Mitochondrial genomes of the key zooplankton copepods Arctic Calanus glacialis and North Atlantic Calanus finmarchicus with the longest crustacean non-coding regions

Agata Weydmann, Aleksandra Przyłucka, Marek Lubośny, Katarzyna S. Walczyńska, Ester A. Serrão, Gareth A. Pearson & Artur Burzyński

We determined the nearly complete mitochondrial genomes of the Arctic Calanus glacialis and its North Atlantic sibling Calanus finmarchicus, which are key zooplankton components in marine ecosystems. The sequenced part of C. glacialis mitogenome is 27,342 bp long and consists of two contigs, while for C. finmarchicus it is 29,462 bp and six contigs, what makes them the longest reported copepod mitogenomes. The typical set of metazoan mitochondrial genes is present in these mitogenomes, although the non-coding regions (NCRs) are unusually long and complex. The mitogenomes of the closest species C. glacialis and C. finmarchicus, followed by the North Pacific C. sinicus, are structurally similar and differ from the much more typical of deep-water, Arctic C. hyperboreus. This evolutionary trend for the expansion of NCRs within the Calanus mitogenomes increases mitochondrial DNA density, what resulted in its similar density to the nuclear genome. Given large differences in the length and structure of C. glacialis and C. finmarchicus mitogenomes, we conclude that the species are genetically distinct and thus cannot hybridize. The molecular resources presented here: the mitogenomic and rDNA sequences, and the database of repetitive elements should facilitate the development of genetic markers suitable in pursuing evolutionary research in copepods.

Copepods of the genus Calanus are widely spread in the world ocean and are particularly important for the ecosystems of polar and sub-polar regions. In the northern hemisphere, the distributions of each species coincide with latitudinal gradients. In the northern North Atlantic Ocean there is C. finmarchicus with its distribution centres in the Labrador and Norwegian Seas. Further north there are C. glacialis, an Arctic shelf water species, and C. hyperboreus, a deep oceanic one. Most zooplankton biomass in the Arctic shelf seas is formed by the grazer C. glacialis, which, in this lipid-based food web, is a key link between low-energy microalgae and higher trophic levels, including zooplankton, fish and planktivorous seabirds. The species are usually sympatric, especially at the limits of their distribution ranges. However, with the recently observed climatic changes, copepods are shifting their spatial distribution northwards, mostly due to seawater temperature rise.

According to Bergmann clines, body sizes of ectotherm species within a genus, as well as of populations within a species, increase along temperature gradients towards colder areas. Such a relationship has also been reported for copepods of the genus Calanus, that generally live longer and have larger body sizes in higher latitudes of the Northern hemisphere. The functional links between body and genome size in copepods, and their physiological,
morphological and ecological consequences have been studied by different authors 13–15 leading to conclusions that genome size is a key trait of organisms, although its adaptiveness is still unclear. 

*Calanus* copepods are influenced by ongoing climate changes, especially the increasing temperatures in the Arctic and sub-Arctic waters and sea-ice loss. Increasing numbers of *C. finmarchicus* are transported into the Arctic with the West Spitsbergen Current, which is the continuation of the North Atlantic Current, carrying boreal species northwards 41. At the same time, due to the sea temperature rise, changes in population structure are observed, possibly leading to earlier spawning and accelerated development, which in turn could lead to food web reorganization in the sub-Arctic regions 46. Its sibling species, *C. glacialis*, which needs energy from the ice algal bloom to fuel its reproduction, is also affected by increasing temperatures 47 and the loss of sea ice. This may lead to a mismatch between the primary production peaks and the reproductive cycle of this key species, with negative consequences for the higher trophic levels 4. Due to the important role of both *Calanus* species in the lipid-driven Arctic and sub-Arctic food webs, major functional changes in these marine ecosystems are anticipated under climate warming.

In spite of the importance of the calanoid copepods for the ecosystems and as indicators of ongoing climatic changes, there is still a lack of consensus about the relationships among their lineages, in which high latitude species are especially under-represented. Mitochondrial DNA is useful for evolutionary studies and for reconstructing the phylogenetic relationships of copepods, which is challenging due to the morphological complexity of their taxonomic features 48,49. Few copepods have their mitochondrial genomes published so far 20, of which only one is an Arctic calanoid, *Calanus hyperboreus* 51. In this genus, within the superorder Gymnoplea 22, only one species of the northwest Pacific, *Calanus sinicus*, has its mitochondrial genome also known 55. Both *Calanus* genomes exhibit some unusual features, including the longest control region reported for a crustacean, a large tRNA gene cluster, and reversed GC (guanine-cytosine content) skews in most of the protein-coding genes in *C. hyperboreus* 51, and the concurrence of multiple non-coding regions and a reshuffled gene arrangement in *C. sinicus* 25. The known mitogenomes of Gymnoplea, having the length of 17,910 bp in the first one and 20,460 bp in the latter species, are also considerably longer than other known copepod mitochondrial genomes, all in the superorder Podoplea. Due to these unusual characters, the conservation of mitochondrial gene order in copepods has been questioned by some authors 23,24, while others claim that known mitochondrial genomes retained pan-crustacean features and conserved calanoid-specific patterns 21 or adapted to harsh environments by mitogenome rearrangements 52,42,53.

