Molecular Ecology Resources
September 2014, Volume 14 (5), Pages 1072-1079
http://dx.doi.org/10.1111/1755-0998.12241

Genome- and transcriptome-assisted development of nuclear insertion/deletion markers for *Calanus* species (Copepoda: Calanoida) identification

I. Smolina[1,*], S. Kollias[1], M. Poortvliet[1], T. G. Nielsen[2,3], P. Lindeque[4], C. Castellani[5], E. F. Møller[6], L. Blanco-Bercial[7] and G. Hoarau[1]

[1] Faculty of Biosciences and Aquaculture, University of Nordland, Bodø, Norway
[2] National Institute of Aquatic Resources, Section of Oceanography and Climate, Technical University of Denmark, Charlottenlund, Denmark
[3] Greenland Climate Research Centre, Greenland Institute of Natural Resources, Nuuk, Greenland
[4] Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth, UK
[5] Sir Alister Hardy Foundation for Ocean Sciences, Plymouth, UK
[6] Department of Bioscience, Aarhus University, Roskilde, Denmark
[7] Department of Marine Sciences, University of Connecticut, Groton, CT, USA

*: Corresponding author : Irina Smolina, fax: +47 75 51 74 57 ; email address : ism@uin.no

Abstract:

Copepods of the genus *Calanus* are key zooplankton species in temperate to arctic marine ecosystems. Despite their ecological importance, species identification remains challenging. Furthermore, the recent report of hybrids among *Calanus* species highlights the need for diagnostic nuclear markers to efficiently identify parental species and hybrids. Using next-generation sequencing analysis of both the genome and transcriptome from two sibling species, *Calanus finmarchicus* and *Calanus glacialis*, we developed a panel of 12 nuclear insertion/deletion markers. All the markers showed species-specific amplicon length. Furthermore, most of the markers were successfully amplified in other *Calanus* species, allowing the molecular identification of *Calanus helgolandicus*, *Calanus hyperboreus* and *Calanus marshallae*.

Keywords: InDels ; Ion Torrent ; marker design ; next-generation sequencing ; species delimitation ; zooplankton

Introduction

Copepods of the genus *Calanus* play an important role both as consumers of primary production and as prey for many ecologically and commercially important species (Lønne & Gulliksen 1989). *Calanus* comprise up to 70 % of the zooplankton biomass (Head *et al.* 2003), with three species dominating the subarctic North Atlantic and Arctic Oceans: *Calanus finmarchicus* (Gunnerus, 1770), *C. glacialis* Jaschnov, 1955 and *C. hyperboreus* Kröyer, 1838 (Søreide *et al.* 2008; Falk-Petersen *et al.* 2009). These species are generally closely associated with the water masses where they originate from (Hirche & Kosobokova 2007), and therefore can be regarded as biological indicators of the North Atlantic sub-Arctic (*Calanus finmarchicus*) and Arctic (*C. glacialis* and *C. hyperboreus*) biomes (Kwasniewski *et al.* 2003; Daase *et al.* 2007). As a response to ocean warming, a northward shift of the subarctic species *C. finmarchicus* has been detected with a parallel decline throughout most of the North Atlantic (Beaugrand *et al.* 2002; Chust et al 2013; Maar *et al.* 2013). A further northward shift of *C. finmarchicus* is predicted into Arctic waters that are dominated by *C. glacialis* and *C. hyperboreus* (Beaugrand *et al* 2013; Wassmann *et al.* 2011), with a subsequent change in the food-web dynamics and secondary production (Falk-Petersen *et al.* 2007). Thus, *Calanus* species are central to many ecological, environmental and climatological studies (Gabrielsen *et al.* 2012) and correct identification of *Calanus* species is essential.

Morphological identification of *Calanus* species is hampered by their subtle morphological differences, especially between early developmental stages (Hirche *et al.* 1994; Lindeque *et al.* 1999; Lindeque *et al.* 2006), resulting in persistent misidentification (Lindeque *et al.* 2006; Parent *et al.* 2011; Gabrielsen *et al.* 2012). Molecular identification of *Calanus* species is mainly based on mitochondrial markers, 16S rDNA (Bucklin *et al.* 1995; Lindeque *et al.* 1999) or cytochrome oxidase I (Hill *et al.* 2001). However, the recent report

of hybrids among *Calanus* species (Parent *et al.* 2012) highlights the need for diagnostic nuclear markers in order to efficiently identify hybrids from parental species. Ten nuclear microsatellite loci have been developed for *C. finmarchicus* for in-depth population genetic studies (Provan et al. 2007), and have revealed hybridization between *C. finmarchicus* and *C. glacialis* in the Canadian Arctic (Parent *et al.* 2012). However, high mutation rate, lack of a mutation model, allelic dropout and difficulties to score alleles (for review see Pompanon *et al.* 2005; Selkoe & Toonen 2006), as well as homoplasy when comparing two species (Chambers & MacAvoy 2000; Curtu *et al.* 2004) can limit their use for species identification. Conversely, nuclear insertion/deletion (InDel) polymorphisms are co-dominant, have a low mutation rate, arise mainly from a single mutation event and provide a reasonably conserved phylogenetic signal (Liu & Cordes 2004; Nagy *et al.* 2012). The usability of InDel markers is further enhanced by ease of genotyping, repeatability, and possible use with degraded DNA (e. g. formalin or museum samples) due to generally small amplicon size (Väli *et al.* 2008; Pepinski *et al.* 2013).

