Mitochondrial genome divergence supports an ancient origin of circatidal behaviour in the *Anurida maritima* (Collembola: Neanuridae) species group

M.J.T.N. Timmermans[1,2*], J.I. Arbea[3], G. Campbell[1], M.C. King[1], A. Prins[1,4], S. Kett[1]

1: Department of Natural Sciences, Middlesex University, London, UK

2: Department of Life Sciences, The Natural History Museum, London, UK

3: Ría de Solía 3, 39610 Astillero, Spain

4: Current address: Rothamsted Research, Harpenden, UK

Email addresses:

Martijn Timmermans (Corresponding author)	m.timmermans@mdx.ac.uk
Javier I. Arbea	jarbeapo@gmail.com
Gabriela Campbell	GC490@live.mdx.ac.uk
Madeleine King	MK1906@live.mdx.ac.uk
Anneke Prins	anneke.prins@rothamsted.ac.uk
Steve Kett	s.kett@mdx.ac.uk

Keywords: phylogenetics, molecular dating analyses, mitochondrial genome, biological rhythm, nanopore sequencing.

Number of figures: 2 **Number of tables:** 1

Abstract:

Animals of the intertidal zone tolerate substantial environmental fluctuations. Survival under such unstable conditions requires specific adaptations. Several intertidal species have evolved endogenous mechanisms that follow tidal rhythms permitting behavioural alignment with periodic inundation. For example, aggregation behaviour in the springtail *Anurida maritima* (Guérin-Méneville, 1836) (Collembola: Neanuridae) is controlled by a free-running clock with a period of ~12.4 hours. This cosmopolitan species is found in the upper intertidal zone where it forages during low tide. Before high tide specimens aggregate in cracks in the substrate or under rocks to survive inundation. Here we report that the closely related intertidal species *A. bisetosa* Bagnall, 1949 displays a similar endogenously-controlled circatidal behaviour. To obtain a minimum age estimate for this shared derived trait we sequenced the full mitochondrial genome of *A. bisetosa*, which was then used for phylogenetic inference and molecular dating. The mitochondrial genomes of *A. maritima* and *A. bisetosa* are highly divergent. This divergence extends throughout the whole mitochondrial genome and is mirrored by a similar pattern in the nuclear internal transcribed spacer 2 (ITS2) region. Dating analyses suggest the species potentially split more than 40 million years ago. Under the assumption that the endogenously-controlled rhythmic behaviour evolved once in an ancestor of the two species, the trait must be of older age. Use of a single genetic marker and a limited number of fossil calibration points constrains accuracy of the age estimate, but nevertheless it offers a glimpse at otherwise 'intangible' palaeoecological and palaeoethological attributes.

Introduction

Earth's orbital dynamics cause environmental conditions to change in a rhythmic fashion. Its rotation relative to the Sun results in the day-night cycle. The predictability of this 24 hour cycle has allowed organisms to evolve a circadian clock; an endogenous oscillator that aligns metabolic and behavioural changes to the solar day (Edgar et al. 2012). Earth's rotation also underlies the tidal cycle. Its continuous change of orientation relative to the gravitational influences of the Sun and the orbiting Moon causes local rhythmic changes in high and low tide with a periodicity of ~12.4 hours. This second rhythm has a profound effect on organisms of the intertidal zone and exposes them to repeated cycles of inundation (de la Iglesia and Johnson 2013). Several intertidal species (Palmer 1973), including the collembolan *Anurida maritima* (Guérin-Méneville, 1836) (Foster and Moreton 1981), show endogenously controlled rhythmicity in their behaviour closely synchronised to the tides and the changing environmental conditions they create.

A. maritima is a terrestrial, sexually reproducing springtail species (Dallai et al. 1999) found in the upper intertidal zone (Fjellberg 1998). It has a cosmopolitan distribution and inhabits rocky seashores and muddy marshes where it scavenges decaying material during low tide (Dexter 1943; Joosse 1966; Manica et al. 2001). Every tidal cycle, one hour before high tide, the animals aggregate in 'nests' in cracks and cavities near the top of the shoreline to avoid being washed away (Foster and Moreton 1981). Whilst submerged, the animals reside and respire within an air bubble captured by their cuticular hairs (Zinkler et al. 1999). The recurring cycle of active locomotion versus strong aggregation behaviour (inactivity) is directed by an endogenous clock that runs free for at least seven days without the input of zeitgebers (i.e. away from the shore and under laboratory conditions) (Foster and Moreton 1981; McMeechan et al. 2000).

The close association of *A. maritima* and several other collembolan species (e.g. *Podura aquatica*) with aquatic environments has led earlier studies to posit a semi-aquatic origin for the Class, and, because of the position of Collembola in the Tree-of-Life, therefore for all hexapods (including the insects) (Haese 2002). Although the Hexapoda are now thought to have evolved from an aquatic crustacean ancestor (Misof et al. 2014), the notion of an independent semi-aquatic origin for Collembola has been refuted (Leo et al. 2019). Instead, their semi-aquatic life-style has been shown to be a derived trait that emerged repeatedly within the clade (Haese 2002). How long *A. maritima* has inhabited the intertidal zone is not known, but the presence of highly specialised adaptive behaviour could indicate a long ecological connection with seashore environments.

Molecular dating analyses allow the time of past phylogenetic events to be estimated. A recent study using full mitochondrial genome sequence data (Leo et al. 2019), examined divergence events within

78 the Class Collembola and estimated it to have originated more than 400 million years ago. Intriguingly,
79 where the authors included more than one member of a species in their analyses, intra-specific
80 divergence estimates of 47-92 Mya were obtained (Leo et al. 2019). These results prompted further
81 taxonomic investigation that revealed that at least one of these genetically diverse taxa (*Friesea*
82 *antarctica*) consists of three previously unrecognised species (Carapelli et al. 2020). These results
83 indicate that morphologically similar lineages can be erroneously assigned to the same biological
84 species. In this respect it is important to note that *A. maritima* has been described as a species group
85 (Arbea 2001; Sun et al. 2018) and that its taxonomy has been under debate. For example Bagnall
86 described the intertidal species *A. bisetosa* in 1949 (Bagnall 1949), but this was later synonymised with
87 *A. maritima* (Goto and Delamare-Deboutteville 1953). Yet, more recent analyses have confirmed the
88 existence of morphologically divergent lineages and Arbea (2001) reinstated *A. bisetosa*.

89 If *A. bisetosa* is shown to have an endogenous tidal activity rhythm (i.e. it is a shared derived trait),
90 genetic divergence among *A. maritima* and *A. bisetosa* can be used to obtain a minimum age estimate
91 for the origin of this intriguing behaviour. Here we confirm that *A. bisetosa*, like its better known close
92 relative *A. maritima*, displays rhythmic behaviour that follows the tide. We subsequently use full
93 mitochondrial genome sequences to estimate divergence times. By assuming that this specific
94 rhythmic trait evolved once in a shared ancestor, these estimates provide a minimum age for the
95 origin of the endogenously-controlled circatidal behaviour. In addition, as the behaviour is only
96 beneficial in intertidal environments, the species will already have inhabited the foreshore at this point
97 in time. The estimate therefore offers a rare insight into otherwise 'intangible' attributes; an ecological
98 association with a particular habitat and its acquisition of a free-running endogenous clock with a
99 period of ~12.4 hours.

100

101 **Material and Methods**

102 **Sample collection**

103 Animals (*Anurida* sp.) were collected on Lundy, an island in the Bristol Channel (United Kingdom),
104 during low tide in June 2018 and June 2019. Animals were sampled in the Landing Bay area
105 (51°09'46.3"N 4°39'23.8"W) using an entomological aspirator. Animals were transferred to Petri
106 dishes, kept on sea water-moistened filter paper and, following McMeechan et al. (2000), fed on
107 cheese. The petri dishes were sealed with cling film. A subset of animals were taken to mainland UK
108 and stored at -20 ˚C until they were used for molecular analyses or taxonomic identification.
109 Specimens were identified to species as described in Arbea (2001). *A. maritima* specimens were

collected in Wells-next-the-Sea, Norfolk, United Kingdom (52°57'25.4"N 0°51'26.2"E)(see also McMeechan et al. 2000).

Aggregation behaviour

Time-lapse photography was applied to investigate rhythms in aggregation behaviour. Animals from Lundy were placed in a Petri dish ~5 hours before high tide and allowed to acclimatise. Time-lapse photography was started ~1.5 hours before high tide and ran for two tidal cycles. Images were taken every thirty minutes using a Samsung Galaxy S6 32GB mobile phone running the Open Camera v1.47.3 Android app (Mark Harman). Five replicate series were run, three replicates contained 20 animals (start: 08/06/2018 at 12:23; finish: 09/06/2018 at 16:51), two replicates contained 19 animals (start: 09/06/2018 at 13.32; finish: 10/06/2018 at 18:07).