We aimed to sequence the mitochondrial genomes of the Arctic *Calanus glacialis* and its sibling species of the North Atlantic, *Calanus finmarchicus*, to study morphological and evolutionary relationships between high latitude copepods, and to obtain a database which will facilitate the development of genetic markers suitable in pursuing evolutionary processes in *Calanus* copepods, which in the future could be linked with climate change in the Arctic and sub-Arctic regions.

**Results**

**Structure and composition of *C. glacialis* and *C. finmarchicus* mitogenomes.** For *C. glacialis*, both NGS assemblers produced similar assembly. Despite a good coverage (approx. 100 fold), they were unable to retrieve the complete mitochondrial genome in one contig from the genomic data set. Instead, two contigs of 9,686 bp and 17,656 bp were generated. Based on the mapping of PE reads and the similarity of repeats at respective contig ends, the two contigs were provisionally joined to form a circular entity. The resulting genetic map of *C. glacialis* mitogenome is similar, only more tandem arrays of TATA repeats are present than in *C. hyperboreus*. There are limited to contig 2 and indicated one additional gene, otherwise this mitogenome may have the same gene order as the MRCA of *C. finmarchicus*.

Non-coding regions. There are at least three NCRs longer than 300 bp within the 17 kbp contig and only 2.5 kb of the 10 kbp contig is coding in *C. glacialis*. In addition, the ends of both contigs contain NCRs characterized by complex repeats, hence their length is uncertain. Interestingly the NCRs are not particularly A + T-rich (Fig. 1). They also do not contain any orphan open reading frames of a significant length. The situation in *C. finmarchicus* mitogenome is similar, only more tandem arrays of TATA repeats are present than in *C. glacialis*. There is no sequence similarity between NCRs of the two species, despite their similar composition.
Figure 1. Mitochondrial genomes organization: (a) of the Arctic copepod Calanus glacialis and (b) of the North Atlantic Calanus finmarchicus. The span of the assembled contigs is indicated by the outer arches. All protein coding genes (CDS) are labelled by the names of the encoded proteins, two rrn genes are labelled 16 S and 12 S for large and small subunit, respectively. The trn genes are labelled by the one letter code of the respective amino acid, distinguishing the two possible anticodons by a number, wherever applicable. The supranumerary trnL<sup>UAA</sup> gene in (a) is distinguished by an asterisk. Direction of transcription is indicated by the position and direction of the arrows, with clockwise transcribed genes on the outside and anticlockwise transcribed genes on the inside of the circle representing the genome. Positions of polyadenylation sites (pA) and microsatellite-like repeats (tata) are also indicated. The two inner circles represent local compositional bias, calculated in a 300 bp long window. The light grey parts of the first circle represent positive AT skew bias while the dark grey parts of this circle represent negative AT skews. The inner, color circle represents local GC (guanine-cytosine) content, relative to the reference 50%.
Ribosomal RNA and tRNA genes. Surprisingly, the most A+T rich parts of the mitogenomes are the two rrn genes (Fig. 1). They are relatively short (1125 and 656 bp in both mitogenomes), but do align well with the respective genes of C. sinicus. The overall nucleotide p-distance between C. glacialis and C. sinicus 16S genes is 20%, for 12S gene the distance is 14%. For the pair of mitogenomes presented here these distances are 14% and 13%, respectively.

The 22 expected trn genes are present in the C. glacialis mitogenome. They form a set typical for the invertebrate mitochondrial genetic code. All four-fold degenerated codons are served by a trn with T at the first anti-codon (wobble) position within trn genes. The two-fold degenerated families have either T or G at this position, with the single exception of the trnM with C at the wobble position. Surprisingly, two copies of trnL2, both with TAA anticodon were found. They both should form typical, stable cloverleaf structures according to arwen, with comparable bitscores (30 and 31). One of the genes was located between nd2 and nd4, in the first contig, while

Figure 2. Dotplot comparison of the two Calanus mitogenomes: C. glacialis and C. sinicus. Forward and reverse dotplots were overlaid to show the cases of reversed fragments. Schematic maps of both mitogenomes are on the margins, with protein coding genes (CDS) in white and rrn genes in grey. It is evident that non-coding parts of the genomes lack homology comparable with the homology of coding parts. The genes are also remarkably similar in length.

