which were 87.5 million base pairs (Mbp) for Wells-next-the-Sea, 91.1 Mbp for Goldhanger, and 105.3 Mbp for Lundy. The data were subsequently assembled, and the completeness of each draft genome was estimated and ranged from 90.5% to 93% (Table 1). These results, when compared to publicly available Collembola genomes and genome size estimates provided by the MetaInvert project (Collins et al. 2023) suggest that species of the genus *Anurida* have relatively small genomes (Figure 2A).

3.2 | Mitochondrial Haplotypes

Animals from Lundy and Goldhanger have previously been identified as *A. bisetosa* sp. dub. Animals from Wells-next-the-Sea have previously been identified as *A. maritima*. Although it is unknown if the Texel material was morphologically assessed, specimens from the population have previously been reported to carry an *A. maritima* mitochondrial haplotype (Timmermans et al. 2022). To determine which mitochondrial lineages (*A. maritima* or *A. bisetosa* sp. dub.) were present in each of the four population samples used here, reads were mapped onto two reference COX1 sequences (MT434145; *A. maritima* originating from Wells-next-the-Sea, and MT434146, *A. bisetosa* sp. dub. originating from Lundy). For Texel and Wells-next-the-Sea, 100% of the reads preferentially mapped back onto the *A. maritima* COX1 sequence (136,871 and 13,958 reads, respectively). For Goldhanger and Lundy, 100% of the reads preferentially mapped back onto *A. bisetosa* sp. dub. COX1 sequence (9405 and 22,058 reads, respectively). Mapped reads were trimmed back to the COX1 reference sequences and assembled. For Texel, Lundy, and Goldhanger, this resulted in a single consensus sequence. The Lundy consensus sequence was 100% identical to the *A. bisetosa* sp. dub. reference sequence, and the Goldhanger consensus differed from it at a single position only. The Texel consensus differed at a single position from the *A. maritima* reference. For the Wells-next-the-Sea dataset, two consensus sequences were obtained. The first sequence differed at a single position from the *A. maritima* reference sequence. The second, which represented 1.9% of the reads, showed 92.2% and 83.4% Pairwise Identity to the *A. maritima* and *A. bisetosa* sp. dub. COX1 references, respectively. However, it was 100% identical to a sequence previously reported from Normandie (France) (Appendix S1).

3.3 | Phylogenetic Analysis

Genome wide comparisons were performed using single-copy orthologous gene (SOG) sequences. In total, 1511 of the 1997 SOG of Sun et al. (2020) were represented in the transcriptome dataset of Misof et al. (2014), of which 1509 passed TransDecoder ORF extraction. These 1509 ORFs were used as 'baits' to extract homologous regions from the three draft genomes and two publicly available *Anurida* genomes that were included as outgroups. Individual alignments were concatenated for a data matrix of 2,604,372 bp, which was used for phylogenetic inference. A topology with strong branch support was obtained. In this

TABLE 1 | Statistics for the three genome assemblies.

	Goldhanger	Lundy	Wells-next-the-Sea
Total length (bp)	101,427,035	116,068,147	91,936,796
N50 length (bp)	1,156,702	1,142,924	4,271,721
BUSCO			
Complete	917 (90.5)	936 (92.4)	942 (93)
Complete single copy	903 (89.1)	909 (89.7)	936 (92.4)
Complete duplicated	14 (1.4)	27 (2.7)	6 (0.6)
Fragmented	37 (3.7)	29 (2.9)	22 (2.2)
Missing	59 (5.8)	48 (4.7)	49 (4.8)
Total	1013	1013	1013

Note: Values in brackets give percentages.
Abbreviations: BUSCO, benchmarking universal single-copy orthologs (Simão et al. 2015); Total: total number of BUSCOs included in the analyses.