Images were opened using the FIJI (Schindelin et al. 2012) distribution of ImageJ v1.52t (Schindelin et al. 2015). Positions of animals were extracted using the multipoint tool and, for each time point, the mean pairwise Euclidian distance was calculated using a custom PERL script. In several cases it was difficult to discriminate animals within aggregates and therefore to count the full number of animals within a Petri dish (i.e. 19 or 20 depending on the replicate). In these instances unaccounted animals were assumed to be part of the aggregation.

DNA extraction and sequencing

DNA was extracted from two specimens per population each using the QIAGEN DNeasy Blood and Tissue Kit using the manufacturers' recommendations. PCR reactions were then performed to amplify the 5' end of the mitochondrial cytochrome *c* oxidase subunit I gene (COI; barcode region) and the Internal Transcribed Spacer 2 (ITS2). For COI primers HCO-2198: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3' and LCO-1490: 5'-GGTCAACAAATCATAAAGATATTGG-3' (Folmer et al. 1994) were used, for ITS2 primers ITS3: 5'-GCATCGATGAAGAACGCAGC-3' and ITS4: 5'-TCCTCCGCTTATTGATATGC-3' (White et al. 1990) were used. PCR was performed using PCRBIO Taq Mix Red (PCRBiosystems) in 20 µl reaction volumes. For all PCR reactions the following conditions were used: 94 °C for 5 min, followed by 30 cycles of 94 °C for 30 sec, annealing at 50 °C for 45 sec, 72 °C for 45 sec and finally 72 °C for 10 min. PCR products were sequenced in both directions on an ABI3100 sequencer.

High-molecular-weight DNA was extracted from ~100 pooled specimens from Lundy collected in June 2019 using the QIAGEN Genomic-tip procedure using G/20 tips and the Qiagen Genomic DNA Buffer Set. A sequencing library was prepared using the Rapid Sequencing Kit (SQK-RAD004; Oxford Nanopore), which was subsequently sequenced using a MinION and a R9.4.1 flow cell. NanoPlot 1.28.0 (De Coster et al. 2018) was used to obtain summary statistics and plot Quality Control graphs.

Sanger sequence analyses

COI and ITS2 sequence trace files were edited using Geneious 8.0.5 (www.biomatters.com). The COI sequences were compared to all Barcode Records on BOLD (Barcode of Life Database) (Ratnasingham and Hebert 2007) and aligned to all *A. maritima* DNA-barcode sequences publicly available in this database (date accessed: 14/4/2020) and p-distances and Kimura-2 parameter distances (Kimura 1980) calculated using MEGA-X (Kumar et al. 2018). Lundy and Wells-next-the-Sea ITS2 sequences were aligned and sequence divergence of the ITS2 fragments visualised using PopART 1.7 (Leigh and Bryant 2015) using the TCS (Clement et al. 2002) algorithm. ITS2 sequences were supplemented by an ITS2 sequence retrieved from a publicly available transcriptome shotgun assembly (NCBI accession code: GAUE00000000) from *A. maritima* collected from the Island of Texel (The Netherlands) (Misof et al. 2014). Sanger sequences were submitted to NCBI Genbank and are available under accession codes: MT431995-MT431998 and MT434144-MT434146.

Phylogenetic analyses and molecular dating

Oxford Nanopore reads were assembled using CANU version 1.8 (Koren et al. 2017) (settings: genome Size=0.015g -nanopore-raw) and the full mitochondrial genome extracted using BLASTn (Altschul et al. 1997). The obtained contig was trimmed to the length of a single mitochondrial genome. The individual sequencing reads were subsequently mapped back onto the mitochondrial genome using Minimap2 (Li 2018). Genetic diversity was expected as the sequencing library was constructed from DNA that was extracted from ~100 pooled animals. To get an overview of the amount of diversity in the Lundy population, variants were called using SAMtools' mpileup and BCFtools' call functions (Li et al. 2009). The distribution of variants and their sequence coverage was visualised using a custom PERL script using the GD::SVG package (https://metacpan.org/pod/GD::SVG).

A second incomplete mitochondrial genome sequence (consisting of all 13 protein coding genes and 21 tRNAs) was obtained from the *A. maritima* transcriptome shotgun assembly GAUE00000000 (NCBI accession code: GAUE02015563.1). The two mitochondrial genomes were aligned using Geneious

8.0.5. The *A. maritima* transcriptome derived mitochondrial genome was then used to correct remaining homopolymer errors in the protein coding genes of the Nanopore derived mitochondrial genome from the Lundy specimens. Divergence data (p-distances) were calculated for the two mitochondrial genome sequences using a sliding window approach (DNA polymorphism option, window size: 100, step size: 25, excluding sites with gaps) as implemented in DnaSP v6 (Rozas et al. 2017). P-distances were plotted in R against the window midpoint. DnaSP v6 was also used to estimate the number of synonymous and non-synonymous changes and calculate Ka and Ks (with Jukes and Cantor correction) for the coding regions, excluding stop codons.

Protein coding genes were extracted and aligned to protein coding genes of a further 26 mitochondrial genomes (23 Collembola + 3 outgroup species). For each gene the sequences were aligned using the RevTrans2 (Wernersson 2003) web-server and non-conserved regions were excluded using GBlocks 0.91b (Castresana 2000) (using default settings and codon-based analyses). The alignments were then concatenated and all third codon positions removed using Geneious 8.0.5. Mutational saturation was evaluated using DAMBE 7.2.43 (Xia 2018) using the test of Xia (Xia and Lemey 2009; Xia et al. 2003).

The final alignment was partitioned by gene and codon position for a total of 26 partitions. The ModelFinder (Kalyaanamoorthy et al. 2017) algorithm of IQ-TREE 1.6.10 (Nguyen et al. 2015) was applied to merge these 26 partitions into an optimal partition scheme and select appropriate substitution models for phylogenetic inference. Relationships were inferred using Maximum Likelihood and robustness estimated using ultrafast bootstrap (1000 replicates). These calculations were performed in IQ-TREE 1.6.10 (Nguyen et al. 2015).

Molecular dating followed the procedure described by Leo et al. (2019) using the software package BEAST v1.10.4 (Suchard et al. 2018). The dataset used here is identical to that study, with the following three exceptions: 1) scaffolds of Leo et al. (2019) are not included here, 2) the following species were added: *Neelides sp.* (MK431893), *Lepidocyrtus fimetarius* (MK431900), *Metisotoma macnamarai* (MN592792), *Isotomurus maculatus* (MK509021), *Pseudachorutes palmiensis* (MN660051) and *Mesaphorura yosii* (MK431897). 3) One internal calibration point was added: *Protodontella minicornis* was used to provide a minimum age for Neanuroidea. This species was described by Christiansen and Nascimbene (2006) from Cenomanian deposits (100.5 - 93.9 mya). For the split Neanuroidea – Hypogastruroidea we used a gamma prior with shape 1.4, scale 35 and offset 94. The other calibration points were set to what was considered most plausible by Leo et al. (2019), with two outgroup calibration points (the root of the tree: 510 +/- 7 Mya, normal prior; split Collembola and other two hexapod lineages: 485 +/- 6 Mya, normal prior) and two internal calibration points (split Isotomidae and Entomobryidae: gamma prior with shape 1.1, scale 20, offset: 391; split between Lepidocyrtinae

and Orchesellinae: gamma prior with shape 1.1, scale 35, offset 280). A partitioned dataset (using the partition scheme determined by IQ-TREE; see above) was used with unlinked substitution (GTR+ F + I + G) and clock models, an uncorrelated log-normal relaxed clock model, and a Yule tree prior. The IQ-TREE Maximum Likelihood tree was used as starting tree. Two analyses were run: with and without the starting tree being fixed. Two independent runs were performed for each. These were run for 50 million generations with parameters logged every 10,000 generations. Convergence was analysed using Tracer v1.7.1 (Rambaut et al. 2018) and trees of the two independent runs combined with LogCombiner v1.10.4 (discarding the first 10% of generations as burn-in) and then summarised using TreeAnnotator v1.10.4 (Maximum clade credibility tree with median heights). The newly derived mitochondrial genome sequence is available on Genbank under accession code: MT430868. Annotations for the transcriptome derived mitochondrial genome are available, in GFF format, in supplementary file 1.

Results

Species identification and aggregation behaviour

Microscopic examination revealed Lundy specimens to display all morphological characters of *A. bisetosa* as described by Arbea (2001). Aggregation analyses of animals within Petri dishes revealed high levels of aggregation at each high tide (Figure 1). Observed mean distances fluctuated widely over the two cycles, however, indicating that (different levels of) aggregation also occur between high tides (as has also been reported in natural populations (Joosse 1966)). In one case, two clusters were formed, resulting in a large average pairwise distance (Figure 1). We suspect such behaviour is common within their natural habitat where animals access and exploit multiple refugia.