Figure 3. Parsimonious evolutionary scenario leading to the four extant gene orders. CREx analysis implies that C. sinicus and C. glacialis gene orders can be derived from each other by as little as three reversed translocations. The gene order represented by C. hyperboreus is substantially more distant, separated by more than seven events. The ancestral gene orders proposed here fit these data well, with two events separating C. sinicus and C. glacialis mitogenomes from their common ancestor and the further four events separating common ancestor of all three species. Simple translocations are represented by black arrows, translocations with reversals by grey arrows. Black asterisks show the positions of major non-coding regions. Note that the non-sequenced part of C. sinicus and C. finmarchicus genomes (indicated by the question marks) will most likely contain the missing trn genes (C, D and R), which also fits this scenario well.
the alternative version was present between cox2 and nd3 in the second contig (L2* in Fig. 1). Since there is no way to dismiss one of them as a non-functional pseudogene, it must be tentatively assumed that there are 23 trn genes in the C. glacialis mitogenome. Twelve of the trn genes are found in one big cluster between cox2 and nd5.

The sequenced part of C. finmarchicus mitogenome does not contain all expected trn genes, the trnC, trnR, and trnD are missing. Notably, according to the analysis of gene order evolution, they are expected to be within the missing parts of the mitogenome (Fig. 3). One supranumerary trn gene was found in C. finmarchicus: the second copy of trnS1 gene, located in exactly the same space where the supranumerary trnL2 was found in C. glacialis (Fig. 1).

Protein-coding genes. All protein coding genes (CDS) typically expected in animal mtDNA are present in the C. glacialis and C. finmarchicus mitogenomes. Given the unusually loose overall structure of these genomes, it is surprising to note that the CDS lengths are largely unaffected, with the overall number of encoded amino acids at 3698-3701, within two amino acid difference of C. sinicus (Table 1). There are also seven new cases of incomplete stop codons, in each of the cases the completion of the stop codon by polyadenylation of the transcript was confirmed by transcript mapping. These incomplete stop codons end cytβ, nd1, and nd5 genes in both species and atp6 in C. finmarchicus.

Genetic distances between the congeneric species were calculated based on concatenated CDS alignments (Fig. 4). The closest relative of C. glacialis is C. finmarchicus, followed by C. sinicus, however the overall genetic

| CDS name | C. glacialis bp | C. finmarchicus bp | C. hyperboreus bp | C. sinicus bp | RSD % |
|----------|----------------|--------------------|-------------------|---------------|-------|
| atp6     | 714            | 712                | 708               | 711           | 0.41  |
| atp8     | 162            | 162                | 159               | 162           | 0.95  |
| cox1     | 1557           | 1557               | 1546              | 1548          | 0.34  |
| cox2     | 705            | 705                | 704               | 705           | 0.00  |
| cox3     | 792            | 792                | 792               | 792           | 0.00  |
| cytβ     | 1133           | 1132               | 1137              | 1137          | 0.15  |
| nd1      | 919            | 919                | 915               | 917           | 0.31  |
| nd2      | 969            | 969                | 969               | 969           | 0.00  |
| nd3      | 354            | 354                | 354               | 354           | 0.00  |
| nd4      | 1302           | 1302               | 1302              | 1302          | 0.00  |
| nd4L     | 336            | 336                | 336               | 336           | 0.00  |
| nd5      | 1711           | 1712               | 1710              | 1723          | 0.39  |
| nd6      | 474            | 483                | 474               | 480           | 0.95  |
| All CDS  | 3698           | 3701               | 3690              | 3700          | 0.14  |

Table 1. Lengths of all mitochondrial protein coding genes (CDS) of four Calanus species (in base pairs, bp) and the lengths of the corresponding amino acid chains (aa), with their relative standard deviation (RSD). The lengths of amino acid chains are obtained by dividing the length of nucleotide sequence by three and subtracting one (for the stop codon), except for the eleven cases of genes with incomplete stop codons (in boldface).

