Received: 19 July 2023 | Revised: 6 October 2023 | Accepted: 16 October 2023

DOI: 10.1111/mec.17183

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Unmasking microsatellite deceptiveness and debunking hybridization with SNPs in four marine copepod species of *Calanus*

Marvin Choquet[1,2] | Apollo M. Lizano[2] | Alan Le Moan[3] | Mark Ravinet[4] | Anusha K. S. Dhanasiri[2,5] | Galice Hoarau[2]

[1]Natural History Museum, University of Oslo, Oslo, Norway

[2]Faculty of Biosciences and Aquaculture, Nord University, Bodø, Norway

[3]CNRS-Sorbonne Université, Station Biologique de Roscoff, Roscoff, France

[4]School of Life Sciences, University of Nottingham, Nottingham, UK

[5]Department of Paraclinical Sciences, Faculty of Veterinary Medicine, Norwegian University of Life Sciences (NMBU), Oslo, Norway

Correspondence
Marvin Choquet, Natural History Museum, University of Oslo, Oslo, Norway.
Email: marvin.choquet@nhm.uio.no

Funding information
Nord University

Handling Editor: Luke Browne

Abstract

Interspecific hybridization events are on the rise in natural systems due to climate change disrupting species barriers. Across taxa, microsatellites have long been the molecular markers of choice to identify admixed individuals. However, with the advent of high-throughput sequencing easing the generation of genome-wide datasets, incorrect reports of hybridization resulting from microsatellite technical artefacts have been uncovered in a growing number of taxa. In the marine zooplankton genus *Calanus* (Copepoda), whose species are used as climate change indicators, microsatellite markers have suggested hybridization between *C. finmarchicus* and *C. glacialis*, while other nuclear markers (InDels) never detected any admixed individuals, leaving the scientific community divided. Here, for the first time, we investigated the potential for hybridization among *C. finmarchicus*, *C. glacialis*, *C. helgolandicus* and *C. hyperboreus* using two large and independent SNP datasets. These were derived firstly from a protocol of target-capture applied to 179 individuals collected from 17 sites across the North Atlantic and Arctic Oceans, including sympatric areas, and second from published RNA sequences. All SNP-based analyses were congruent in showing that *Calanus* species are distinct and do not appear to hybridize. We then thoroughly re-assessed the microsatellites showing hybrids, with the support of published transcriptomes, and identified technical issues plaguing eight out of 10 microsatellites, including size homoplasy, paralogy, potential for null alleles and even two primer pairs targeting the same locus. Our study illustrates how deceptive microsatellites can be when applied to the investigation of hybridization.

KEYWORDS
Calanus finmarchicus, *Calanus glacialis*, homoplasy, hybridization, microsatellite artefacts, target-capture

1 | INTRODUCTION

The mechanisms of reproductive isolation and speciation are of central interest because of their role in determining gene flow between formerly interbreeding populations (Coyne & Allen Orr, 1998). In the case of hybridizing species, reproductive barriers are more or less permeable, allowing the transfer of genetic material between species. Interspecific hybridization appears to be on the rise in natural systems as a consequence of climate change disrupting species barriers (Canestrelli et al., 2017; Chunco, 2014; Larson et al., 2019).

Therefore, more than ever, accurately recording events of natural hybridization and monitoring their consequences, both in the short- and the long-term, are essential to understand the impacts of climate change on ecosystems.

Many reports of interspecific hybridization are based on results from microsatellites. While microsatellites can effectively be used to identify genetically introgressed individuals (i.e. hybrids), under the condition of validating their reliability with simulations (e.g. Coyer et al., 2007), it is important to exercise caution since microsatellites can be misleading (Henriques et al., 2016; Miralles et al., 2023; Parejo et al., 2018; Poelstra et al., 2022). Indeed, the high risk of homoplasy (Chambers & MacAvoy, 2000; Henriques et al., 2016), the frequent occurrence of null alleles (Dakin & Avise, 2004), high mutation rate and difficulties scoring alleles (Pompanon et al., 2005; Selkoe & Toonen, 2006), are well-documented issues inherent to the nature of microsatellites, but are often ignored, potentially leading to false reports of hybridization (e.g. Cairns et al., 2023; Poelstra et al., 2022).