Next generation sequencing (NGS) has revolutionized speed and availability of data acquisition in biology. For non-model species, NGS has dramatically facilitated the development of molecular markers such as microsatellites (Nie *et al.* 2012; Hunter & Hart 2013), InDels (Choi *et al.* 2013) and single nucleotide polymorphisms (SNPs) (Everett *et al.* 2011; Nussberger *et al.* 2013). Therefore, in the present study we relied on NGS to generate both genomic and transcriptomic sequences to develop a robust panel of nuclear InDel markers for the ease and versatile identification of *Calanus* spp.

Material and methods

Genome Ion Torrent PGM sequencing

Samples of *Calanus finmarchicus* and *C. glacialis* for genome sequencing were collected from two areas with minimum sympatry: *C. finmarchicus* close to Svinøy island (Norwegian Sea) and *C. glacialis* – in Rijpfjorden (Svalbard) (Table S1). Samples were collected using a WP2 net from 200m to the surface and stored in 95 % non-denatured ethanol. Preliminary identification of species was conducted by eye using prosome length criteria (Kwasniewski *et al.* 2003) and individuals with a prosome length far below/above the delimitation threshold were selected for further DNA extraction.

DNA was extracted individually using the E.Z.N.A.® Insect DNA Kit (Omega Bio-Tek) according to manufacturer´s instruction. To ensure correct species identification, we used 6 microsatellite loci previously developed for *Calanus finmarchicus*: EL696609, EL585922, EH666870 (Provan *et al.* 2007); FK868270, FK670364, FK867682 (Parent *et al.* 2012). PCR reactions were performed in a total volume of 5 µl and contained 3 ng of DNA, reverse and fluorescently labeled forward primers with a final concentration of 0.25 µM each and 1X AmpliTaq Gold® PCR Master Mix (LifeTechnologies). A Veriti 96-Well Fast Thermal Cycler (LifeTechnologies) was used for PCR reactions with following cycling parameters: an initial denaturation step at 95 °C (10 min) followed by 40 cycles of 95 °C (20 s), 54 °C (for EL585922 and FK670364) or 56 °C (remaining 4 loci) (20 s), 72 °C (20 s) and a final extension at 72 °C (10 min). Fragment analysis was performed for all 6 loci in a single multiplex run on a 3500xL Genetic Analyzer (LifeTechnologies) following the manufacturer´s protocol. Alleles were scored using GENEMAPPER 3.7 (LifeTechnologies) and checked by eye. Species identification was based on results of STRUCTURE (v. 2.3.4) (Pritchard *et al*. 2000).

A single individual of each species was selected and 100 ng of DNA was subsequently used for library preparation. The Ion Plus Fragment Library Kit (Life Technologies) was used

for the construction of 200 bp libraries according to manufacturer´s protocol. One library was constructed per species. Emulsion PCR was done in the Ion One Touch System, using the Ion One Touch 200 Template Kit (LifeTechnologies) followed by enrichment and quality control following the manufacturer´s protocol. Sequencing of each library was performed in the Ion Personal Genome Machine® (PGM™) System (LifeTechnologies) using the Ion PGM 200 Sequencing Kit and one Ion 316 chip (LifeTechnologies) per library according to manufacturer´s protocols.

Transcriptome Ion Torrent PGM sequencing

Samples for transcriptome sequencing were obtained from Disko Bay (West Greenland) (Table S1). Live females were morphologically identified using prosome length criteria and pigmentation/redness (Nielsen *et al*. submitted) and preserved in RNAlater (Qiagen). RNA and DNA were extracted simultaneously from each individual using the E.Z.N.A. DNA/RNA Isolation Kit (Omega Bio-Tek) according to manufacturer´s instruction.

Microsatellite genotyping was performed as previously described to ensure correct species identification. An equal amount of total RNA was pooled from 30 individuals per species. Pooled RNAs were treated with Baseline-ZERO DNase (Epicentre) and subsequently cleaned using the RNA Clean & Concentrator kit (Zymo Research, USA). Isolation of mRNA from total RNA was performed using two rounds of Mag-Bind mRNA Enrichment (Omega Bio-Tek, USA) with subsequent cleaning and concentration using the RNA Clean & Concentrator kit (Zymo Research, USA). Libraries were constructed using 40 ng of Poly(A) RNA and the Ion Total RNA-Seq Kit for AB Library Builder System (LifeTechnologies). Emulsion PCR was done in the Ion One Touch 2 System, using the Ion PGM Template OT2 200 Kit (LifeTechnologies) followed by enrichment and quality control following the

manufacturer´s protocol. Sequencing of each library was performed in the Ion PGM System (LifeTechnologies) using the Ion PGM 200 Sequencing Kit and one Ion 318 chip (LifeTechnologies) for each species.