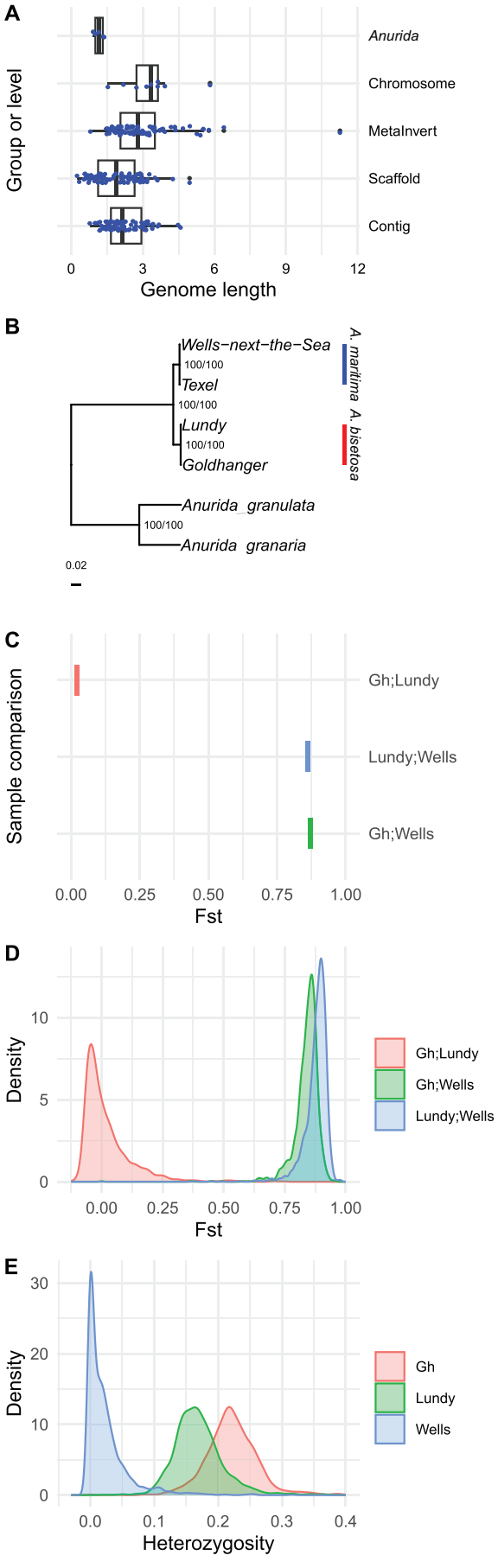

FIGURE 2 | (A) The sizes of five *Anurida* genome assemblies (three genomes presented here, plus *Anurida granaria* and *Anurida granulata*) compared to genome size estimates given by the MetaInvert soil invertebrate genome resource for 77 Collembolan species (Collins et al. 2023) and to chromosome ($n = 9$), scaffold ($n = 79$) and contig ($n = 80$) level assemblies available via NCBI Datasets (last accessed: 22/10/2024). Horizontal axis: Values x 100,000. Boxes represent interquartile ranges, vertical lines within boxes are medians, whiskers represent the range of the data and dots given beyond the whiskers represent outliers. Note that the MetaInvert genomes were submitted to NCBI Datasets and hence there is redundancy. The MetaInvert boxplot displays genome sizes estimated by the authors. The other boxplots show lengths given in NCBI Datasets. (B) ML phylogenetic tree based on 1509 single copy orthologous gene fragments. Values give branch supports with ultrafast bootstrap approximation on the left and SH-like approximate likelihood ratio test values on the right. The *A. maritima* and *A. bisetosa* sp. dub. lineages are indicated with a blue and a red line, respectively. *A. granaria* and *A. granulata* are included as outgroup species. (C) Pairwise Fst as obtained from the full (merged) dataset with standard errors estimated using block-jackknife. (D) Density plot of pairwise Fst values estimated for individual gene fragments. (E) Density plot of within population heterozygosities for individual gene fragments for each of the four samples under study. Gh, Goldhanger; Wells, Wells-next-the-Sea.

topology, the Wells-next-the-Sea and Texel samples grouped together and the Lundy and Goldhanger samples grouped together (Figure 2B). This topology was also strongly supported by the individual gene-based analyses (alignments with data on all 6 taxa only, $n = 1337$), with more than 90% of the obtained gene trees matching the taxon arrangement (Appendix S2).