Sequence divergence

The COI sequences were aligned to publicly available *A. maritima* DNA barcode sequences and pairwise distances calculated (Table 1). This revealed comparatively high sequence divergence (>17%) between the *A. bisetosa* sample from Lundy and all the other samples, including those from Wells-next-the-Sea. When the Lundy sequence was compared directly to the BOLD repository, it was shown to be 99.8% identical to four unpublished/private *A. maritima* sequences within the database (database last accessed: 04/05/2021).

ITS2 sequences generated using primer 'ITS4' were unreadable in most cases due to the presence of length variation (indels) and therefore data from only one strand were used for further analysis (318 bp fragment). Inspection of this 318 bp fragment revealed further intra-genomic variation for one of the two Wells-next-the-Sea samples. For this sample, three positions showed heterozygous peaks. The sequence was manually phased, using the information from the other Wells-next-the-Sea sequence (that lacked such heterozygous peaks) and then aligned to the other sequences. The transcriptome derived ITS2 sequence from Texel was 100% identical to sequences obtained from Wells-next-the-Sea specimens. The two Lundy samples differed at 10 positions from the other samples (Supplementary Figure 2) (p-distance: 0.0346; K2P distance: 0.0354).

Phylogenetic analyses and molecular dating

A total of 90655 NanoPore reads were obtained, with a mean quality score of 10.8 and a total length of 311,331,170 bp. Data were assembled and a full length mitochondrial genome with a length of 16,055 bp was obtained. The COI region as sequenced on the Oxford Nanopore MinION was 100% identical to the Lundy Sanger sequence described above. Reads were mapped back to investigate coverage and variation within the genome. Coverage was shown to be more than sufficient with each position covered by ~400 reads. A total of 153 variants were observed (~1% of all positions) which were more or less equally distributed over the genome (Supplementary Figure 3). The genome contains 13 protein coding genes, 22 tRNAs and two rRNAs which are found in the ancestral pancrustacean gene order. This gene order is the one most often found among collembolans and has also been observed for other neanurids (Leo et al. 2019). Comparison of the two *Anurida* mitochondrial genomes (tRNAs and protein coding genes) showed the divergence observed in COI extends throughout the rest of the mitochondrial genome (Supplementary figure 4). The majority of substitutions were synonymous (silent) (Supplementary file 5).

Data from the 13 protein coding genes of the mitochondrial genome were used for phylogenetic analysis. The test of Xia (Xia et al. 2003) indicated that there was little saturation in the dataset (full dataset: Iss 0.31, Iss.c 0.81, $p < 0.001$; first codon positions: Iss 0.40, Iss.c 0.80, $p < 0.001$; second codon positions: Iss 0.24, Iss.c 0.80, $p < 0.001$). The dataset was used to obtain a Maximum Likelihood tree. The branching pattern of this tree (Figure 2) was very similar to the one reported by Leo et al. (2019). Within a monophyletic Collembola, the Neelipleona was the first lineage to diverge. Poduromorpha is sister to Symphypleona + Entomobryomorpha, albeit with low support. Within the Entomobryomorpha, Isotomidae is monophyletic, as is Entomobryidae. Within the Poduromorpha, *A. maritima* and *A. bisetosa* group with the other neanurids. In contrast to Cucini et al. (2020), Leo et al.

(2019) and Sun et al. (2020), *P. aquatica* (Poduridae) is not recovered as sister to Neanuridae, but *Gomphiocephalus hodgsoni* (Hypogastruridae) is.

The dataset was subsequently used for molecular dating analyses. The analyses were performed with and without a fixed starting tree. The topology of the analysis without a fixed starting tree differed from the one obtained using Maximum Likelihood, mainly in the placement of the four (monophyletic) orders. In this Bayesian tree, Entomobryomorpha was the first clade to branch off. Symphypleona recovered as sister to Poduromorpha + Neelipleona. These relationships, however, were recovered with very low support. The relationships within the four orders were very similar to the Maximum Likelihood tree, as were the results from the dating analyses, with an estimate of 44 Mya for the two *Anurida* lineages for the analyses that used a fixed topology, and an estimate of 43 Mya for the analysis where the topology was not fixed.

Discussion

Our analyses revealed a high level of mitochondrial genome divergence within the *Anurida maritima* species group and provide the first molecular evidence that *A. bisetosa* is a 'good' species. It has to be mentioned, however, that the level of divergence reported is not uncommon for intra-specific comparisons within the Class. High levels of divergence have been observed within numerous collembolan species, from temperate, desert and arctic environments (Bennett et al. 2016; Collins et al. 2019; Porco et al. 2012; Timmermans et al. 2005; von Saltzwedel et al. 2016, 2017), even at limited geographic scales (Cicconardi et al. 2010; Collins et al. 2020; Garrick et al. 2004). These studies clearly emphasise the need for additional genetic, morphological and behavioural examination of *A. bisetosa* to further elucidate its status and its relationship to *A. maritima*. It is interesting to note that in contrast to the system presented here, two morphologically distinct intertidal species of the genus *Thalassaphorura* (*T. debilis* and *T. thalassophila*) only showed low levels of genetic divergence (Sun et al. 2018). The lack of divergence between these two species was postulated to be a consequence of their unique association to the tidal environment (Sun et al. 2018). Our observations on the *A. maritima* species group seem to refute this latter hypothesis and suggest that living in the intertidal zone does not prevent divergence *per se*.

Life evolved in the oceans in the Archean, more than 3.5 billion years ago. The first colonisation of land by multicellular organisms occurred much later, and was achieved independently by several lineages (Loron et al. 2019). The Hexapoda originated ~479 million years ago in the Ordovician and first diversified in coastal areas (Misof et al. 2014). *A. maritima* and *A. bisetosa* are not remnants from these early diversification events. Instead, it is generally accepted that the species are much younger and that their semi-aquatic lifestyle evolved secondarily (Haese 2002). Yet it remains unclear when the species adapted to life in the intertidal zone. The genetic divergence observed here, combined with the observation that *A. bisetosa* shows internally-modulated circatidal behaviour, provided a unique opportunity to obtain a minimum age estimate for this event. As the ~12.4 hour rhythmic behaviour likely evolved once in an ancestor of these closely related species and it is beneficial only in the intertidal zone, our observations suggest that both lineages were already adapted to the foreshore environment when they split millions of years ago. Our estimate is only a minimum age estimate and the association with the intertidal zone could have emerged long before the two species split, especially when considering that there are at least eight intertidal species (Joosse 1976) within this speciose genus (Bellinger et al. 1996-2021), whose relationships are not well understood. A more accurate estimate will require a comprehensive phylogeny of the genus that includes representative samples of intertidal and non-intertidal members, supplemented with behavioural observations. Such

a phylogeny will then also reveal whether the endogenous circatidal rhythms evolved once within the wider genus or multiple times independently.

The age estimate presented here is to our knowledge the first one for a secondarily evolved circatidal behaviour. A population genetic study on the marine midge *Clunio marinu* revealed evidence for a postglacial origin of tide specific timing adaptations (Kaiser et al. 2010), but this study did not determine when its genetically determined rhythms first originated. We must, however, emphasise that our current age estimate is based upon the mitochondrial genome, a set of strongly linked genes that behaves as a single genetic marker, has a smaller effective population size and is affected by introgression (Avise 1994; Hurst and Jiggins 2005). As a consequence there is a potential mismatch between nuclear and mitochondrial genomes, and it may be that the latter merely reflects a discordant gene history rather than representing the species' true evolutionary history. In this respect, the relatively lower genetic distances observed in the nuclear ITS2 signal a need for caution. Although data on this marker were in agreement with the mitochondrial genomes, the genetic distance was less pronounced and with a K2P distance of 0.035 similar to distances regularly observed for intra-specific comparisons (Yao et al. 2010). Nuclear genome analyses will be needed to further test the validity of our estimate for the split between *A. maritima* and *A. bisetosa*. Recent, whole genome sequencing efforts applied to Collembola and the development of conserved genome wide markers (e.g. single-copy orthologous sets) (Sun et al. 2020) can assist this.

It is a rare moment when the origin of a specific palaeobiological behaviour not evidenced by morphology can be ascribed a date, however putative. Carnivory may be implied by dentition, flight by wings, even herding may be revealed via ichnofacies and prey choice via coprolite. Other behaviours are 'soft' – they leave no fossil imprint. Were fossil *A. maritima* and *A. bisetosa* to be found, their activity patterns could not be discerned via morphological analyses. Independently-running circatidal behaviour is a 'soft behaviour', yet in this case its origins might just be discernible; not by morphology but via molecules. *Anurida's* molecules offer one timepiece upon which to measure the origin of another, the circatidal clock.

Acknowledgements

We thank the Lundy warden Dean W. Jones for giving permission to collect *Anurida sp.* in the Lundy Landing Bay area. We would also like to thank Tom Dickins for organising the Middlesex University Lundy field trip and the Faculty of Science and Technology for financially supporting it. The authors declare no conflict of interest.