Bioinformatics

Sequences were analyzed using GENEIOUS 6.1.6 (Biomatters, available from http://www.geneious.com). Reads of genomes and transcriptomes for each species were analyzed separately using the following strategy. Reads were quality trimmed with an error probability limit of 0.05 and *de novo* assembled with medium-low sensitivity settings into four separate assemblies. To identify homologous regions between the two species, the 1000 longest genomic contigs and the 700 longest transcriptomic contigs of *C. finmarchicus* were used for BLAST searches (Megablast, default settings) against custom databases of genomic and transcriptomic contigs of *C. glacialis* respectively. Aligned regions between contigs were then visually scanned for large insertions/deletions between the two species using the criteria: InDel size > 7 nucleotides for genome, ≥ 3 for transcriptome, and suitable conserved flanking regions for primer development. Primers were then designed using PRIMER3 (Koressaar & Remm 2007; Untergrasser *et al.* 2012) incorporated into GENEIOUS 6.1.6 (Biomatters).

Marker optimization and multiplexing

InDel optimization and multiplexing have been performed with DNA of two *C. finmarchicus* and two *C. glacialis* individuals identified as specimens for transcriptome Ion Torrent PGM sequencing (Table S1). PCR amplification of the candidate markers were tested using a Veriti 96-Well Fast Thermal Cycler (LifeTechnologies) in a total volume of 5 µl with 3 ng of DNA, 0.25 µM of each primer and 1X AmpliTaq Gold PCR Master Mix

(LifeTechnologies). Cycling parameters included an initial denaturation step at 95 °C (10 min) followed by 40 cycles of 95 °C for 20 s, 55 °C for 20 s, 72 °C for 25 s, and a final extension at 72 °C for 20 min. Amplifications were checked on 2.7% agarose gels, and markers failing to amplify, showing multiple bands or no size difference between the two species were discarded. For the rest of the markers, new forward PCR primers were obtained with fluorescent labeling (VIC, 6-FAM, PET or NED) and tested using a 3500XL Genetic Analyzer (LifeTechnologies).

InDels test

To test validity of the new marker panel, 12 samples of *C. finmarchicus* and *C. glacialis* were obtained from both the West Greenland Sea and East Greenland Sea (Table S1). *Calanus* individuals were sampled with a WP2 net and preserved in 75 % non-denatured ethanol. Molecular identification based on a mitochondrial 16S rDNA fragment was performed according to Lindeque *et al.* (1999; 2006) with minor modifications and using half of the copepod. Modifications consisted of GoTAQ DNA polymerase (Promega) usage for PCR with the following cycling parameters: initial denaturation step at 94 °C (5 min) followed by 40 cycles of 94 °C (1 min), 45 °C (2 min), 72 °C (1 min), and a final annealing phase at 45 °C (2 min) and an extension phase at 72 °C (5 min). From the remaining half of each copepod DNA was extracted using the E.Z.N.A. Insect DNA Kit (Omega Bio-Tek) according to manufacturer´s instruction. Microsatellite loci were genotyped as described above in the Genome Ion Torrent PGM sequencing section. InDels were genotyped in four multiplex PCRs as follow: (1) G_150, G_155, G_461 and G_701; (2) T_595, T_1338 and T_6474; (3) T_1301 and T_1966; (4) T_4700, T_3133 and T_461. PCR conditions were as described above in the Marker optimization and multiplexing section.

InDel sequencing and annotation

To ensure that homologous regions were amplified in both *C. finmarchicus* and *C. glacialis*, two individuals of each species (Table S1) were sequenced for each marker. PCR products were cleaned with ExoSAP-IT (Affymetrix), and sequenced using a BigDye Terminator v3.1 Cycle Sequencing Kit (LifeTechnologies) on a 3500XL Genetic Analyzer (LifeTechnologies) following the manufacturer´s protocols. Sequences were analyzed and aligned in GENEIOUS 6.1.6 (Biomatters) and deposited in the Genbank. BLASTN searches against nucleotide and EST databases at NCBI (http://blast.ncbi.nlm.nih.gov) were used to annotate the regions/genes containing the InDels. Hits with E-value > 1e-10 were ignored.

Cross-amplification with other Calanus spp.