3.4 | Pool-Seq Analyses

To include within-sample allelic variation in the analyses, sample pairwise Fst and within-sample heterozygosities were estimated using the Pool-seq framework. For this, the ONT reads were mapped back onto the 1509 single-copy ORFs. SNPs were called in 1498 ORFs. High levels of population pairwise Fst were observed when comparing the *A. bisetosa* sp. dub. samples (Lundy and Goldhanger) with the *A. maritima* sample (Wells-next-the-Sea). Compared to these estimates, Fst values for the comparison that involved the same lineage were low (Figure 2C,D). The two *A. bisetosa* sp. dub. samples showed higher heterozygosity than the *A. maritima* sample (Figure 2E).

4 | Discussion

It is not uncommon for Collembola to display high levels of intraspecific diversity, to such an extent that many morphospecies might consist of unrecognised sibling species (Emerson et al. 2011). In this study, genome-wide divergence between two representatives of the *Anurida maritima* species group (*Anurida maritima* and *Anurida bisetosa* sp. dub.) was investigated. Draft genome sequences were obtained and genome sizes estimated to be around 100 Mb in length. This is at the lower

end of the range reported for other Collembola and smaller than that of the congeneric *A. granulata* and *A. gregaria* (Collins et al. 2023). Phylogenetic analysis using ~1500 gene fragments was in agreement with previous mitogenome-based analyses of Timmermans et al. (2022). It confirmed the existence of two divergent lineages within the species as it is currently recognised and supports the separation of *A. bisetosa* sp. dub. and *A. maritima*. Allele-based analyses corroborated this finding. High levels of genetic differentiation were uncovered when comparing samples of different lineages, suggesting that gene flow is severely restricted between the populations involved. Prior to reinstating *A. bisetosa* sp. dub., genetic and morphological information on further populations should be gathered, however, as only a limited number of samples were included here.

The cause of the genetic split and the distributions and phylogeographic structures of *A. maritima* and *A. bisetosa* sp. dub. remain unknown. Phylogeographic structure arises from a combination of historical processes and the migration capacity of species (Avise 2009). It has been suggested that for many soil arthropods, present-day structures result from processes acting on co-occurring pre- and post-glacial lineages (Rosenberger et al. 2013). Recent phylogeographic studies of European collembolans indicate that during the glaciations of the Quaternary period (2.58 Ma onwards) species were not necessarily restricted to prominent ice-free areas in Southern Europe, such as those in present-day Italy and the Balkans, but might also have survived in more northerly areas (Faria et al. 2019; von Saltzwedel et al. 2016). Survival in such areas was likely enabled by the groups' close association to the temperature-buffering soil environment and their diverse diets (von Saltzwedel et al. 2016). During these cold periods, populations were isolated and reduced in size, resulting in the erosion of genetic diversity, potentially amplifying the creation of genetically distinct and geographically separated lineages (Avise 2009; Waters et al. 2013). In this respect, it is intriguing that the two *A. bisetosa* sp. dub. populations (Goldhanger and Lundy) showed higher heterozygosity than the *A. maritima* population from Wells-next-the-Sea. Although no conclusions can be drawn due to the limited number of populations analysed, the observation could suggest that North-Western Europe was a central area of initial expansion after the glaciations for *A. bisetosa* sp. dub. (Ibrahim et al. 1996) or that the lineage held a larger effective population size than *A. maritima* over time (Kimura 1968).

Migration potential differs widely within the Collembola (Ponge 2020). Many species prefer wet or damp environments, and the risk of desiccation severely limits their dispersal ability. However, passive long-distance dispersal has been reported or inferred for various species (Collins et al. 2019; Potapov et al. 2020; Schuppenhauer et al. 2019; Van Der Wurff et al. 2003). For example, some Collembola naturally float on water and have been reported to survive marine dispersal via passive drift (Coulson et al. 2002; Hensel et al. 2016). *A. maritima* has a hydrophobic cuticle and is often found floating on water surfaces in large numbers (Imms 1906). Therefore, it is plausible that *A. maritima* could also withstand long-distance marine dispersal. This notion is corroborated by the observation that a mitochondrial haplotype known to occur in Normandie (France) is present in the Wells-next-the-Sea population. The haplotype is most similar to those of *A. maritima*, and although it was represented by a relatively small number of sequencing reads (< 2%; potentially originating from a single specimen in the pool of 30), it clearly suggests there has been genetic exchange between the two regions. The other populations investigated consisted of one lineage only. This could indicate a 'founder-takes-all' process of colonisation. Under this model, founding individuals rapidly increase in numbers and prevent other lineages from invading (Waters et al. 2013). The observed pattern might, however, also be a result of competitive exclusion, where two species competing for the same limited resources will not coexist (Hardin 1960). This latter notion could potentially be tested using controlled, artificial environments (e.g., see Timmermans et al. 2024).