Genome-wide data have much greater power to detect interspecific hybrids compared to a few microsatellites (see examples in Melville et al., 2017), and can provide direct evidence of hybridization without necessary sampling recent hybrid individuals (Fraïsse et al., 2022; Le Moan et al., 2021; Stankowski et al., 2023). Rapid developments in sequencing technologies and protocols now ease the generation of large single nucleotide polymorphism (SNP) datasets for non-model organisms. With their higher resolution, SNPs can estimate frequencies of hybridization more accurately (Cairns et al., 2023; Parejo et al., 2018) and solve contradictory estimates obtained from small numbers of short markers (Miralles et al., 2023).

Copepod species of the genus *Calanus* dominate the zooplankton biomass in the North Atlantic and Arctic Oceans, and are extensively used as climate change indicators (Hays et al., 2005; Jaschnov, 1972; Kosobokova et al., 2011; Reid et al., 2003; Wassmann et al., 2015). Associated with distinct habitats, four species play a key role in marine pelagic food webs as links for energy transfer between primary producers and higher trophic levels (Bonnet et al., 2005; Falk-Petersen et al., 2009). The pseudo-oceanic *Calanus helgolandicus* is found in warmer temperate waters of the North Atlantic (Bonnet et al., 2005; Helaouët & Beaugrand, 2007). The oceanic *C. finmarchicus* lives in colder temperate waters of the North Atlantic (Falk-Petersen et al., 2009; Rees, 1957). *Calanus glacialis* is usually associated with Arctic shelf environments (Conover, 1988), while *C. hyperboreus* is defined as a sub-Arctic and Arctic oceanic species (Broms et al., 2009; Conover, 1988). As marine zooplankton, *Calanus* spp. are among the fastest organisms to respond to climate variations by shifting their distribution range (Beaugrand et al., 2002; Chust et al., 2013; Poloczanska et al., 2016; Villarino et al., 2015), which can lead to dire consequences for local ecosystems (e.g. Beaugrand et al., 2003). Closely monitoring changes in *Calanus* spp. distributions are therefore critical to detect the impacts of climate change.

However, this task has proved difficult due to the morphological likeness among *Calanus* species, challenging the process of species identification (Bucklin et al., 1995; Fleminger & Hulsemann, 1977). Hence, increasingly, molecular markers have been used to reliably identify *Calanus* species, revealing inaccuracies in traditional morphology-based identification methods and helping to redraw the genus' biogeography in the North Atlantic and Arctic Oceans (Choquet et al., 2017, 2018; Gabrielsen et al., 2012; Lindeque et al., 2004, 2006; Parent et al., 2011). By using six nuclear insertion–deletion markers (InDels; Choquet et al., 2017; Smolina et al., 2014), multiple previously overlooked areas of sympatry were unveiled, where up to four *Calanus* species co-occur and reproduce locally (Choquet et al., 2017, 2020; Schultz et al., 2023). Whether these newly found areas of sympatry are ancient or resulted from recent climate change-induced species range shifts remains to be determined. Regardless, these findings raise questions about the potential for hybridization within the genus *Calanus*. Indeed, earlier studies based on microsatellites concluded that *C. finmarchicus* and *C. glacialis* can frequently interbreed and produce fertile hybrids in the East-Canadian Arctic (Parent et al., 2012, 2015). However, recent studies based on six InDel markers (Choquet et al., 2017, 2020) examined over 4400 individuals from 83 locations, including areas of sympatry, as well as 1126 individuals collected monthly in two areas of sympatry on the Norwegian coast during the main reproductive season, but no hybrid was ever detected. The contrast between the results of microsatellites reporting high rates of hybridization in one location and InDels detecting none despite extensive spatio-temporal sampling is striking. The scientific community is currently divided on the reliability of both sets of molecular markers (see Choquet et al., 2021; Parent et al., 2021); hence, a genome-wide approach is necessary to address this issue.