Data availability

The full mitochondrial genome and COI sequences generated and/or analysed during this study are available in the GenBank (https://www.ncbi.nlm.nih.gov/genbank/; see material and methods section for accession codes) or the BOLD repository (https://www.boldsystems.org/). All other datasets generated or used are available from the corresponding author on request.

References

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, W. Miller, & 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 (GUÉRIN, 1839) (COLLEMBOLA: NEANURIDAE) EN LA PENÍNSULA IBÉRICA. *Revta. Aragon. Ent*, (9), 37–42.

Avise, J. C. (1994). *Molecular markers, natural history and evolution*. New York: Chapman & Hall.

Bagnall, R. S. (1949). Notes on British Collembola. *Ent. Mon. Mag.*, *85*(1017), 51–61.

Bellinger, P. F., Christiansen, K. A., & Janssens, F. (1996). Checklist of the Collembola of the World. http://www.collembola.org. Accessed 11 May 2021

Bennett, K. R., Hogg, I. D., Adams, B. J., & Hebert, P. D. N. (2016). High levels of intraspecific genetic divergences revealed for Antarctic springtails: evidence for small-scale isolation during Pleistocene glaciation. *Biological Journal of the Linnean Society*, *119*(1), 166–178. https://doi.org/10.1111/bij.12796

Carapelli, A., Cucini, C., Fanciulli, P. P., Frati, F., Convey, P., & Nardi, F. (2020). Molecular Comparison among Three Antarctic Endemic Springtail Species and Description of the Mitochondrial Genome of Friesea gretae (Hexapoda, Collembola). *Diversity*, *12*(12), 450. https://doi.org/10.3390/d12120450

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*, 540–552.

Christiansen, K., & Nascimbene, P. (2006). Collembola (Arthropoda, Hexapoda) from the mid Cretaceous of Myanmar (Burma). *Cretaceous Research*, *27*(3), 318–363. https://doi.org/10.1016/j.cretres.2005.07.003

Cicconardi, F., Nardi, F., Emerson, B. C., Frati, F., & Fanciulli, P. P. (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*(2), 386–400. https://doi.org/10.1111/j.1365-294X.2009.04457.x

Clement, M., Snell, Q., Walke, P., Posada, D., & Crandall, K. (2002). TCS: estimating gene genealogies. In *Proceedings 16th International Parallel and Distributed Processing Symposium* (p. 7 pp). Presented at the Proceedings 16th International Parallel and Distributed Processing Symposium. IPDPS 2002, Ft. Lauderdale, FL: IEEE. https://doi.org/10.1109/IPDPS.2002.1016585

Collins, G. E., Hogg, I. D., Baxter, J. R., Maggs-Kölling, G., & Cowan, D. A. (2019). Ancient landscapes of the Namib Desert harbor high levels of genetic variability and deeply divergent lineages for Collembola. *Ecology and Evolution*, *9*(8), 4969–4979. https://doi.org/10.1002/ece3.5103

Collins, G. E., Hogg, I. D., Convey, P., Sancho, L. G., Cowan, D. A., Lyons, W. B., et al. (2020). Genetic diversity of soil invertebrates corroborates timing estimates for past collapses of the West Antarctic Ice Sheet. *Proceedings of the National Academy of Sciences*, *117*(36), 22293–22302. https://doi.org/10.1073/pnas.2007925117

Cucini, C., Fanciulli, P. P., Frati, F., Convey, P., Nardi, F., & Carapelli, A. (2020). Re-Evaluating the Internal Phylogenetic Relationships of Collembola by Means of Mitogenome Data. *Genes*, *12*(1), 44. https://doi.org/10.3390/genes12010044

Dallai, R., Fanciulli, P. P., & Frati, F. (1999). Chromosome elimination and sex determination in springtails (Insecta, Collembola). *The Journal of Experimental Zoology*, *285*(3), 215–225. https://doi.org/10.1002/(sici)1097-010x(19991015)285:3<215::aid-jez4>3.0.co;2-5

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics*, *34*(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

de la Iglesia, H. O., & Johnson, C. H. (2013). Biological Clocks: Riding the Tides. *Current Biology*, *23*(20), R921–R923. https://doi.org/10.1016/j.cub.2013.09.006

399 Dexter, R. W. (1943). Anurida maritima: An Important Sea-Shore Scavenger. *Journal of Economic*
400 *Entomology*, *36*(5), 797–797. https://doi.org/10.1093/jee/36.5.797
401 Edgar, R. S., Green, E. W., Zhao, Y., van Ooijen, G., Olmedo, M., Qin, X., et al. (2012). Peroxiredoxins
402 are conserved markers of circadian rhythms. *Nature*, *485*(7399), 459–464.
403 https://doi.org/10.1038/nature11088
404 Fjellberg, A. (1998). *The Collembola of Fennoscandia and Denmark, Part I: Poduromorpha* (Vol. 35).
405 Leiden, The Netherlands: Brill.
406 Folmer, O., M. Black, W. Hoeh, R. Lutz, & R. Vrijenhoek. (1994). DNA primers for amplification of
407 mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.
408 *Molecular Marine Biology and Biotechnology*, *3*, 294–299.
409 Foster, W. A., & Moreton, R. B. (1981). Synchronization of activity rhythms with the tide in a
410 saltmarsh collembolan Anurida maritima. *Oecologia*, *50*(2), 265–270.
411 https://doi.org/10.1007/BF00348049
412 Garrick, R. C., Sands, C. J., Rowell, D. M., Tait, N. N., Greenslade, P., & Sunnucks, P. (2004).
413 Phylogeography recapitulates topography: very fine-scale local endemism of a saproxylic
414 'giant' springtail at Tallaganda in the Great Dividing Range of south-east Australia:
415 PHYLOGEOGRAPHY OF A SAPROXYLIC SPRINGTAIL. *Molecular Ecology*, *13*(11), 3329–3344.
416 https://doi.org/10.1111/j.1365-294X.2004.02340.x
417 Goto, H. E., & Delamare-Deboutteville, C. (1953). Anurida bisetosa Bagnall a synonym of A. maritima
418 (Guérin)(Collemb., Hypogastruridae). *The Entomologist's*, *89*, 249–250.
419 Haese, C. A. D. (2002). Were the first springtails semi-aquatic? A phylogenetic approach by means of
420 28S rDNA and optimization alignment. *Proceedings of the Royal Society of London. Series B:*
421 *Biological Sciences*, *269*(1496), 1143–1151. https://doi.org/10.1098/rspb.2002.1981
422 Hurst, G. D. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population,
423 phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings*

of the Royal Society B: Biological Sciences, *272*(1572), 1525–1534. https://doi.org/10.1098/rspb.2005.3056

Joosse, E. N. G. (1966). Some observations on the biology of Anurida maritima (Gu�rin), (collembola). *Zeitschrift f�r Morphologie und Ökologie der Tiere*, *57*(3), 320–328. https://doi.org/10.1007/BF00407599

Joosse, E. N. G. (1976). Littoral apterygotes (Collembola and Thysanura). In L. Cheng (Ed.), *Marine insects* (pp. 151–186). North-Holland Publishing Company.

Kaiser, T. S., Neumann, D., Heckel, D. G., & Berendonk, T. U. (2010). Strong genetic differentiation and postglacial origin of populations in the marine midge Clunio marinus (Chironomidae, Diptera): GENETIC DIFFERENTIATION IN CLUNIO MARINUS. *Molecular Ecology*, *19*(14), 2845–2857. https://doi.org/10.1111/j.1365-294X.2010.04706.x

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589. https://doi.org/10.1038/nmeth.4285

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. https://doi.org/10.1007/BF01731581

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Leigh, J. W., & Bryant, D. (2015). POPART : full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

Leo, C., Carapelli, A., Cicconardi, F., Frati, F., & Nardi, F. (2019). Mitochondrial Genome Diversity in Collembola: Phylogeny, Dating and Gene Order. *Diversity*, *11*(9), 169. https://doi.org/10.3390/d11090169

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Loron, C. C., François, C., Rainbird, R. H., Turner, E. C., Borensztajn, S., & Javaux, E. J. (2019). Early fungi from the Proterozoic era in Arctic Canada. *Nature*, *570*(7760), 232–235. https://doi.org/10.1038/s41586-019-1217-0

Manica, A., McMeechan, F. K., & Foster, W. A. (2001). An aggregation pheromone in the intertidal collembolan *Anurida maritima*. *Entomologia Experimentalis et Applicata*, *99*(3), 393–395. https://doi.org/10.1046/j.1570-7458.2001.00840.x

McMeechan, F. K., Manica, A., & Foster, W. A. (2000). Rhythms of activity and foraging in the intertidal insect *Anurida maritima* : coping with the tide. *Journal of the Marine Biological Association of the United Kingdom*, *80*(1), 189–190. https://doi.org/10.1017/S0025315499001770