The usability of the 12 InDel markers was tested for *Calanus* species found in the North Atlantic and in the Arctic Oceans. DNA was extracted from two specimens of each of the following species: *C. hyperboreus* (Disko Bay), *C. helgolandicus* (Claus, 1863) (Bay of Biscay) and *C. marshallae* Frost, 1974 (Alaska) (Table S1) using the E.Z.N.A. Insect DNA Kit (Omega Bio-Tek). To ensure correct species identification we sequenced 400 bp of the 16S rDNA site using the following primers: C_16s_Fd1 (5´-GCCGCGTTAGTGYTAAGGTAGCA-3´) and C_16s_Rd1 (5´-AGAAACCAATCTGACTTRCGTCGA-3´). Amplification reactions were performed in a total volume of 10 µl with 6 ng of DNA, 0.25 µM of each primer and 1X AmpliTaq Gold PCR Master Mix (LifeTechnologies). Cycling parameters included an initial denaturation step at 95 °C (10 min) followed by 40 cycles of 95 °C, 62 °C, 72 °C for 20 s each step, and a final extension at 72 °C for 3 min. Amplifications were checked on 1 % agarose gel, and cleaned

with ExoSAP-IT (Affymetrix). Sequencing was as described above and sequences were compared against nucleotide NCBI databases (http://blast.ncbi.nlm.nih.gov) using BLASTN and deposited in the Genbank. InDel amplification and genotyping was performed as described above in the Marker optimization and multiplexing section.

Following the results of cross-species amplification, we tested different combinations of markers and amplification conditions to provide an easy and inexpensive protocol for high-throughput *Calanus* species identification in a single PCR capillary electrophoresis on automatic sequencers.

Results

For genome and transcriptome sequencing only individuals with morphological identification confirmed by nuclear microsatellites (probability of belonging to the cluster of either species at 99 %) were used. The sequencing of genomic libraries resulted in 3,118,080 reads for *C. finmarchicus* and in 1,920,496 reads for *C. glacialis*. After quality trimming, for *C. finmarchicus* 2,754,339 reads (mean length = 188 nt) were assembled into 36,134 contigs (mean contig length = 835 nt), and for *C. glacialis* 1,042,632 reads (mean length = 83.2 nt) were assembled into 52,597 contigs (mean contig length = 185 nt). The sequencing of transcriptome libraries of *C. finmarchicus* and *C. glacialis* resulted in 4,894,166 and 3,412,784 reads respectively. Quality trimming reduced number of reads to 3,548,728 (mean length = 108 nt) and 2,843,610 (mean length = 112 nt) respectively, and following assembly produced 251,042 (mean contig length = 225 nt) and 242,602 (mean contig length = 298 nt) contigs.

Primers were developed for 48 markers (InDel size ≥ 7nt) located in 43 genomic contigs and 31 markers (InDel size ≥ 3 nt) located in 30 transcriptomic contigs. Among those,

four genome-based and eight transcriptome-based markers showed clear bands of a consistently different size between *C. finmarchicus* and *C. glacialis* (Table 1) and was consistent with other identification methods (Table S1). Sequencing the fragments for both species confirmed the homology of the amplicons. Only two markers could be annotated using BLAST: T_595 similar to H+ transporting ATP synthase gene (Genbank FK040981.1, E-value = 7e-26), and T_1966 similar to CDA02 protein (EL696767.1, E-value= 1e-33).

All 12 InDel markers successfully amplified in all 48 samples of *C. finmarchicus* and *C. glacialis* from Greenland. The length of the PCR product for each marker was species specific, and species identification based on InDels was in agreement with morphological and molecular (mtDNA and microsatellite) identifications. No hybrids were found in our samples.

Sequencing of the 16S rDNA region of *C. hyperboreus*, *C. helgolandicus* and *C. marshallae* confirmed their morphological identification. Most of the InDel loci also amplified in these three species, and the results of cross species amplification are presented in Table 1. The following loci were diagnostic: G_155 and T_461 for *C. hyperboreus*, T_3133 for *C. helgolandicus* and *C. marshallae.*

We were able to multiplex six InDel markers in a single PCR, allowing efficient discrimination of all five species. Amplification was performed in a single PCR reaction in a total volume of 5 µl with 3 ng of DNA, 1X AccuStart™ II PCR ToughMix (Quanta BioSciences) and the following combination of markers and primer concentrations (equal forward and reverse): G_150 (13 µM), T_461 (27 µM), T_1338 (22 µM), T_1966 (16 µM), T_3133 (25 µM) and T_4700 (22 µM). Cycling protocol was shortened to 40 minutes in total and included an initial denaturation step at 94 °C (2 min) followed by 35 cycles of 94 °C (10 s), 55 °C (10 s), 72 °C (10 s), and a final extension at 72 °C for 5 min.

Discussion

In less than a decade, next-generation sequencing (NGS) technologies have fundamentally changed our approach to the genomes of non-model species. NGS approaches considerably reduce the per-base sequencing cost, while dramatically increasing the number of bases sequenced, by sequencing DNA in a massively parallel fashion (Metzker 2009). Initially reserved to large-scale projects, such as whole human genomes, the recent introduction of a new generation of instruments (Ion Torrent, MiSeq, 454 Junior) has opened up the use of NGS to smaller and cheaper projects (Quail *et al*. 2012). Among the new instruments, the major benefits of the Ion Torrent platform are sequencing speed and low upfront and operating costs (Perkel 2011). Indeed, sequencing costs on the Ion Torrent are <1000 USD for 1 Gb (Quail *et al*. 2012) with run time < 1/2 day. It is now economically feasible for individual laboratories to employ genome/transcriptome sequencing of non-model organisms in order to develop relevant molecular markers.