Figure 4. Estimates of intragenic polymorphism in Calanus and Tigriopus. Four Calanus (C. finmarchicus, C. glacialis, C. sinicus, and C. hyperboreus labelled as Cf, Cg, Cs and Ch, respectively) and two Tigriopus (T. japonicus and T. californicus, Tj and Tc) species were compared. The dataset comprised concatenated amino acid alignments of all mitogenome CDS. All positions with gaps and missing data were eliminated. There were a total of 3306 positions in the final dataset. The number of amino acid differences per site (p-distances) between sequences, with standard error estimates are shown. The average, uncorrected amino acid p-distance between Calanus and Tigriopus genera is slightly higher than 0.5.
distance between them is still substantial. Uncorrected amino acid p-distances are larger than 12% in all cases. The distance between the two \textit{Tigriopus} species, shown for comparison, is nevertheless even higher.

It is important to consider the local compositional biases in the highly rearranged and variable structure of the \textit{Calanus} mitogenomes. Local AT skew (Fig. 1) shows excess of T over A in all CDS transcripts. However, subtle mutational pressures can only be traced by calculating skews at four-fold degenerated (selectively neutral) sites, as these should not be affected by compositional biases of encoded proteins. Particularly the GC skew at neutral sites is considered indicative of the orientation and proximity of regulatory elements. In the studied \textit{Calanus} mitogenomes, the GC skew at neutral sites is highly irregular: the values are small and change both within each genome and between the congeneric genomes for the same genes (Fig. 5).

\textbf{Nuclear background.} Nuclear genomes of the two \textit{Calanus} species are expected to be complex and sequencing them will require much more effort than a single NGS run. Nevertheless, the sequencing gave us the opportunity to check some basic properties of \textit{C. glacialis} and \textit{C. finmarchicus} nuclear genomes.

In principle, it should be possible to estimate the size of the haploid genome from the distribution of unique k-mers\textsuperscript{26}, providing the sequencing coverage is adequate. With $4.5 \times 10^{10}$ bp obtained in the present study, the k-mer distribution analysis should yield a meaningful result if the sampled genome was similar to the human genome in size. The distributions of k-mer frequencies in our data did not have non-trivial maxima (data not shown), indicating large genomes, rich in repetitive sequences. An attempt has been made to assemble the most repetitive parts of the nuclear genomes. The two species are remarkably similar, but \textit{C. glacialis} has more repetitive elements. Both the number of identified elements was higher: 4582 vs. 4204 for \textit{C. finmarchicus} (at the same arbitrary assembly threshold of 45 $\times$ ), and they represented bigger fraction of the genome (30\% vs 25\%). The expansion of LINE elements in the genome of \textit{C. glacialis} is probably responsible for this difference (Fig. 6).

The most abundant repetitive element in both genomes is the rDNA gene cluster. The cluster apparently forms one or more arrays of tandem repeats, each containing the four rRNA genes (Fig. 7). The genetic distance between rRNA genes of the two studied \textit{Calanus} species is remarkably low: there are no differences in 5.8S, one substitution in 18S and 5S and 34 substitutions in the longest 28S gene (still, more than 99\% sequence similarity). The average coverage of the consensus sequence was very high: 8600 $\times$ for \textit{C. glacialis} and 5300 $\times$ for \textit{C. finmarchicus}, again indicating more redundant structure of the former genome.

\textbf{Discussion} The \textit{C. glacialis} and \textit{C. finmarchicus} mitogenomes, both longer than 27 kbp are the longest reported mitogenomes, not only within the superorder Gymnoidea: 17,910 bp in \textit{C. hyperboreus}\textsuperscript{21} and $>20,460$ bp in \textit{C. sinicus}\textsuperscript{21}, but also

\textbf{Figure 5.} Compositional bias at four-fold degenerate sites in CDS of the four \textit{Calanus} species and \textit{T. californicus}. GC (guanine-cytosine) skew was calculated separately for each gene, based only on the most neutral sites, that is four-fold degenerated positions. This index is particularly sensitive to mutational pressure associated with the relative position of the gene and the origin of replication. Native directions and relative positions of the genes were used, the charts start at the position of a major NCR in each species.

\textbf{Figure 6.} The classification of repetitive elements from the nuclear genome of \textit{C. glacialis} (blue) and \textit{C. finmarchicus} (red). The elements represented by k-mer coverage greater than 45 were assembled by \textit{tedna}, resulting in 4582 elements from \textit{C. glacialis} and 4204 elements from \textit{C. finmarchicus}. These elements were then roughly classified by \textit{Teclass}.
among Copepoda in general, including the species from the superorder Podoplea, which are relatively short and range from 13,440 bp in *Caligus clemensi* to 15,981 bp in *Paracyclopina nana*. Such extraordinary lengths of the *C. glacialis* and *C. finmarchicus* mitogenomes are caused by an unusual length of NCRs, a feature that was also observed in *C. sinicus*. The 20,674 bp long sequence of *C. glacialis* mitochondrial genome, announced recently, is almost identical to contig 2 presented by us (>99% sequence similarity), extended at 3' end with low complexity repeatable region of substantial length, therefore contrary to the description given by the authors, we believe that in fact this contig does not contain the complete set of mitochondrial genes (compare Fig. 1a).