Two intertidal species closely related to *A. maritima* have been described from Florida (*Anurida ashbyae*) and Costa Rica (*Anurida mara*) (Christiansen and Bellinger 1988). The species are morphologically nearly indistinguishable and the degree of similarity is so high that historical records of *A. maritima* from Florida are likely to be *Anurida ashbyae* (Christiansen and Bellinger 1988). Together, the species are sometimes treated as a species group (e.g., Arbea 2001) as they display various characters that clearly separate them from other congeneric species. To gain a more comprehensive understanding of *A. maritima*, the next step will be to sample more densely and genetically characterise individual springtails from a wider geographical range. Especially with the inclusion of the North and Middle American species, these studies will resolve the phylogeographic structure, phylogenetic relationships and speciation processes of the group.

Acknowledgements

The author would like to thank Maldon District Council, the Landmark Trust and Natural England for providing permission to sample animals in the Blackwater estuary and on Lundy. Ethical approval was obtained from the Natural Science (Middlesex University) Research Ethics Committee. The author would like to thank Stephen Kett for valuable feedback on an earlier version of the text and Sarder Kamal for computational support. The work was supported by funding from the Faculty of Science and Technology, Middlesex University.

Data Availability Statement

The newly generated Goldhanger and Lundy sequences have been uploaded to the NCBI SRA database at https://www.ncbi.nlm.nih.gov/sra/ (BioProject ID: PRJNA1196671). The three draft genome assemblies are available at: https://doi.org/10.22023/mdx.28001105.v1.

References

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, W. Miller, and D. J. Lipman. 1997. "Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs." *Nucleic Acids Research* 25: 3389–3402.

Arbea, J. 2001. "Las Especies del Grupo de *Anurida maritima* (Guerin, 1839) (Collembola:Neanuridae) en la Peninsula Iberica." *Revista Aragonesa de Entomologia* 9: 9.

Avise, J. C. 2009. "Phylogeography: Retrospect and Prospect." *Journal of Biogeography* 36: 3–15.

Bagnall, R. S. 1949. "Notes on British Collembola." *Entomologist's Monthly Magazine* 85: 1017.

Bailey, T. L., J. Johnson, C. E. Grant, and W. S. Noble. 2015. "The MEME Suite." *Nucleic Acids Research* 43, no. W1: W39–W49. https://doi.org/10.1093/nar/gkv416.

Bedre, R. 2022. "Reneshbedre/Bioinfokit: Bioinformatics Data Analysis and Visualization Toolkit (Version 2.0.9) [Computer Software]." *Zenodo*: 3698145. https://doi.org/10.5281/ZENODO.3698145.

Bellinger, P. F., K. A. Christiansen, and F. Janssens. 1996. "Checklist of the Collembola of the World." http://www.collembola.org.

Bellini, B. C., W. M. Weiner, and B. R. Winck. 2023. "Systematics, Ecology and Taxonomy of Collembola: Introduction to the Special Issue." *Diversity* 15, no. 2: 221. https://doi.org/10.3390/d15020221.

Borowiec, M. L. 2016. "AMAS: A Fast Tool for Alignment Manipulation and Computing of Summary Statistics." *PeerJ* 4: e1660. https://doi.org/10.7717/peerj.1660.

Carapelli, A., C. Cucini, P. P. Fanciulli, F. Frati, P. Convey, and F. Nardi. 2020. "Molecular Comparison Among Three Antarctic Endemic Springtail Species and Description of the Mitochondrial Genome of *Friesea gretae* (Hexapoda, Collembola)." *Diversity* 12, no. 12: 450. https://doi.org/10.3390/d12120450.