Misof, B., Liu, S., Meusemann, K., Peters, R. S., Donath, A., Mayer, C., et al. (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science*, *346*(6210), 763–767. https://doi.org/10.1126/science.1257570

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Palmer, J. D. (1973). Tidal rhythms: the clock control of the rhythmic physiology of marine organisms. *Biological Reviews*, *48*(3), 377–418. https://doi.org/10.1111/j.1469-185X.1973.tb01008.x

Porco, D., Potapov, M., Bedos, A., Busmachiu, G., Weiner, W. M., Hamra-Kroua, S., & Deharveng, L. (2012). Cryptic Diversity in the Ubiquist Species Parisotoma notabilis (Collembola, Isotomidae): A Long-Used Chimeric Species? *PLoS ONE*, *7*(9), e46056. https://doi.org/10.1371/journal.pone.0046056

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Systematic Biology*, *67*(5), 901–904. https://doi.org/10.1093/sysbio/syy032

Ratnasingham, S., & Hebert, P. D. N. (2007). BARCODING: bold: The Barcode of Life Data System (http://www.barcodinglife.org): BARCODING. *Molecular Ecology Notes*, *7*(3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302. https://doi.org/10.1093/molbev/msx248

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schindelin, J., Rueden, C. T., Hiner, M. C., & Eliceiri, K. W. (2015). The ImageJ ecosystem: An open platform for biomedical image analysis: T HE I MAGEJ E COSYSTEM. *Molecular Reproduction and Development*, *82*(7–8), 518–529. https://doi.org/10.1002/mrd.22489

Suchard, M. A., Lemey, P., Baele, G., Ayres, D. L., Drummond, A. J., & Rambaut, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution*, *4*(1). https://doi.org/10.1093/ve/vey016

Sun, X., Bedos, A., & Deharveng, L. (2018). Unusually low genetic divergence at COI barcode locus between two species of intertidal *Thalassaphorura* (Collembola: Onychiuridae). *PeerJ*, *6*, e5021. https://doi.org/10.7717/peerj.5021

Sun, X., Ding, Y., Orr, M. C., & Zhang, F. (2020). Streamlining universal single-copy orthologue and ultraconserved element design: A case study in Collembola. *Molecular Ecology Resources*, *20*(3), 1755–0998.13146. https://doi.org/10.1111/1755-0998.13146

Sun, X., Yu, D., Xie, Z., Dong, J., Ding, Y., Yao, H., & Greenslade, P. (2020). Phylomitogenomic analyses on collembolan higher taxa with enhanced taxon sampling and discussion on method selection. *PLOS ONE*, *15*(4), e0230827. https://doi.org/10.1371/journal.pone.0230827

Timmermans, M. J. T. N., Ellers, J., Marien, J., Verhoef, S. C., Ferwerda, E. B., & Van Straalen, N. M. (2005). Genetic structure in Orchesella cincta (Collembola): strong subdivision of European populations inferred from mtDNA and AFLP markers. *Molecular Ecology*, *14*, 2017–2024.

von Saltzwedel, H., Scheu, S., & Schaefer, I. (2016). Founder events and pre-glacial divergences shape the genetic structure of European Collembola species. *BMC Evolutionary Biology*, *16*(1), 148. https://doi.org/10.1186/s12862-016-0719-8

von Saltzwedel, H., Scheu, S., & Schaefer, I. (2017). Genetic structure and distribution of Parisotoma notabilis (Collembola) in Europe: Cryptic diversity, split of lineages and colonization patterns. *PLOS ONE*, *12*(2), e0170909. https://doi.org/10.1371/journal.pone.0170909

Wernersson, R. (2003). RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Research*, *31*(13), 3537–3539. https://doi.org/10.1093/nar/gkg609

White, T. J., T. Bruns, S. Lee, & J. Taylor. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In M. A. Innis, D. H. Gelfand, J. Sninsky, & T. J. White (Eds.), *PCR protocols: A guide to methods and applications* (pp. 315–322). San Diego: Academic Press.

524 Xia, X., & Lemey, P. (2009). Assessing substitution saturation with DAMBE. In *The Phylogenetic*
525 *Handbook: A Practical Approach to DNA and Protein Phylogeny* (2nd ed., pp. 615–630).
526 Cambridge University Press.
527 Xia, Xuhua. (2018). DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and
528 Evolution. *Molecular Biology and Evolution*, *35*(6), 1550–1552.
529 https://doi.org/10.1093/molbev/msy073
530 Xia, Xuhua, Xie, Z., Salemi, M., Chen, L., & Wang, Y. (2003). An index of substitution saturation and its
531 application. *Molecular phylogenetics and evolution*, *26*(1).
532 Yao, H., Song, J., Liu, C., Luo, K., Han, J., Li, Y., et al. (2010). Use of ITS2 Region as the Universal DNA
533 Barcode for Plants and Animals. *PLoS ONE*, *5*(10), e13102.
534 https://doi.org/10.1371/journal.pone.0013102
535 Zinkler, D., Ruessbeck, R., Biefang, M., & Baumgaertl, H. (1999). Intertidal respiration of *Anurida*
536 *maritima* (Collembola: Neanuridae). *EJE*, *96*(2), 205–209.
537
538

Figure captions:

Figure 1: Aggregation behaviour in *Anurida bisetosa*. *A. bisetosa* is a collembolan from the intertidal zone. It has 5+5 ocelli (top left inset). Bottom graph: Aggregation behaviour over time. Mean distance between animals was calculated at 30 minutes intervals. Five replicates were run for ~27 hours each (each indicated using a different colour. Black vertical lines indicates time of local high tide. A) typical distribution during low tide, B) typical distribution during high tide, C) aggregation in two clusters during high tide. Animals were within 90 millimetre Petri dishes.

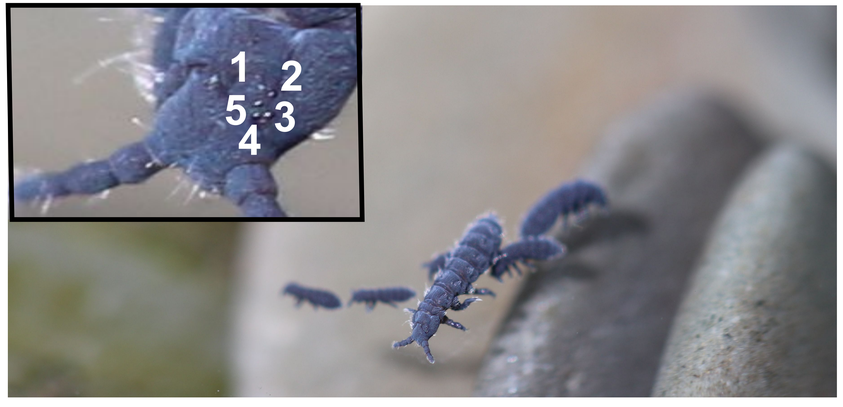
Figure 2: Phylogenetic and molecular dating analyses. Top: Bayesian divergence time estimation using fixed starting tree. The starting tree is shown on the right and is a Maximum Likelihood (ML) tree obtained using IQ-TREE. Numbers at the nodes of this ML tree are ultrafast bootstrap support values (%), only values <100 are given. Bottom: Divergence time estimation using the same starting tree, but with Bayesian tree optimization. Numbers at the nodes indicate posterior probabilities, only values of <1.0 are given. NCBI accession codes are given at the branch labels. Numbers under the time scales indicate 95% Highest Posterior Density of estimated age of *Anurida maritima and A. bisetosa* (in My). The four orders are indicated with differently coloured branches.

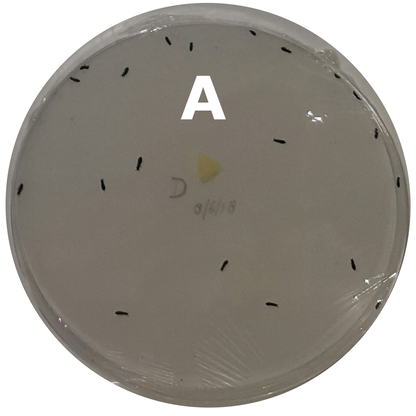
555 **Table 1: COI sequence variation.** All publicly available *Anurida maritima* COI DNA barcodes were
556 downloaded from the BarCode of Life Database (BOLD), aligned and pairwise p-distances (above the
557 diagonal) and K2P distances (under the diagonal) calculated. Identical sequences were removed
558 (Sequence GBCO1851-14 is identical to SDP258013-15 (USA); ECHUB341-11 is identical to ECHUB343-
559 11 and ECHUB339-11 (France). Comparisons with *A. bisetosa* (Lundy) data are given in bold. Note that
560 comparisons with ECHUB339-11 also show relative high divergence, albeit much lower than the
561 comparisons with the Lundy sample. *) This study.