The *C. glacialis* and *C. finmarchicus* mitogenomes are structurally similar to the mitochondrial genome of the North Pacific *Calanus sinicus* and differ from the much more typical mitogenome of the deep-water, Arctic *Calanus hyperboreus*. However, given the relatively large differences in sequence and structure of *C. glacialis* and *C. finmarchicus* mitogenomes, it must be concluded that these two species are genetically distinct and have evolved independently for quite a long time. Thus, it seems unlikely that *C. glacialis* and *C. finmarchicus* can hybridize in the areas of sympatry.

The structural similarities in copepod mitogenomes parallel the genetic distances in this case (Fig. 3 and Fig. 4), reflecting the general evolutionary trend for the expansion of NCRs within mitochondrial genomes of *Calanus* species. An interesting consequence of this expansion is the unexpected property of their mitochondrial DNA: the AT-content of the sequenced part of the mitogenome appears to be similar to the average of the reported nuclear loci, and lies within the range recorded for the sequenced nuclear DNA (Fig. 8), confirming that the separation of mitochondrial and nuclear DNA by density gradients is impossible.

The structure of mitogenome in arthropods is considered relatively stable. The gene order close to the one found in the mitogenome of *Homarus americanus* is usually seen as a typical, ground pattern. Copepods are exceptional arthropods because, with only a few complete mitogenomes published, they already present many deviations from the reference gene order. However, the mitogenomes presented here are exceptional even among copepods. There is only one other known example of a congeneric mitogenomes in Copepoda: the two *Tigriopus* species from the order Hapacticoidea and these mitogenomes are much more stable. Even though they substantially differ at the sequence level (Fig. 4), their structure is similar. In fact, the only difference between these two genomes is the position of *trnW* gene, but a closer look at the actual alignment of the mitogenomes suggests that even this small difference is only caused by inconsistent annotation and not a real, structural difference. So, in the genus *Tigriopus*, in an evolutionary time long enough to accumulate substitutions in more than...
25% of mitochondrial amino acid positions, there was no change in gene order. Contrary to that, the rate of rearrangements in the genus *Calanus* must be much greater, since despite the lower genetic distance (less than 20% as p-distance, Fig. 4), there must have been at least four genomic rearrangements (Fig. 3), two involving large CDS and not just single trn genes as is usually the case.

The expectation that the rearrangement and substitution rates are generally linked has been dismissed, so in theory, this effect could be the result of slower-than-average amino acid substitution rate in *Calanus*. However, mitogenome rearrangements have also a noticeable effect on the local compositional biases, particularly if they involve NCRs. Negative GC skew at neutral sites has been used as an argument for reversal of the origin of replication in several species, including *Tigriopus*. In the mitogenome of *C. hyperboreus* the GC skew at neutral sites is also negative for all but the two shortest genes (Fig. 5). Similar conclusions have been reached when the data for all sites, not just the neutral ones, were used. Still, the absolute skew values were smaller in *C. hyperboreus* than in *T. californicus*, suggesting less stable mutational pressure in *Calanus*. This is further confirmed by GC skew at neutral sites in the three *Calanus* species with multiple NCRs: *C. finmarchicus* and *C. glacialis*. The absolute GC skews are even smaller in these species, and the sign of the skew frequently differs between genes, even the neighbouring ones (Fig. 5). This pattern suggests independent replication from different NCRs and/or frequent change of the mutational pressure associated with mitochondrial DNA replication. Therefore, the increased rearrangement rate and not the slower substitution rate is responsible for the observed pattern of polymorphisms in *Calanus*.

Our inability to close all the gaps by PCR is not totally unexpected, as similar difficulties were encountered by other teams sequencing *Calanus* mitogenomes. Their results can be viewed as a limitation of PCR technology or as a proof of mitogenome fragmentation in *C. finmarchicus*. Similar considerations apply to the presented mitogenomes of *C. glacialis* and *C. finmarchicus*. The assembled contigs may actually represent separate mitochondrial "chromosomes", existing without a constant physical linkage, even if they can be mapped on a circular mitogenome. The lack of consistent compositional biases could be used as an argument supporting such hypothesis. Fragmented, linear mitochondrial genomes of medusozoa cnidarians often evolve gene duplications and are terminated by inverted repeats. Nearly all the contigs assembled from *Calanus* mitogenomes are terminated with palindromic repeats. Moreover, in the case of *C. glacialis* mitogenome we report the existence of two trnL2 genes, one per each contig. So there are some arguments supporting atypical structure of the reported mitogenomes, but they are not conclusive. Whatever the cause, these difficulties will not be solved without a breakthrough in sequencing technology allowing the reads to continue past difficult to amplify regions.