Christiansen, K., and P. Bellinger. 1988. "Marine Littoral Collembola of North and Central America." *Bulletin of Marine Science* 42, no. 2: 2.

Cicconardi, F., P. P. Fanciulli, and B. C. Emerson. 2013. "Collembola, the Biological Species Concept and the Underestimation of Global Species Richness." *Molecular Ecology* 22, no. 21: 5382–5396. https://doi.org/10.1111/mec.12472.

Cicconardi, F., F. Nardi, B. C. Emerson, F. Frati, and P. P. Fanciulli. 2010. "Deep Phylogeographic Divisions and Long-Term Persistence of Forest Invertebrates (Hexapoda: Collembola) in the North-Western Mediterranean Basin: MEDITERRANEAN SPRINGTAIL PHYLOGEOGRAPHY." *Molecular Ecology* 19, no. 2: 386–400. https://doi.org/10.1111/j.1365-294X.2009.04457.x.

Collins, G., C. Schneider, L. L. Boštjančić, et al. 2023. "The MetaInvert Soil Invertebrate Genome Resource Provides Insights Into Below-Ground Biodiversity and Evolution." *Communications Biology* 6, no. 1: 1241. https://doi.org/10.1038/s42003-023-05621-4.

Collins, G. E., I. D. Hogg, J. R. Baxter, G. Maggs-Kölling, and D. A. Cowan. 2019. "Ancient Landscapes of the Namib Desert Harbor High Levels of Genetic Variability and Deeply Divergent Lineages for Collembola." *Ecology and Evolution* 9, no. 8: 4969–4979. https://doi.org/10.1002/ece3.5103.

Coulson, S. J., I. D. Hodkinson, N. R. Webb, and J. A. Harrison. 2002. "Survival of Terrestrial Soil-Dwelling Arthropods on and in Seawater: Implications for Trans-Oceanic Dispersal." *Functional Ecology* 16, no. 3: 353–356. https://doi.org/10.1046/j.1365-2435.2002.00636.x.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10, no. 2: giab008. https://doi.org/10.1093/gigascience/giab008.

De Coster, W., S. D'Hert, D. T. Schultz, M. Cruts, and C. Van Broeckhoven. 2018. "NanoPack: Visualizing and Processing Long-Read Sequencing Data." *Bioinformatics* 34, no. 15: 15. https://doi.org/10.1093/bioinformatics/bty149.

De Coster, W., and R. Rademakers. 2023. "NanoPack2: Population-Scale Evaluation of Long-Read Sequencing Data." *Bioinformatics* 39, no. 5: btad311. https://doi.org/10.1093/bioinformatics/btad311.

Dexter, R. W. 1943. "*Anurida maritima*: An Important Sea-Shore Scavenger." *Journal of Economic Entomology* 36, no. 5: 5. https://doi.org/10.1093/jee/36.5.797.

Emerson, B. C., F. Cicconardi, P. P. Fanciulli, and P. J. A. Shaw. 2011. "Phylogeny, Phylogeography, Phylobetadiversity and the Molecular Analysis of Biological Communities." *Philosophical Transactions of the Royal Society, B: Biological Sciences* 366, no. 1576: 2391–2402. https://doi.org/10.1098/rstb.2011.0057.

Faria, C. M. A., P. Shaw, and B. C. Emerson. 2019. "Evidence for the Pleistocene Persistence of Collembola in Great Britain." *Journal of Biogeography* 46, no. 7: 1479–1493. https://doi.org/10.1111/jbi.13610.

Gautier, M., R. Vitalis, L. Flori, and A. Estoup. 2022. "F-Statistics Estimation and Admixture Graph Construction With Pool-Seq or Allele Count Data Using the R Package Poolfstat." *Molecular Ecology Resources* 22, no. 4: 1394–1416. https://doi.org/10.1111/1755-0998.13557.

Goto, H. E., and C. Delamare Deboutteville. 1953. "*Anurida bisetosa* Bagnall a Synonym of *A. maritima* (Guérin) (Collemb., Hypogastruridae)." *Entomologist's Monthly Magazine* 89: 249–250.