A protocol of genome-reduced representation, based on target-capture, was recently developed for *C. finmarchicus* and proved useful in generating high numbers of SNPs in both *C. finmarchicus* and *C. glacialis* (Choquet et al., 2019). In addition, whole transcriptomes of several *Calanus* species were recently made available and represent a valuable resource to mine for SNPs (Lizano et al., 2022). Using both assets, we here generated two large and independent SNP datasets to investigate the species boundaries of four *Calanus* species collected in various regions of the North Atlantic and Arctic Oceans, including areas of sympatry. We also re-investigated the microsatellite loci originally used to describe hybrids in *Calanus* in order to estimate their power to do so.

2 | MATERIALS AND METHODS

2.1 | Zooplankton sampling and species identification

Zooplankton samples were collected from 17 locations across the North Atlantic and Arctic Oceans (Figure 1) by vertical towing of WP2-type nets with mesh sizes varying from 150 to 200 μM. Samples were immediately preserved in 80%–90% undenatured ethanol, with subsequent change of ethanol after 24 hours, and placed in a freezer at −20°C. The proportion of each *Calanus* species within these samples was assessed genetically in earlier studies (Choquet et al., 2017, 2020; species composition reported here

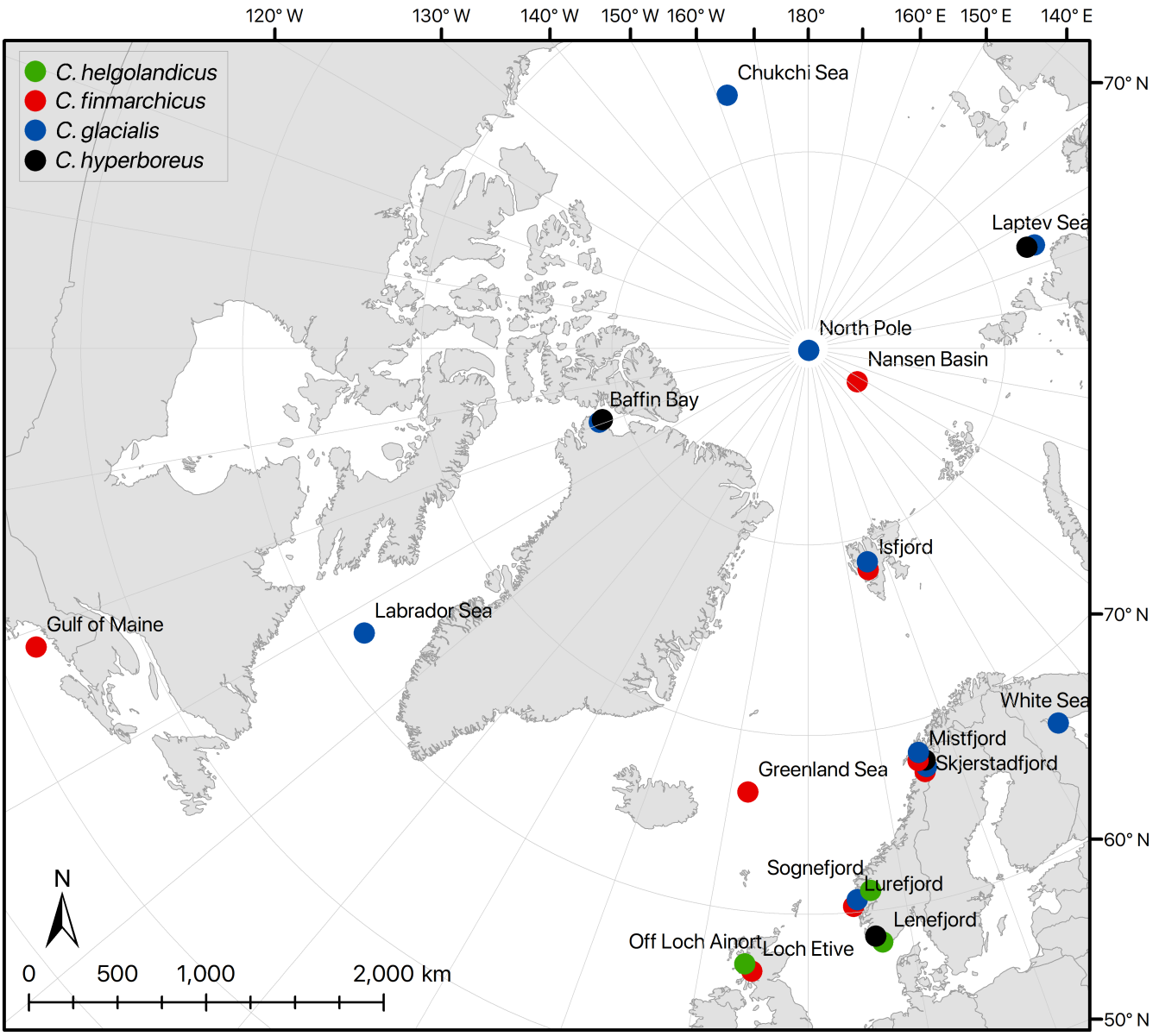

FIGURE 1 Sampling map. Four species of *Calanus* were collected from 17 locations across the North Atlantic and Arctic Oceans to cover various areas of each *Calanus* species distributional ranges, with an effort to include sympatric and allopatric areas for each species. Coloured circles indicate sampling sites per species, in green for *C. helgolandicus*, in red for *C. finmarchicus*, in blue for *C. glacialis*, and in black for *C. hyperboreus*.

in Table 1). In brief, individuals of *Calanus* spp. from developmental stages CIV, CV and adult females were sorted out from each sample and DNA was extracted from their antennas. Six InDel molecular markers (Smolina et al., 2014) were amplified for each specimen and analysed for species identification following the protocol described in Choquet et al. (2017).