It has been argued that there is a relationship between body and genome size of keystone marine copepods. Larger genomes were observed in cold waters species, a feature linked to slower growth and metabolic rates at lower temperatures. The DNA content of a single nucleus is very high in the two studied species: 13 pg in *C. finmarchicus* and 24.2 pg in *C. glacialis*. According to this translates to 1.27 and 2.37 × 10^15 bp per genome, respectively. Clearly, our sequencing effort (at 4.5 × 10^15 bp) was not enough to achieve a more precise estimate of the genome size. However, if either of the two genomes was in fact highly polyploid, the achieved coverage could have been adequate to estimate the size of the haploid equivalent. This was not the case, hence no direct support for polyploidization. According to the two species differ in the concentration of rDNA in their genomes – the species with larger genome has also the higher concentration of rDNA, by approximately 17%. This suggests that the existing relationship between body and genome sizes in these species cannot be explained by simple polyploidization. Our data suggest even a greater disproportion in the concentration of rDNA between the two genomes (60% more rDNA in *C. glacialis*) and the expansion of certain classes of repeatable elements in this species.

In conclusion, large differences in the length and structure of *C. glacialis* and *C. finmarchicus* mitochondrial genomes indicate that the species are genetically distinct, so it is unlikely that they can hybridize. The molecular and evolutionary processes leading to the rapid structural evolution of both nuclear and mitochondrial genomes of these zooplankton species, key for the polar and sub-polar marine ecosystems, remain to be explained. We believe that the molecular resources presented here: the mitogenomic and rDNA sequences of the Arctic *C. glacialis* and its North Atlantic sibling *C. finmarchicus*, as well as the database of repetitive elements should facilitate the development of genetic markers suitable in pursuing this subject.

**Methods**

**Sampling, DNA extraction and sequencing.** *Calanus glacialis* was collected from the Hornsund fjord (station 77°00.54’N 16°28.3’E, western Svalbard Archipelago) in August 2012 (250 individuals) and 2013 (150 individuals) and *C. finmarchicus* from the Norwegian Sea (73°30.06’N 13°04.1’E) in June 2013 (700 individuals), during the routine scientific cruises onboard r/v Oceania. To facilitate species identification and handling, mainly the fifth copepodites were hand-picked and kept in 95% ethanol until use.

The strategy to sequence mitogenomes of the two species assumed that it will be possible to start with the sequences of mitochondrial transcripts, then design PCR primers and close the gaps by direct sequencing of PCR products. To identify mitotranscripts we used RNA-Seq data of – SRA project accession number SRP053198 for *C. glacialis* and the data of SRA project accession number SRP058899 for *C. finmarchicus*. Transcripts assembly and identification followed the established pipeline and led to the identification of all CDS and rRNA mitotranscripts. Several PCR primers were designed and most gaps between the transcriptomic contigs were closed, as planned. In this round, we worked with 250 individuals of *C. glacialis* and 450 of *C. finmarchicus* from which we extracted DNA using a modified CTAB protocol. However, technical difficulties associated with working with limited amount of starting material and frustrating failures in several seemingly well-designed PCRs precluded the realization of the original strategy. Moreover, since the overall genetic polymorphism in mitochondrial DNA of *C. glacialis* turned out to be very low, we decided to continue the work with a heterogeneous sample
of high quality DNA. Consequently, an attempt has been made to isolate pure mitochondrial DNA by density gradient centrifugation.

The procedure described by Lang and Burger\(^2\) was applied, isopyknotic caesium chloride gradient centrifugation was run for 16 h in a ultracentrifuge Beckman Coulter Optima L-90K with a vertical rotor. However, under long-wavelength UV we observed only a single DNA band, instead of expected separate mitochondrial and genomic DNA bands. Therefore, the procedure was repeated, starting with additional 150 individuals of C. glacialis and 250 of C. finnarchicus, with extended centrifugation time to 24 h. Again, only a single DNA band was obtained.

This highly purified DNA was used in NGS, in hope that the concentration of mitochondrial DNA would be high enough to allow a mitogenome assembly. To maximize our chances, one of the highest-throughput technologies available was used, the Illumina HiSeq X-Ten. Because of the high concentration and purity of DNA, the sequencing library was prepared using PCR-Free kit (MacroGen). The library has average insert size of 440 bp. The resulting set of approx. $3 \times 10^5$, 150 bp long, paired end (PE) sequence reads, was used in all subsequent analyses.