Guindon, S., J.-F. Dufayard, V. Lefort, M., W. Hordijk, and O. Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59, no. 3: 307–321. https://doi.org/10.1093/sysbio/syq010.

Hardin, G. 1960. "The Competitive Exclusion Principle: An Idea That Took a Century to Be Born has Implications in Ecology, Economics, and Genetics." *Science* 131, no. 3409: 1292–1297. https://doi.org/10.1126/science.131.3409.1292.

Hending, D. 2025. "Cryptic Species Conservation: A Review." *Biological Reviews* 100, no. 1: brv.13139. https://doi.org/10.1111/brv.13139.

Hensel, R., C. Neinhuis, and C. Werner. 2016. "The Springtail Cuticle as a Blueprint for Omniphobic Surfaces." *Chemical Society Reviews* 45, no. 2: 323–341. https://doi.org/10.1039/C5CS00438A.

Hivert, V., R. Leblois, E. J. Petit, M. Gautier, and R. Vitalis. 2018. "Measuring Genetic Differentiation From Pool-Seq Data." *Genetics* 210, no. 1: 315–330. https://doi.org/10.1534/genetics.118.300900.

Hopkin, S. P. 1997. *Biology of Springtails: Insecta: Collembola*. Oxford University Press.

Hopkin, S. P. 2007. *A Key to the Collembola (Springtails) of Britain and Ireland*. 1st ed. Field Studies Council.

Ibrahim, K. M., R. A. Nichols, and G. M. Hewitt. 1996. "Spatial Patterns of Genetic Variation Generated by Different Forms of Dispersal During Range Expansion." *Heredity* 77: 282–291.

Imms, A. D. 1906. *Anurida: Vol. XIII*, edited by W. A. Herdman. Williams and Norgate.

Joosse, E. N. G. 1966. "Some Observations on the Biology of *Anurida maritima* (Guérin) (collembola)." *Zeitschrift für Morphologie und Ökologie der Tiere* 57, no. 3: 320–328. https://doi.org/10.1007/BF00407599.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin. 2017. "ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates." *Nature Methods* 14, no. 6: 587–589. https://doi.org/10.1038/nmeth.4285.

Katoh, K., G. Asimenos, and H. Toh. 2009. "Multiple Alignment of DNA Sequences With MAFFT." In *Methods in Molecular Biology*, edited by D. Posada, vol. 537, 39–64. Humana Press.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780. https://doi.org/10.1093/molbev/mst010.

Katz, A. D., R. Giordano, and F. N. Soto-Adames. 2015. "Operational Criteria for Cryptic Species Delimitation When Evidence Is Limited, as Exemplified by North American Entomobrya (Collembola: Entomobryidae): Collembola Species Delimitation." *Zoological Journal of the Linnean Society* 173, no. 4: 818–840. https://doi.org/10.1111/zoj.12220.

Kimura, M. 1968. "Evolutionary Rate at the Molecular Level." *Nature* 217, no. 5129: 624–626. https://doi.org/10.1038/217624a0.

Knyshov, A., E. R. L. Gordon, and C. Weirauch. 2021. "New Alignment-Based Sequence Extraction Software (ALiBaSeq) and Its Utility for Deep Level Phylogenetics." *PeerJ* 9: e11019. https://doi.org/10.7717/peerj.11019.

Kolmogorov, M., J. Yuan, Y. Lin, and P. A. Pevzner. 2019. "Assembly of Long, Error-Prone Reads Using Repeat Graphs." *Nature Biotechnology* 37, no. 5: 540–546. https://doi.org/10.1038/s41587-019-0072-8.

Koren, S., B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, and A. M. Phillippy. 2017. "Canu: Scalable and Accurate Long-Read Assembly via Adaptive k -Mer Weighting and Repeat Separation." *Genome Research* 27, no. 5: 722–736. https://doi.org/10.1101/gr.215087.116.

Kováč, Ľ., M. Žurovcová, N. Raschmanová, N. Jureková, V. Papáč, and A. Parimuchová. 2023. "Troglomorphic Adaptations on the Northern European Frontier: The Phylogeny of the Cave Pseudosinella (Hexapoda, Collembola) in the Western Carpathians." *Frontiers in Ecology and Evolution* 11: 1169911. https://doi.org/10.3389/fevo.2023.1169911.