Our approach shows that both genome and transcriptome sequencing on the Ion Torrent can be used as useful tools for the development of molecular markers for copepods. Although we did not formally quantify the presence of InDels, they appeared to be more frequent and longer in the genomic contigs compared to the transcriptome based ones. However, it was generally easier to develop primers for the transcriptome based markers and amplification was also more successful. If resources were limited we would thus recommend relying on transcriptome sequencing. One of the advantages of genome sequencing is that only a small amount of DNA is required for library preparation (ca. 100 ng) compared to the amount of total RNA required (ca. 2000 ng) for transcriptome sequencing. Furthermore tissue preservation for RNA extraction can be more difficult than for DNA.

The panel of 12 nuclear markers we have developed allows the identification of all species of *Calanus* present in the North Atlantic/Arctic Oceans. The markers are easy to amplify and genotyping can be done in relatively high throughput using an automatic sequencer. Furthermore, the absence of stutter and the size difference between species allows reliable automatic scoring using software such as GENEMAPPER or GENEIOUS. For some of the markers, the size difference is sufficient for species discrimination using agarose gels (Fig. 1). This will permit the use of these markers in a low tech/low cost setting such as in the field or on board of a research vessel.

The large number of diagnostic markers between *C. finmarchicus* and *C. glacialis* we have developed will contribute to the in-depth study of hybridization between the two taxa. Hybrids between *C. finmarchicus* and *C. glacialis* have been reported in Canada (Parent *et al.* 2012), but we did not detect any in our samples from Greenland. However, the current microsatellite loci available for *Calanus* have insufficient power to separate the different classes of hybrids (*e.g.* F1, F2, backcrosses etc…). The twelve fully diagnostic markers we have developed would allow the resolution of the different hybrid classes with low error (Anderson & Thompson 2002).

In conclusion, the present study illustrates the utility of NGS to easily develop molecular markers for ecologically important non-model species. The correct identification of *Calanus* species in both modern and historical samples is a necessary condition for understanding the response of the North-Atlantic/Arctic ecosystems to climate change.

Acknowledgements

We thank the captains and crews of the following research vessels whom provided great assistance during sampling: R. V. Porsild, R/V “Johan Hjort” (Institute of Marine

Research, Bergen) and R/V “Helmer Hanssen” (University of Tromsø, Tromsø). We would like to thank Mette D. Agersted for collecting the *Calanus* samples from the east coast of Greenland, and organizers and participants of AB-820 Arctic Marine Zooplankton of Svalbard Waters course at UNIS (The University Centre in Svalbard). We are grateful to present and past staff of SAHFOS who have contributed to the maintenance of the continuous plankton recorder time series. The study was funded by European Commission FP7 EURO-BASIN (Grant Agreement: 264 933), the Greenland Climate Research Centre (project 6505) and the Norwegian Research Council (project HAVKYST 216578).

References

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217-1229.

Beaugrand G, Reid F, Ibañez F *et al.* (2002) Reorganization of North Atlantic marine copepod biodiversity and climate. *Science*, **296**, 1692-1694.

Beaugrand G, Mackas D, Goberville E (2013) Applying the concept of the ecological niche and a macroecological approach to understand how climate influences zooplankton: Advantages, assumptions, limitations and requirements. *Progress in Oceanography*, **111**, 75–90.

Bucklin A, Frost B, Kocher T (1995) Molecular systematics of six *Calanus* and three *Metridia* species (Calanoida, Copepoda). *Marine Biology*, **121**, 655-664.

Chambers G, MacAvoy E (2000) Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B*, **126**, 455-476.

Chust G, Castellani C, Licandro P, Ibaibarriaga L, Sagarminaga Y, Irigoien X (2013) Are *Calanus* spp. shifting poleward in the North Atlantic? A habitat modelling approach. *ICES Journal of Marine Science* (In press) doi:10.1093/icesjms/fst147.

Choi J-W, Liao X, Park S et al. (2013) Massively parallel sequencing of chikso (Korean brindle cattle) to discover genome-wide SNPs and InDels. *Molecules and Cells*, **36**, 203-211.

Curtu A, Finkeldey R, Gailing O (2004) Comparative sequencing of a microsatellite locus reveals size homoplasy within and between european oak species (*Quercus* spp.). *Plant Molecular Biology Reporter*, **22**, 339-346.

Daase M, Vik J, Bagoien E *et al.* (2007) The influence of advection on *Calanus* near Svalbard: statistical relations between salinity, temperature and copepod abundance. *Journal of Plankton Research*, **29**, 903-911.

Everett M, Grau E, Seeb J (2011) Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. *Molecular Ecology Resources*, **11**, 93-108.

Falk-Petersen S, Pavlov V, Timofeev S *et al.* (2007) Climate variability and possible effects on arctic food chains: The role of *Calanus*. In: Arctic alpine ecosystems and people in a changing environment ed. Ørbæk JB, Kallenborn R *et al*.), pp. 147-166. Springer-Verlag, Berlin.