**Gene identification and genome analysis.** The preliminary identification of mitochondrial transcripts in RNA-Seq datasets\(^3\) was performed using estwise algorithm implemented in wise2 software\(^3\), using hmm profiles of all mitochondrial protein coding sequences (CDS). Identified transcripts were then used as baits in MITObim\(^4\) and NOVOPlasty\(^5\) assemblers to retrieve the mitogenomes from the sets of NGS sequencing reads.

Gene annotations have been done semi-automatically. De novo prediction of all protein-coding genes was attempted using a set of algorithms implemented in CRITICA\(^6\), GLIMMER3\(^7\) and wise2. For prediction of RNA genes, ARWEN was used\(^8\). The available reasonable complete mitogenomes of the copepods from the same genus, C. hyperboreus\(^9\) accession number: JX678968, C. sinicus\(^1\) accession number: GU355641, were used to verify and fine-tune the annotations. The transcripts were mapped back onto the mitogenome contigs in CLC Genomics Workbench (version 9.5.4, QiAGENE) to verify consistency of gene annotations. Amino acid sequences for protein-coding genes were obtained using the invertebrate mitochondrial genetic code (NCBI translation table 5).

The obtained raw sequence reads are deposited in SRA database under the following accession numbers: SRR6063609 (C. glacialis) and SRR6065686 (C. finnarchicus), the assemblies are accessible through BioProject accessions: PRJNA411905 (C. glacialis) and PRJNA411907 (C. finnarchicus), the individual rDNA contigs are deposited in GenBank under accession numbers MF993123 and MF993124, mitochondrial contigs under MG001883 - MG001890.

There are several complete copiot Mitogenome deposited in GenBank available for comparative analysis. However, they have a different organization and are also substantially distant from the mitogenome reported here, making meaningful comparison problematic. Therefore, the two congeneric sequences earlier were selected for a comparative analysis. The mitogenomic sequences of two harpacticoid copiot species from the genus Tigriopus were also used in some comparative analyses: Tigriopus californicus DQ917374\(^9\) and Tigriopus sp. AY959338\(^2\). All comparative analyses were performed following nucleotide alignment in MEGA\(^7\), all nucleotide alignments were adjusted to correspond to amino acid alignments wherever possible, sites with alignment gaps were eliminated and bootstrap procedure with 1000 replicates was used to estimate standard errors.

For sequence manipulation such as extraction of genes and dotplot comparisons, EMBOS suite was used\(^1\). Analysis of gene order data was performed using some heuristics implemented in CREx web server\(^2\), Analysis of k-mer frequency distributions were done in Jellyfish\(^3\) and the assembly of repeatable DNA was done in Tedna\(^5\), the provisional classification of the assembled elements was done by TEclass\(^4\). All mapping and manipulation of genomic reads were done in CLC Genomics Workbench.