2.2 | Genomic DNA library preparation and target-capture

In total, 179 genetically identified specimens were used in a protocol of target-capture for genomic analyses (Table 2; Supplementary Material 1 in Data S1). For the first part of the study, involving four species (*C. helgolandicus*, *C. finmarchicus*, *C. glacialis* and *C. hyperboreus*), we prepared 90 DNA libraries representing five individuals of each occurring *Calanus* species from 11 locations (upper part of Table 2). The second part of the study, focusing on the pair *C. finmarchicus*–*C. glacialis*, included 144 specimens of the targeted species, with eight individuals per species from a total of 17 locations (lower part of Table 2).

The 179 DNA libraries were prepared following the protocol described in Choquet et al. (2019). First, genomic DNA was extracted from genetically identified specimens using the E.Z.N.A. Insect DNA Kit (Omega Bio-Tek). Secondly, DNA libraries were prepared individually using the NEXTflex Rapid Pre-Capture Combo Kit (Bioo

TABLE 1 Sampling locations where *Calanus* spp. were collected for the target-capture part of this study.

Location	Lat.	Long.	Date	Species composition
Gulf of Maine	42.98	−68.86	22.08.2012	CF
Loch Etive	56.45	−5.183	19.09.2016	CF
Off Loch Ainort	56.85	−5.88	31.07.2008	Ch/CF
Greenland Sea	66.7	−7.77	08.05.2013	CF/CG/CH
Lenefjord	58.08	7.16	18.07.2016	Ch/CF/CH
Lurefjord	60.72	5.07	22.06.2016	CF/CG/CH
Sognefjord	61.18	6.58	22.06.2016	Ch/CF
Labrador Sea	62.22	−57.35	26.05.2013	CF/CG/CH
White Sea	66.55	33.72	22.08.2016	CG
Skjerstadfjord	67.18	15.43	26.02.2016	Ch/CF/CG/CH
Mistfjord	67.45	14.83	23.02.2017	Ch/CF/CG/CH
Chukchi Sea	76.41	−162.2	27.07.2016	CG/CH
Laptev Sea	77.31	114.59	10.09.2013	CF/CG/CH
Isfjord	78.32	15.15	05.06.2016	CF/CG/CH
Baffin Bay	78.7	−70.72	06.09.2013	CG/CH
Nansen Basin	87	55.78	04.10.2016	CF/CG
North Pole	89.89	14.3	09.04.2012	CG/CH