Falk-Petersen S, Mayzaud P, Kattner G *et al.* (2009) Lipids and life strategy of Arctic *Calanus*. *Marine Biology Resources*, **5**, 18-39.

Gabrielsen T, Merkel B, Søreide J *et al.* (2012) Potential misidentifications of two climate indicator species of the marine arctic ecosystem: *Calanus glacialis* and *C. finmarchicus*. *Polar Biology*, **35**, 1621-1628.

Head E, Harris L, Yashayaev I (2003) Distributions of *Calanus* spp. and other mesozooplankton in the Labrador Sea in relation to hydrography in spring and summer (1995–2000). *Progress in Oceanography*, **59**, 1-30.

Hill R, Allen L, Bucklin A (2001) Multiplexed species-specifc PCR protocol to discriminate four N. Atlantic *Calanus* species, with an mtCOI gene tree for ten *Calanus* species. *Marine Biology*, **139**, 279-287.

Hirche H, Hagen W, Mumm N *et al.* (1994) The northeast water Polynya, Greenland Sea. 3. Mesozooplankton and macrozooplankton distribution and production of dominant herbivorous copepods during spring. *Polar Biology*, **14**, 491-503.

Hirche H, Kosobokova K (2007) Distribution of *Calanus finmarchicus* in the northern North Atlantic and Arctic Ocean—Expatriation and potential colonization. *Deep-Sea Research II*, **54**, 2729-2747.

Hunter M, Hart K (2013) Rapid microsatellite marker development using next generation pyrosequencing to inform invasive burmese python—*Python molurus bivittatus*—management. *International Journal of Molecular Sciences*, **14**, 4793-4804.

Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics*, **23**, 1289-1291.

Kwasniewski S, Hop H, Falk-Petersen S *et al.* (2003) Distribution of *Calanus* species in Kongsfjorden, a glacial fjord in Svalbard. *Journal of Plankton Research*, **25**, 1-20.

Lindeque P, Harris R, Jones M *et al.* (1999) Simple molecular method to distinguish the identity of *Calanus* species (Copepoda: Calanoida) at any developmental stage. *Marine Biology*, **133**, 91-96.

Lindeque P, Hay S, Heath M *et al.* (2006) Integrating conventional microscopy and molecular analysis to analyse the abundance and distribution of four *Calanus* congeners in the North Atlantic. *Journal of Plankton Research*, **28**, 221-238.

Liu Z, Cordes J (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture*, **238**, 1-37.

Lønne O, Gulliksen B (1989) Size, age and diet of polar cod, *Boreogadus Saida* (Lepechin 1773), in ice covered waters. *Polar Biology*, **9**, 187-191.

Maar M, Møller E, Gürkan Z *et al.* (2013) Sensitivity of *Calanus* spp. copepods to environmental changes in the North Sea using life-stage structured models. *Progress in Oceanography*, **111**, 24-37.

Metzker ML (2009) Sequencing technologies — the next generation. *Nature Reviews Genetics*, **11**, 31-46.

Nagy L, Kocsube S, Csanadi Z *et al.* (2012) Re-mind the gap! Insertion – deletion data reveal neglected phylogenetic potential of the nuclear ribosomal internal transcribed spacer (ITS) of fungi. *PloS ONE*, **7**, e49794.

Nie X, Wang L, Liu P *et al.* (2012) Development of chromosome-arm-specific microsatellite markers in *Triticum aestivum* (Poaceae) using NGS technology. *American Journal of Botany*, e369-e371.

Nussberger B, Greminger M, Grossen C *et al.* (2013) Development of SNP markers identifying European wildcats, domestic cats, and their admixed progeny. *Molecular Ecology Resources*, **13**, 447-460.

Parent G, Plourde S, Turgeon J (2011) Overlapping size ranges of *Calanus* spp. off the Canadian Arctic and Atlantic Coasts: impact on species' abundances. *Journal of Plankton Research*, **33**, 1654-1665.

Parent G, Plourde S, Turgeon J (2012) Natural hybridization between *Calanus finmarchicus* and *C. glacialis* (Copepoda) in the Arctic and Northwest Atlantic. *Limnology and Oceanography*, **57**, 1057-1066.

Pepinski W, Abreu-Glowacka M, Koralewska-Kordel M *et al.* (2013) Population genetics of 30 INDELs in populations of Poland and Taiwan. *Molecular Biology Reports*, **40**, 4333-4338.

Perkel J (2011) Making contact with sequencing's fourth generation. *BioTechniques*, **50**, 93-95.

Pompanon F, Bonin A, Bellemain E *et al.* (2005) Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics*, **6**, 847-859.

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945-959.

Provan J, Beatty G, Maggs C *et al.* (2007) Expressed sequence tag-derived microsatellites for the cool-water marine copepod *Calanus finmarchicus*. *Molecular Ecology Notes*, **7**, 1369-1371.

Quail MA, Smith M, Coupland P *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*, **13**, 341.

Selkoe KA, Toonen R (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615-629.