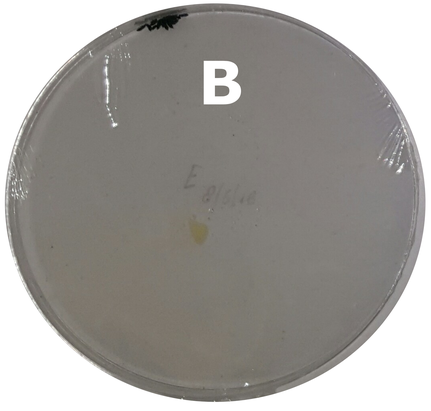
562

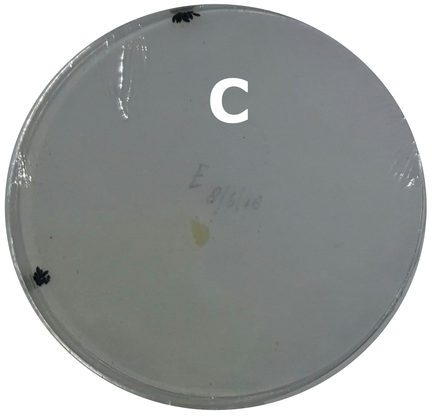
	A. bisetosa Lundy*	*A. maritima* Wells-next-the-Sea* (UK)	*A. maritima* Wells-next-the-Sea* (UK)	*A. maritima* GBCO1656-13 (France)	*A. maritima* NORCO164-14 (Norway)	*A. maritima* NORCO165-14 (Norway)	*A. maritima* GBCO1851-14 (USA)	*A. maritima* ECHUB339-11 (France)
Lundy*		**0.179**	**0.179**	**0.181**	**0.176**	**0.176**	**0.173**	**0.166**
Wells-next-the-Sea*	**0.214**		0.005	0.006	0.008	0.005	0.011	0.079
Wells-next-the-Sea*	**0.214**	0.005		0.005	0.006	0.003	0.009	0.078
GBCO1656-13	**0.216**	0.006	0.005		0.008	0.005	0.011	0.079
NORCO164-14	**0.209**	0.008	0.006	0.008		0.003	0.009	0.074
NORCO165-14	**0.209**	0.005	0.003	0.005	0.003		0.006	0.074
GBCO1851-14	**0.205**	0.011	0.009	0.011	0.009	0.006		0.068
ECHUB339-11	**0.194**	0.086	0.084	0.086	0.080	0.080	0.073	