Li, H. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics* 34, no. 18: 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lin, M. F., O. Rodeh, J. Penn, et al. 2018. "GLnexus: Joint Variant Calling for Large Cohort Sequencing." *bioRxiv*: 343970. https://doi.org/10.1101/343970.

Liu, B., Y. Shi, J. Yuan, et al. 2013. "Estimation of Genomic Characteristics by Analyzing k-Mer Frequency in De Novo Genome Projects (Version 2)." *arXiv*: 13082012. https://doi.org/10.48550/ARXIV.1308.2012.

Malcicka, M., M. P. Berg, and J. Ellers. 2017. "Ecomorphological Adaptations in Collembola in Relation to Feeding Strategies and Microhabitat." *European Journal of Soil Biology* 78: 82–91. https://doi.org/10.1016/j.ejsobi.2016.12.004.

Misof, B., S. Liu, K. Meusemann, et al. 2014. "Phylogenomics Resolves the Timing and Pattern of Insect Evolution." *Science* 346: 763–767. https://doi.org/10.1126/science.1257570.

Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32, no. 1: msu300. https://doi.org/10.1093/molbev/msu300.

Paradis, E., and K. Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics* 35, no. 3: 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Ponge, J.-F. 2020. "Move or Change, an Eco-Evolutionary Dilemma: The Case of Collembola." *Pedobiologia* 79: 150625. https://doi.org/10.1016/j.pedobi.2020.150625.

Porco, D., A. Bedos, P. Greenslade, et al. 2012. "Challenging Species Delimitation in Collembola: Cryptic Diversity Among Common Springtails Unveiled by DNA Barcoding." *Invertebrate Systematics* 26, no. 6: 470. https://doi.org/10.1071/IS12026.

Potapov, A., B. Bellini, S. Chown, et al. 2020. "Towards a Global Synthesis of Collembola Knowledge—Challenges and Potential Solutions." *Soil Organisms* 92, no. 3: 161–188. https://doi.org/10.25674/SO92ISS3PP161.

Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus, and A. Korobeynikov. 2020. "Using SPAdes De Novo Assembler." *Current Protocols in Bioinformatics* 70, no. 1: e102. https://doi.org/10.1002/cpbi.102.

R Core Team. 2013. "R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (Computer Software)." http://www.R-project.org/.

Ratnasingham, S., C. Wei, D. Chan, et al. 2024. "BOLD v4: A Centralized Bioinformatics Platform for DNA-Based Biodiversity Data." In *DNA Barcoding*, edited by R. DeSalle, vol. 2744, 403–441. Springer US. https://doi.org/10.1007/978-1-0716-3581-0_26.

Rosenberger, M., M. Maraun, S. Scheu, and I. Schaefer. 2013. "Pre- and Post-Glacial Diversifications Shape Genetic Complexity of Soil-Living Microarthropod Species." *Pedobiologia* 56, no. 2: 79–87. https://doi.org/10.1016/j.pedobi.2012.11.003.

Rusek, J. 1998. "Biodiversity of Collembola and Their Functional Role in the Ecosystem." *Biodiversity and Conservation* 7, no. 9: 1207–1219. https://doi.org/10.1023/A:1008887817883.

Schliep, K. P. 2011. "Phangorn: Phylogenetic Analysis in R." *Bioinformatics* 27, no. 4: 592–593. https://doi.org/10.1093/bioinformatics/btq706.

Schuppenhauer, M. M., R. Lehmitz, and W. E. R. Xylander. 2019. "Slow-Moving Soil Organisms on a Water Highway: Aquatic Dispersal and Survival Potential of Oribatida and Collembola in Running Water." *Movement Ecology* 7, no. 1: 20. https://doi.org/10.1186/s40462-019-0165-5.

Sigovini, M., E. Keppel, and D. Tagliapietra. 2016. "Open Nomenclature in the Biodiversity Era." *Methods in Ecology and Evolution* 7, no. 10: 1217–1225. https://doi.org/10.1111/2041-210X.12594.