**References**

1. Fleminger, A. & Hulsemann, K. Geographical range and taxonomic divergence in North Atlantic Calanus (C. helgolandicus, C. finnarchicus and C. glacialis). Mar. Biol. 200, 233–248, https://doi.org/10.1007/BF00390879 (1977).
2. Conover, R. J. Comparative life histories in the genera Calanus and Neocalanus in high latitudes of the northern hemisphere. Hydrobiol. 167/168, 127–142, https://doi.org/10.1007/BF00026299 (1988).
3. Tande, K. S. Calanus in North Norwegian fjords and in the Barents Sea. Polar Res. 10, 389–408, https://doi.org/10.3402/polar.v10i02.6754 (1991).
4. Søreide, J. E., Leu, E., Berge, J., Graeve, M. & Falk-Petersen, S. Timing of blooms, algal food quality and Calanus glacialis reproduction and growth in a changing Arctic. Glob. Change Biol. 16, 3154–3163, https://doi.org/10.1111/j.1365-2486.2010.02175.x (2010).
5. Weidmann, A. et al. Ice-related seasonality in zooplankton community composition in a high Arctic fjord. J. Plankton Res. 35, 831–842, https://doi.org/10.1016/j.plankt.2013.01.001 (2013).
6. Węsławski, J. M., Kosztelnik, J., Kwasniewski, S., Stepieniewicz, L. & Malinga, M. Summer food resources of the little auk, Alle alle (L.) in the European Arctic seas. Polish Polar Res. 20, 387–403 (1999).
7. Karmowski, N. J., Węsławski, J. M., Kwasniewski, S., Walkusz, W. & Bęczynska-Moeller, A. Foraging behavior of little auks in heterogeneous environment. Mar. Ecol. Prog. Ser. 253, 389–393, https://doi.org/10.3354/meps253389 (2003).
8. Falk-Petersen, S., Timoferj, S., Pavlov, V. & Sargent, J. R. Climate variability and possible effects on Arctic food chains: the role of Calanus in Arctic-Alpine ecosystems and people in a changing environment (ed. Örback, J. B. et al.). 147–166 (Springer, 2007).
9. Reygondeau, G. & Beaugrand, G. Ecosystem-driven shifts in distribution of Calanus finnarchicus. Glob. Change Biol. 17, 756–766, https://doi.org/10.1111/j.1365-2486.2010.02310.x (2011).
10. Chust, G. et al. Are Calanus spp. shifting poleward in the North Atlantic? A habitat modelling approach. ICES J. Mar. Sci. 71, 241–253, https://doi.org/10.1093/icesjms/fst147 (2014).
11. Weidmann, A. et al. Shift towards the dominance of boreal species in the Arctic: inter-annual and spatial zooplankton variability in the West Spitsbergen Current. Mar. Ecol. Prog. Ser. 501, 41–52, https://doi.org/10.3354/meps10694 (2014).
12. Atkinson, D. Temperature and organism size, a biological law for ectotherms? Adv. Ecol. Res. 25, 1–58, https://doi.org/10.1016/0001-89 | 7: 13702 | DOI:10.1038/s41598-017-13807-0 | 9
51. Weydmann, A. et al. Postglacial expansion of the Arctic keystone copepod Calanus glacialis. Mar. Biodiv. doi:10.1007/s12526-017-0774-4 (2017).
52. Lang, B. F. & Burger, G. Purification of mitochondrial and plastid DNA. Nat. Protoc. 2, 652–660, https://doi.org/10.1038/nprot.2007.58 (2007).
53. Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. Genome Res. 14, 988–995, https://doi.org/10.1101/gr.186504 (2004).
54. Hahn, C., Bachmann, L. & Chevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 41, e129–e139, https://doi.org/10.1093/nar/gkt371 (2013).
55. Dierckxsens, N., Mardulyn, P. & Smits, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 45, e18–e18, https://doi.org/10.1093/nar/gkw955 (2017).
56. Badger, J. H. & Olsen, G. J. CRITICA: coding region identification tool invoking comparative analysis. Mol. Biol. Evol. 16, 512–524, https://doi.org/10.1093/oxfordjournals.molbev.a026133 (1999).
57. Delcher, A. L., Harmon, D., Kasif, S., White, O. & Salzberg, S. L. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27, 4636–4641 (1999).
58. Laslett, D. & Canback, B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24, 172–173, https://doi.org/10.1093/bioinformatics/btm573 (2008).
59. Burton, R. S., Byrne, R. J. & Rawson, P. D. Three divergent mitochondrial genomes from California populations of the copepod Tigriopus californicus. Gene 403, 53–59, https://doi.org/10.1016/j.gene.2007.07.026 (2007).
60. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874, https://doi.org/10.1093/molbev/msw054 (2016).
61. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 16, 276–277, https://doi.org/10.1016/S0168-9525(00)02024-2 (2000).
62. Bernt, M. et al. CREx: inferring genomic rearrangements based on common intervals. Bioinformatics 23, 2957–2958, https://doi.org/10.1093/bioinformatics/btm468 (2007).
63. Zytnicki, M., Akhunov, E. & Quesneville, H. Tedna: a transposable element de novo assembler. Bioinformatics 30, 2656–2658, https://doi.org/10.1093/bioinformatics/btu365 (2014).
64. Abrusan, G., Grundmann, N., DeMester, L. & Makalowski, W. TEllass—a tool for automated classification of unknown eukaryotic transposable elements. Bioinformatics 25, 1329–1330, https://doi.org/10.1093/bioinformatics/btp084 (2009).

Acknowledgements
This research project was supported by the Polish Ministry of Science and Higher Education [Juventus Plus grant no. IP2014 050573]; and FCT-CCMAR Portugal [Multi/04326/2013]. We thank prof. Igor Konieczny and dr Katarzyna Węgrzyn from the Laboratory of Molecular Biology at the Intercollegiate Faculty of Biotechnology UG&MUG for the possibility to use the ultracentrifuge. All computationally intense tasks were supported by PL-GRID infrastructure.

Author Contributions
A.W., A.B., E.A.S. and G.A.P. conceived the experiment. A.W. collected samples and identified species. A.W., A.P., M.L. and K.S.W. conducted the lab work. A.B. and A.W. analyzed the results and wrote the manuscript. All authors reviewed the manuscript.

Additional Information
Competing Interests: The authors declare that they have no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017