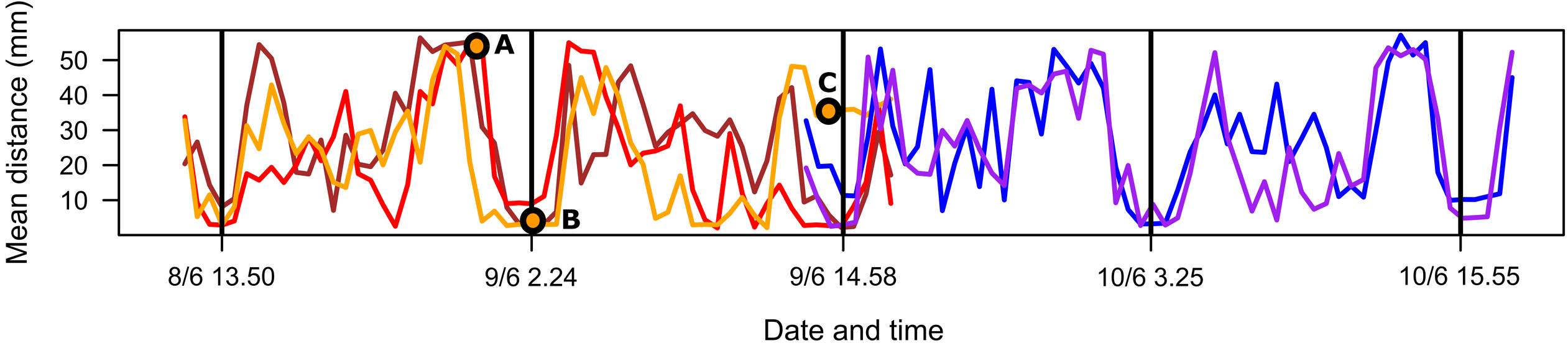
563

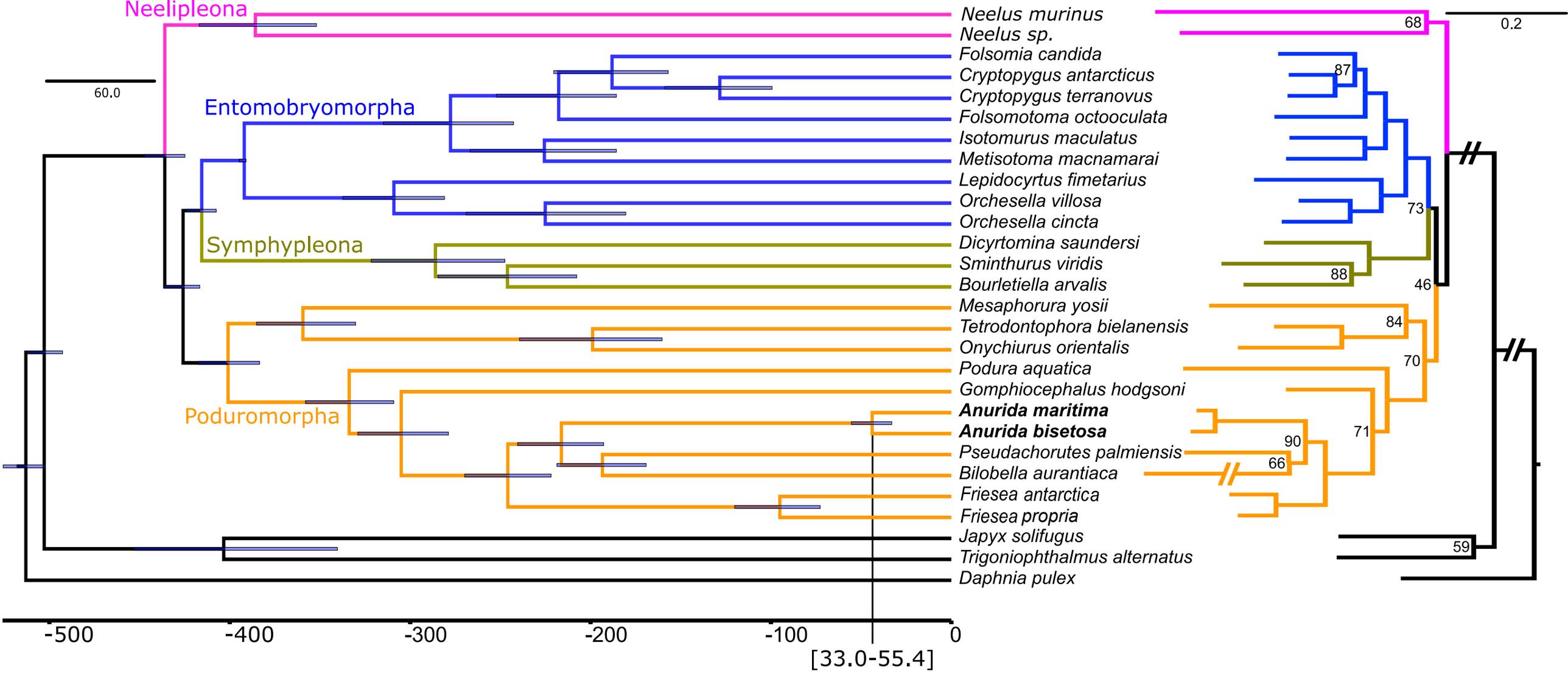

1
2
5
3
4

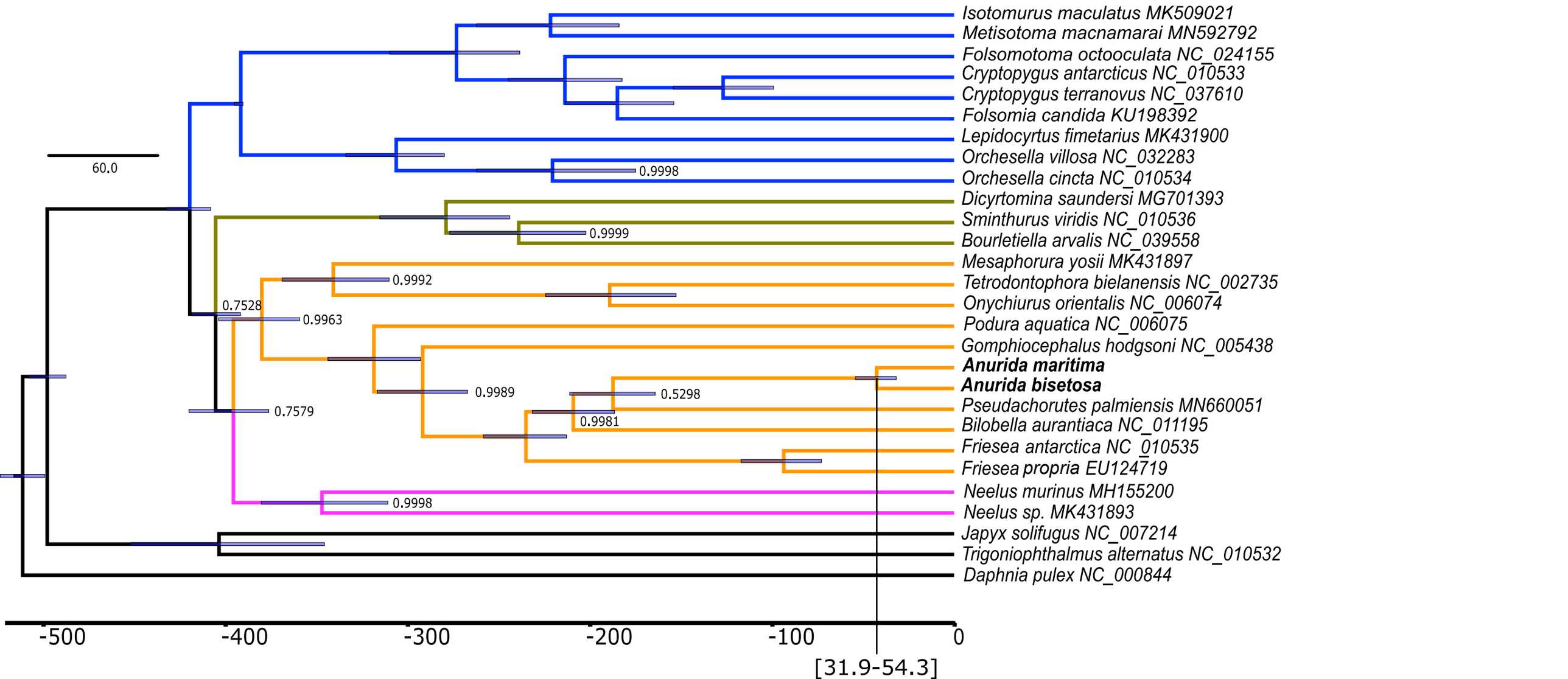

A


B


C

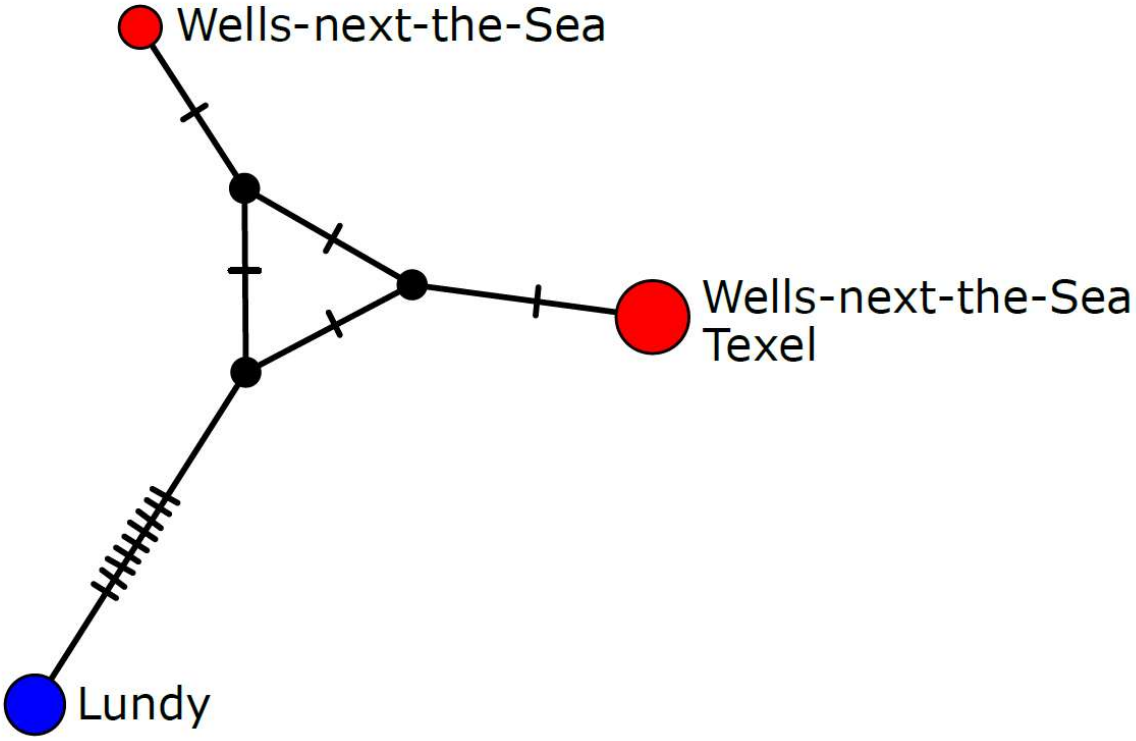

Mean distance (mm)
50
40
30
20
10
A
B
C
8/6 13.50
9/6 2.24
9/6 14.58
10/6 3.25
10/6 15.55
Date and time


Neelipleona
Entomobryomorpha
Symphypleona
Poduromorpha
60.0
Neelus murinus
Neelus sp.
Folsomia candida
Cryptopygus antarcticus
Cryptopygus terranovus
Folsomotoma octooculata
Isotomurus maculatus
Metisotoma macnamarai
Lepidocyrtus fimetarius
Orchesella villosa
Orchesella cincta
Dicyrtomina saundersi
Sminthurus viridis
Bourletiella arvalis
Mesaphorura yosii
Tetrodontophora bielanensis
Onychiurus orientalis
Podura aquatica
Gomphiocephalus hodgsoni
Anurida maritima
Anurida bisetosa
Pseudachorutes palmiensis
Bilobella aurantiaca
Friesea antarctica
Friesea propria
Japyx solifugus
Trigoniophthalmus alternatus
Daphnia pulex
68
87
73
88
46
84
70
71
90
66
59
0.2
-500
-400
-300
-200
-100
0
[33.0-55.4]

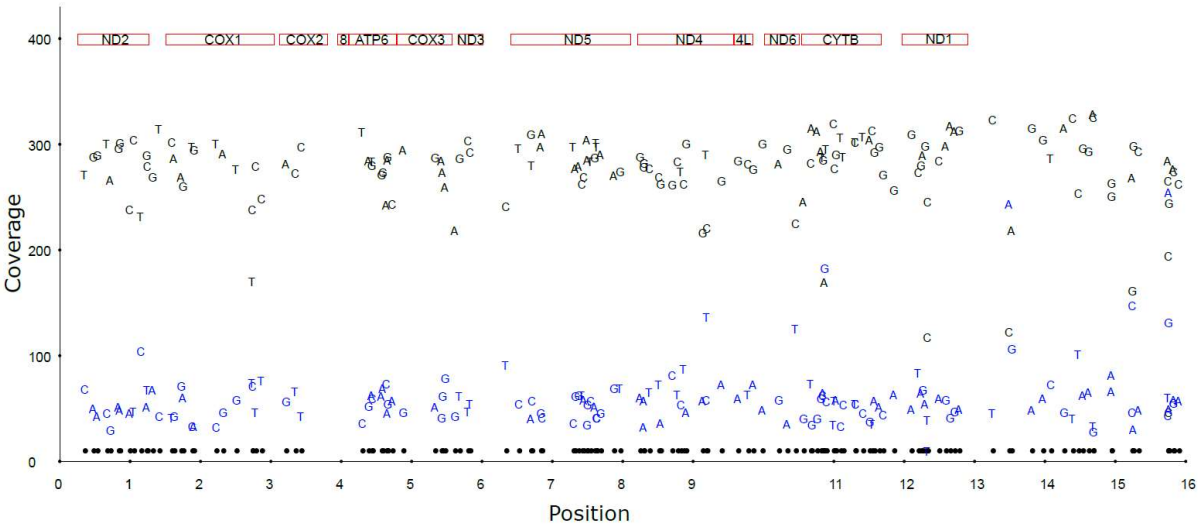

60.0
Isotomurus maculatus MK509021
Metisotoma macnamarai MN592792
Folsomotoma octooculata NC_024155
Cryptopygus antarcticus NC_010533
Cryptopygus terranovus NC_037610
Folsomia candida KU198392
Lepidocyrtus fimetarius MK431900
Orchesella villosa NC_032283
Orchesella cincta NC_010534
Dicyrtomina saundersi MG701393
Sminthurus viridis NC_010536
Bourletiella arvalis NC_039558
Mesaphorura yosii MK431897
Tetrodontophora bielanensis NC_002735
Onychiurus orientalis NC_006074
Podura aquatica NC_006075
Gomphiocephalus hodgsoni NC_005438
Anurida maritima
Anurida bisetosa
Pseudachorutes palmiensis MN660051
Bilobella aurantiaca NC_011195
Friesea antarctica NC_010535
Friesea propria EU124719
Neelus murinus MH155200
Neelus sp. MK431893
Japyx solifugus NC_007214
Trigoniophthalmus alternatus NC_010532
Daphnia pulex NC_000844
0.9998
0.9999
0.9992
0.7528
0.9963
0.9989
0.5298
0.7579
0.9981
0.9998
-500
-400
-300
-200
-100
0
[31.9-54.3]

Supplementary File 1: Annotations for *A. maritima* partial mitochondrial genome (GenBank accession code: GAUE02015563.1) in GFF format.