Note: Information on latitude ('Lat.') and longitude ('Long.') of sampling sites are provided, together with the date of collection and the name of the collaborator who collected the sample. *Calanus* species composition was assessed genetically at these 17 specific sites in Choquet et al. (2017, 2020). 'CF' corresponds to *C. finmarchicus*, 'Ch' to *C. helgolandicus*, 'CG' to *C. glacialis*, 'CH' to *C. hyperboreus*.

TABLE 2 Sampling sites where *Calanus* spp. were collected for target-capture.

C. helgolandicus	*C. finmarchicus*	*C. glacialis*	*C. hyperboreus*
–	Loch Etive	–	–
Off Loch Ainort	–	–	–
Lenefjord	–	–	Lenefjord
–	Lurefjord	Lurefjord	–
Sognefjord	–	–	–
–	–	Labrador Sea	–
–	–	White Sea	–
–	Skjerstadfjord	Skjerstadfjord	Skjerstadfjord
–	–	Laptev Sea	Laptev Sea
–	Isfjord	Isfjord	–
–	–	Baffin Bay	Baffin Bay
–	Gulf of Maine	–	–
–	Greenland Sea	–	–
–	Mistfjord	Mistfjord	–
–	–	Chukchi Sea	–
–	Nansen Basin	–	–
–	–	North Pole	–

Note:The upper half of the table displays locations included in the first part of the study, involving four species, while the lower half of the table displays additional locations included for the second part of the study, focusing exclusively on *C. finmarchicus* and *C. glacialis*.

Scientific) following the recommendations from the producer, with a few adjustments detailed in Choquet et al. (2019). Individually barcoded libraries were then pooled per species and a set of target-capture probes, initially designed for *C. finmarchicus* and targeting 2656 loci of sizes ranging from 302 to 1500 bp (total length 2,106,591 bp—Choquet et al., 2019), was used to perform two

successive capture reactions on the species-specific pools following the Mycroarray Mybaits protocol v3 and modifications reported in Choquet et al. (2019). Finally, captured library pools were sequenced in paired-ends on a NextSeq 500 (Illumina), with NextSeq 500/550 2×150 bp mid-output kits v2.5.

2.3 | SNP mining

2.3.1 | Target-capture data

Generated sequences were demultiplexed with bcl2fastq (Illumina) v1.8.4, adapter trimmed with cutadapt v1.18 (Martin, 2011) and mapped to the genomic assembly from which the capture probes were designed (i.e. the MaSuRCA assembly in Choquet et al., 2019) using BWA-MEM v0.7.16 (Li, 2013). We used samtools v1.9 (Li et al., 2009) to only retain reads mapping back uniquely, concordantly and in pairs to the reference. Duplicates were removed using Picard tools v2.21.7 (Broad Institute, 2019).

Variants were called separately for each dataset (the 4-species dataset with 90 individuals; and the 2-species dataset with 144 individuals), on all individuals and species together at once using HaplotypeCaller from GATK v4.1.4.1 (Van der Auwera et al., 2013). In each dataset, all variant genotypes were then combined with CombineGVCFS and jointly genotyped with GenotypeGVCFS implemented in GATK. To avoid a potential bias in our analyses if the process of target-capture included mitochondrial regions (unsuitable for analyses of genetic introgression), we performed a BLAST search (Johnson et al., 2008) of the 2656 unique contigs used to design our target-capture probes (Choquet et al., 2019) against a custom-made database comprising all known mitochondrial sequences of *C. finmarchicus* from NCBI (https://www.ncbi.nlm.nih.gov). One 794-bp long contig was identified as potentially mitochondrial and variants identified along it were removed from our datasets.