Søreide J, Falk-Petersen S, Hegseth E *et al.* (2008) Seasonal feeding strategies of *Calanus* in the high arctic Svalbard region. *Deep- Sea Research Part II Topical Studies in Oceanography*, **55**, 2225-2244.

Untergrasser A, Cutcutache I, Koressaar T *et al.* (2012) Primer3 - new capabilities and interfaces. *Nucleic Acids Research*, **40**, e115.

Väli Ü, Brandström M, Johansson M *et al.* (2008) Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. *BMC Genetics*, **9**, 8.

Wassmann P, Duarte C, Agusti S *et al.* (2011) Footprints of climate change in the Arctic marine ecosystem. *Global Change Biology*, **17**, 1235-1249.

Data Accessibility

DNA sequences: Genbank accessions KF913026-KF913049 for InDels; KF956848-KF956853 for 16S rDNA; NCBI SRA: SRX456648, SRX456652, SRX456654, SRX456657, SRX456658, SRX456659 (*C. finmarchicus* transcriptome); SRX456669, SRX456672, SRX456674, SRX456676, SRX456677, SRX456679 (*C. glacialis* transcriptome); SRX456026 (*C. finmarchicus* genome); SRX456144 (*C. glacialis* genome).

Genome and transcriptome contigs and genotypes (microsatellites) of individuals for next-generation sequencing: DRYAD entry doi: 10.5061/dryad.n8f30.

Author Contribution

I.S. collected *Calanus* samples, performed NGS, bioinformatics analysis, and genotyping. S.K. performed NGS and genotyping. M.P. performed genotyping. T.G.N., P.L., E.F.M, C. C and L.B-B collected samples. P.L. performed mtDNA analysis. G.H. designed the study and contributed to all aspect of the research. All authors contributed to the manuscript.

Supplementary Data

Table S1. Sampling location and species identification of *Calanus*

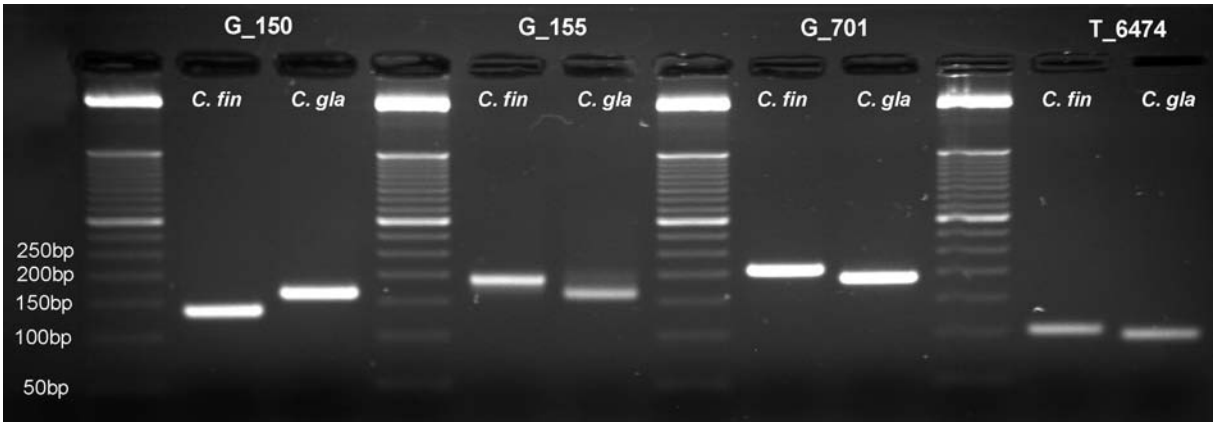

Figure 1. *Calanus* species identification using a 2.7% agarose gel and 4 InDels markers.

Table 1. *Calanus* InDel markers.

Marker	Primers	Amplicon size					Accession number
		C. finmarchicus	*C. glacialis*	*C. helgolandicus*	*C. hyperboreus*	*C. marshallae*	
G_150	F: GACGCCATTGACCATCCAGT R: GCTCCAGCGGTTAGGTTTCT	131	161	na	na	161	KF913026, KF913030
G_155	F: AGAACAACTTGAGCTTATGGA R: CAGCACAATCTTCACATTCA	188	161	170, 188	169	161	KF913027, KF913031
G_461	F: CCTGTTGCTTCAAGGTCAAA R: CTCAGGTGGATCAACCCCC	166	157	166	166	157,170	KF913028, KF913032
G_701	F: GTGGACATAGTTTACTGAAAA R: GTGAGAATGTGAGTAGAGGGCA	209	180, 191	na	na	180, 191	KF913029, KF913033
T_461	F: TGTGAAATGGCGGCCTAACA R: ACAGTACATTAAAATTATGAGCTCGCA	143	136	na	170	136	KF913034, KF913042
T_595	F: GACTTCTCCGTGAGCTCTCC R: ACCGATGACAGAGACAACCTG	68	65	68	na	65	KF913035, KF913043
T_1301	F: CCTCCTCACCCAACTCATCC R: GGATTCATATATTCAAACAAGATATCC	66	76	na	na	76	KF913036, KF913044
T_1338	F: GACTACTCCACCATCTCCGAC R: AAGACTACGGCATGTGTTG	75	72	na	na	72	KF913037, KF913045