Simão, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, and E. M. Zdobnov. 2015. "BUSCO: Assessing Genome Assembly and Annotation Completeness With Single-Copy Orthologs." *Bioinformatics* 31, no. 19: 19. https://doi.org/10.1093/bioinformatics/btv351.

Sun, X., D. Yu, Z. Xie, et al. 2020. "Phylomitogenomic Analyses on Collembolan Higher Taxa With Enhanced Taxon Sampling and Discussion on Method Selection." *PLoS One* 15, no. 4: e0230827. https://doi.org/10.1371/journal.pone.0230827.

Thibaud, J.-M., and E. Christian. 1997. "Biodiversity of Interstitial Collembola (Insecta) in Sand Sediments." *European Journal of Soil Biology* 33: 123–127.

Timmermans, M. J. T. N., J. I. Arbea, G. Campbell, M. C. King, A. Prins, and S. Kett. 2022. "Mitochondrial Genome Divergence Supports an Ancient Origin of Circatidal Behaviour in the *Anurida maritima* (Collembola: Neanuridae) Species Group." *Organisms Diversity & Evolution* 22, no. 1: 131–140. https://doi.org/10.1007/s13127-021-00503-1.

Timmermans, M. J. T. N., M. King, D. Purchase, B. J. A. Dickins, T. E. Dickins, and S. Kett. 2024. "Turning the Tide: Rhythmic Aggregation Behaviour in *Anurida maritima* (Collembola) is Entrained by Inundation." *Journal of Experimental Marine Biology and Ecology* 581: 152062. https://doi.org/10.1016/j.jembe.2024.152062.

Timmermans, M. J. T. N., H. Prabha, and S. Kett. 2023. "Wolbachia and Spiroplasma Endosymbionts in the *Anurida maritima* (Collembola) Species Group." *Evolutionary Journal of the Linnean Society* 2, no. 1: 1. https://doi.org/10.1093/evolinnean/kzad001.

Trifinopoulos, J., L.-T. Nguyen, A. von Haeseler, and B. Q. Minh. 2016. "W-IQ-TREE: A Fast Online Phylogenetic Tool for Maximum Likelihood Analysis." *Nucleic Acids Research* 44, no. W1: W1. https://doi.org/10.1093/nar/gkw256.

Van Der Wurff, A. W. G., J. A. Isaaks, G. Ernsting, and N. M. Van Straalen. 2003. "Population Substructures in the Soil Invertebrate *Orchesella cincta*, as Revealed by Microsatellite and TE-AFLP Markers." *Molecular Ecology* 12, no. 6: 1349–1359. https://doi.org/10.1046/j.1365-294X.2003.01811.x.

Vierstraete, A. R., and B. P. Braeckman. 2022. "Amplicon_sorter: A Tool for Reference-Free Amplicon Sorting Based on Sequence Similarity and for Building Consensus Sequences." *Ecology and Evolution* 12, no. 3: e8603. https://doi.org/10.1002/ece3.8603.

von Saltzwedel, H., S. Scheu, and I. Schaefer. 2016. "Founder Events and Pre-Glacial Divergences Shape the Genetic Structure of European Collembola Species." *BMC Evolutionary Biology* 16, no. 1: 1. https://doi.org/10.1186/s12862-016-0719-8.

Wang, H., B. Liu, Y. Zhang, et al. 2020. "Estimation of Genome Size Using k-Mer Frequencies From Corrected Long Reads (Version 1)." *arXiv*: 11817. https://doi.org/10.48550/ARXIV.2003.11817.

Waters, J. M., C. I. Fraser, and G. M. Hewitt. 2013. "Founder Takes All: Density-Dependent Processes Structure Biodiversity." *Trends in Ecology & Evolution* 28, no. 2: 78–85. https://doi.org/10.1016/j.tree.2012.08.024.

Wickham, H., and C. Sievert. 2016. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. Springer.

Zheng, Z., S. Li, J. Su, A. W.-S. Leung, T.-W. Lam, and R. Luo. 2022. "Symphonizing Pileup and Full-Alignment for Deep Learning-Based Long-Read Variant Calling." *Nature Computational Science* 2, no. 12: 797–803. https://doi.org/10.1038/s43588-022-00387-x.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.