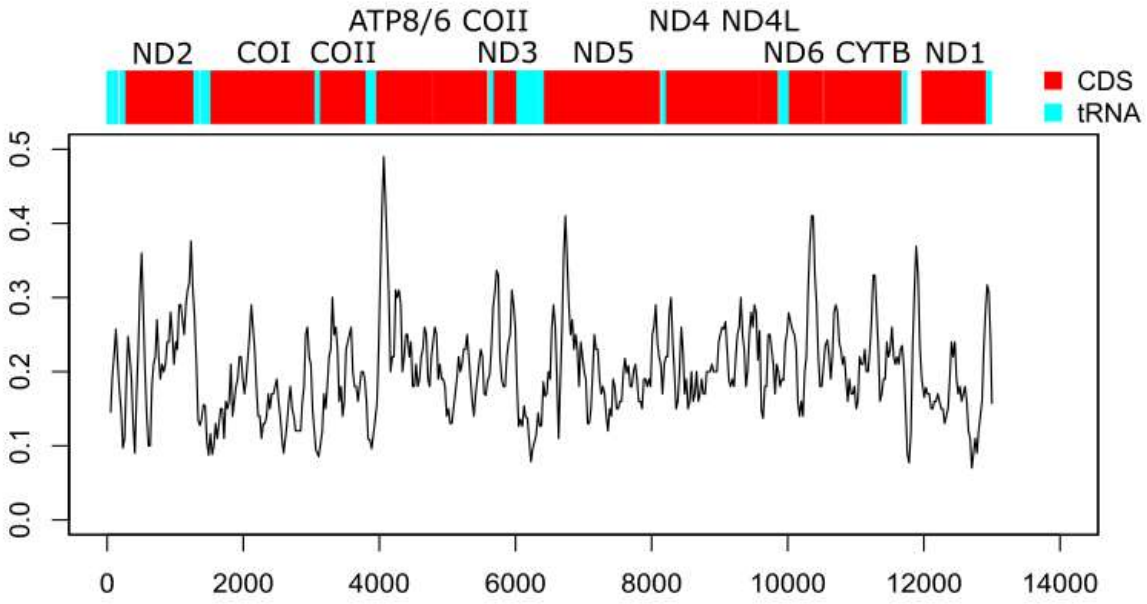
```
##gff-version 3
##source-version geneious 2021.0.3
##sequence-region     GAUE02015563.1 1       13671
GAUE02015563.1 Geneious       CDS     7060    8762    .       -       .       Name=ND5 CDS
GAUE02015563.1 Geneious       CDS     2175    3707    .       +       .       Name=COX1 CDS
GAUE02015563.1 Geneious       CDS     8857    10221   .       -       .       Name=ND4 CDS
GAUE02015563.1 Geneious       CDS     11200   12333   .       +       .       Name=CYTB CDS
GAUE02015563.1 Geneious       CDS     914     1921    .       +       .       Name=ND2 CDS
GAUE02015563.1 Geneious       CDS     12618   13553   .       -       .       Name=ND1 CDS
GAUE02015563.1 Geneious       CDS     5443    6231    .       +       .       Name=COX3 CDS
GAUE02015563.1 Geneious       CDS     3769    4449    .       +       .       Name=COX2 CDS
GAUE02015563.1 Geneious       CDS     4758    5432    .       +       .       Name=ATP6 CDS
GAUE02015563.1 Geneious       CDS     10683   11171   .       +       .       Name=ND6 CDS
GAUE02015563.1 Geneious       CDS     6319    6664    .       +       .       Name=ND3 CDS
GAUE02015563.1 Geneious       CDS     10229   10495   .       -       .       Name=ND4L CDS
GAUE02015563.1 Geneious       CDS     4597    4764    .       +       .       Name=ATP8 CDS
GAUE02015563.1 Geneious       tRNA    4455    4532    .       +       .       Name=tRNALys
GAUE02015563.1 Geneious       tRNA    738     808     .       -       .       Name=tRNAGln
GAUE02015563.1 Geneious       tRNA    843     913     .       +       .       Name=tRNAMet
GAUE02015563.1 Geneious       tRNA    6665    6731    .       +       .       Name=tRNAAla
GAUE02015563.1 Geneious       tRNA    2037    2108    .       -       .       Name=tRNACys
GAUE02015563.1 Geneious       tRNA    10582   10651   .       -       .       Name=tRNAPro
GAUE02015563.1 Geneious       tRNA    1931    2002    .       +       .       Name=tRNATrp
GAUE02015563.1 Geneious       tRNA    651     718     .       +       .       Name=tRNAIle
GAUE02015563.1 Geneious       tRNA    6791    6857    .       +       .       Name=tRNAAsn
GAUE02015563.1 Geneious       tRNA    6856    6924    .       +       .       Name=tRNASer
GAUE02015563.1 Geneious       tRNA    2107    2173    .       -       .       Name=tRNATyr
GAUE02015563.1 Geneious       tRNA    10519   10584   .       +       .       Name=tRNAThr
GAUE02015563.1 Geneious       tRNA    13577   13642   .       -       .       Name=tRNALeu
GAUE02015563.1 Geneious       tRNA    6729    6793    .       +       .       Name=tRNAArg
GAUE02015563.1 Geneious       tRNA    4532    4596    .       +       .       Name=tRNAAsp
GAUE02015563.1 Geneious       tRNA    6996    7060    .       -       .       Name=tRNAPhe
GAUE02015563.1 Geneious       tRNA    3720    3768    .       +       .       Name=tRNALeu
GAUE02015563.1 Geneious       tRNA    6927    6985    .       +       .       Name=tRNAGlu
GAUE02015563.1 Geneious       tRNA    12352   12409   .       +       .       Name=tRNASer
GAUE02015563.1 Geneious       tRNA    6257    6314    .       +       .       Name=tRNAGly
GAUE02015563.1 Geneious       tRNA    8797    8847    .       -       .       Name=tRNAHis
```



Supplementary File 2: Network showing relationships among ITS2 haplotypes. The network was generated using the TCS algorithm as implemented in PopArt. Colour coding: blue circles: *Anurida bisetosa*; haplotype from Lundy, red circles: *A. maritima*; haplotypes from Wells-next-the-Sea and Texel, black circles: putative haplotypes not present in the dataset. Hatches on connecting branches represent single nucleotide substitutions.



Supplementary File 3: Coverage of called variants. *Anurida bisetosa* mitochondrial genome coverage and position of variants with reference alleles given in black and alternative alleles in blue. Horizontal axis: mitochondrial genome position in 1000 bp. Location of protein coding genes is given at the top (red boxes). In total 153 variants were called (~1%).



Supplementary File 4: Mitochondrial genome divergence *Anurida bisetosa* – *A. maritima*. Sliding window p-distance (window size: 100, step size:25) plotted against mitochondrial genome position. The position of the 13 protein coding genes and of 21 tRNAs are given on top.

Supplementary File 5: Estimated number of synonymous and non-synonymous changes and corresponding Ka and Ks

Gene	Gene length	Synonymous positions	Synonymous differences	Ks	Non-Synonymous Positions	Non-Synonymous Differences	Ka	Ka/Ks
ND1	936	224.83	118.00	0.90	708.17	39.00	0.06	0.06
ND2	1008	242.50	157.00	1.49	762.50	79.00	0.11	0.07
ND3	346	85.00	61.50	2.51	260.00	26.50	0.11	0.04
ND4	1365	329.5	213	1.48	1032.5	82	0.08	0.06
ND4L	267	65.83	38.00	1.10	198.17	15.00	0.08	0.07
ND5	1703	418	247.83	1.17	1283	99.17	0.08	0.07
ND6	489	119.67	70.83	1.17	366.33	44.17	0.13	0.11
ATP6	675	168.00	125.33	3.93	504.00	27.67	0.06	0.01
ATP8	168	39.67	32.50	n.a.	125.33	27.50	0.26	n.a.
COX1	1533	374.67	232.00	1.31	1155.33	20.00	0.02	0.01
COX2	681	151.33	126.50	n.a.	526.67	10.50	0.02	n.a.
COX3	789	190.33	132.00	1.94	595.67	14.00	0.02	0.01
CYTB	1133	267.33	211.50	n.a.	863.67	37.50	0.04	n.a.

n.a.: not applicable. JC correction cannot be computed as the proportion of differences between the two sequences is higher than 75 percent.