T_1966	F: GCCATCGCTCACAAGATCCA R: CCCTTGCTTCCTGGGACATAG	102	99	102	102	99	KF913038, KF913046
T_3133	F: ACGTGAATCTCCTGTGATTTCTGA R: TTTCAAATCAGTCGAAAGCCGT	115	107	110	na	126	KF913039, KF913047
T_4700	F: TGAGGAGGAACACGTACAAGG R: TGTAGATCACCGTGCCATTGG	70	67	70	na	67	KF913040, KF913048
T_6474	F: CAAGCGCTCTCCCTCAAGAT R: GGAGATTTAGACCTGGATCTGGAT	97	91	86, 91	na	91	KF913041, KF913049

na – no amplification, G = genomic origin and T = transcriptome origin. The first accession number is given for *C. finmarchicus* sequence; the second is for *C. glacialis* sequence

Table S1 Sampling location and species identification of *Calanus*.

Designated purpose	Species	Number of specimens	Location of collection	Coordinates of location	Collection date	Collected by	Morphological identification	Molecular identification	Accession to generated sequences
Genome sequencing	*C. finmarchicus*	1	Norway: Svinøy Island	62.3685 N, 05.2027 E	août-2012	I. Smolina	Prosome length	6 microsatellites	KF913026-KF913029
Genome sequencing	*C. glacialis*	1	Norway: Rijpfjorden	80.3028 N, 22.3017 E	sept-2012	I. Smolina	Prosome length	6 microsatellites	KF913030-KF913033
Transcriptome sequencing	*C. finmarchicus*	30	Greenland: Disko Bay	69.2333 N, 53.3833 W	avr-2012	E. Moller and P. Lindeque	Prosome length, redness	6 microsatellites	KF913034-KF913041
Transcriptome sequencing	*C. glacialis*	30	Greenland: Disko Bay	69.2333 N, 53.3833 W	avr-2012	E. Moller and P. Lindeque	Prosome length, redness	6 microsatellites	KF913042-KF913049
Marker optimization	*C. finmarchicus*	2	Greenland: Disko Bay	69.2333 N, 53.3833 W	avr-2012	E. Moller and P. Lindeque	Prosome length, redness	6 microsatellites	Not applicable
Marker optimization	*C. glacialis*	2	Greenland: Disko Bay	69.2333 N, 53.3833 W	avr-2012	E. Moller and P. Lindeque	Prosome length, redness	6 microsatellites	Not applicable
InDels test	*C. finmarchicus*	12	West Greenland Sea	69.2333 N, 53.3833 W	mai-2011	T. G. Nielsen	Prosome length, redness	6 microsatellites, 16S rDNA mtDNA, 12 InDels (present study)	Not applicable
InDels test	*C. glacialis*	12	West Greenland Sea	69.2333 N, 53.3833 W	mai-2011	T. G. Nielsen	Prosome length, redness	6 microsatellites, 16S rDNA mtDNA, 12 InDels (present study)	Not applicable
InDels test	*C. finmarchicus*	12	East Greenland Sea	74.3097 N, 20.2507 W	août-2012	M.D. Agersted	Prosome length, redness	6 microsatellites, 16S rDNA mtDNA, 12 InDels (present study)	Not applicable
InDels test	*C. glacialis*	12	East Greenland Sea	74.3097 N, 20.2507 W	août-2012	M.D. Agersted	Prosome length, redness	6 microsatellites, 16S rDNA mtDNA, 12 InDels (present study)	Not applicable
Cross-species amplification	*C. hyperboreus*	2	Greenland: Disko Bay	69.2333 N, 53.3833 W	avr-2012	E. Moller and P. Lindeque	General morphology	16S rDNA mtDNA, 12 InDels (present study)	KF956850, KF956851
Cross-species amplification	*C. helgolandicus*	2	Spain: Bay of Biscay	43.7000 N, 6.1500 W	mai-2003	L. Blanco-Bercia	General morphology	16S rDNA mtDNA, 12 InDels (present study)	KF956848, KF956849
Cross-species amplification	*C. marshallae*	2	USA: Alaska	55.3722 N, 168.175 W	sept-1999	B. Frost	General morphology	16S rDNA mtDNA, 12 InDels (present study)	KF956852, KF956853
InDel sequencing	*C. finmarchicus*	2	Greenland: Disko Bay	69.2333 N, 53.3833 W	avr-2012	E. Moller and P. Lindeque	Prosome length, redness	6 microsatellites, 12 InDels (present study)	KF913026-KF913029, KF913034-KF913041
InDel sequencing	*C. glacialis*	2	Greenland: Disko Bay	69.2333 N, 53.3833 W	avr-2012	E. Moller and P. Lindeque	Prosome length, redness	6 microsatellites, 12 InDels (present study)	KF913030-KF913033, KF913042-KF913049