For subsequent analyses, only high-quality SNPs were kept. Thus, relevant filtering thresholds were determined for each dataset after examination of a subsample of data. Based on this initial analysis, one *C. glacialis* individual from the Baffin Bay location (CG_WGr_92) was excluded from both datasets, due to very low sequencing depth leading to excessive amount of missing data. Two additional *C. glacialis* individuals were excluded from the 2-species dataset, namely 'Chuk_79' from Chukchi Sea, for very low sequencing depth, and 'CG_Mis_652' from Mistfjord for large proportion of missing data compared to other individuals (Supplementary Materials 1 & 2 in Data S1). The 4-species dataset was filtered, using vcftools v0.1.15 (Danecek et al., 2011), as to include only SNPs present in at least 75% of genotypes, with a quality score above 30 and a sequencing depth comprised between 5× and 56× (corresponding to mean + std dev. * 2 of the sequencing depth distribution). The 2-species dataset, focusing on *C. finmarchicus* and *C. glacialis*, was filtered as to include only bi-allelic SNPs present in at least 80% of genotypes, with a quality score above 30, and covered at least 5× and no more than 89× (again, corresponding to mean + std dev. * 2 of the sequencing depth distribution). True multi-allelic SNPs are not expected to be observed frequently, even less when comparing only two species, hence we considered it safer to remove multi-allelic sites from the 2-species dataset as these may reflect sequencing artefacts.

To assess the usefulness of the target-capture protocol developed initially for *C. finmarchicus* and hereby applied on three other species, we counted the number of raw SNPs yielded specifically for each species (using the 4-species dataset) and compared these numbers to the level of identity between each species and the reference. We used vcftools v0.1.15 (Danecek et al., 2011) to split the SNP dataset per species and count SNPs. To estimate the genetic distances between each species and the reference (*C. finmarchicus* draft genomic assembly) to which their sequences were aligned, we used the alignment tool NextGenMap v0.5.0 (Sedlazeck et al., 2013) to map our trimmed FASTQ sequences against the assembly. NextGenMap outputs a score of pairwise identity for each alignment that we averaged per species.

2.3.2 | RNA-seq data

RNA-seq raw reads were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov) for three individuals of each of the four focal species (Table 3). These were adapter and quality trimmed with cutadapt v1.18 (Martin, 2011). Then, reads were aligned to the same genomic assembly as mentioned above using the splice-aware mapper STAR v2.5.4 (Dobin et al., 2013). Alignment files were cleaned using samtools v1.9. (Li et al., 2009) to keep only reads mapping uniquely, concordantly and in pairs to the reference. Duplicates were removed using Picard tools v2.21.7 (Broad Institute, 2019). Variant calling was performed on all 12 individuals together at once following the GATK pipeline of best practices for identification of short variants from RNA-seq data (https://gatk.broadinstitute.org/hc/en-us/articles/360035531192-RNAseq-short-variant-discovery-SNPs-Indels-). The resulting variant dataset was filtered with vcftools v0.1.15 (Danecek et al., 2011), as to only include SNPs present in at least 80% of genotypes, with a quality score above 30 and covered at least 5× and no more than 172×. These filtering thresholds were determined based on an empirical analysis of the generated set of SNPs (Supplementary Material 2 in Data S1).

2.4 | Species boundaries analyses across four *Calanus* species

The 4-species SNP dataset obtained from target-capture was used to reconstruct a neighbour-joining (NJ) tree. All SNPs with a minor allele count of at least two were pruned for linkage disequilibrium (LD) using Plink v1.90 (https://www.cog-genomics.org/plink/; Chang et al., 2015) to generate sets of independent markers of variation. Pruning was performed in sliding windows of 50 SNPs, with

TABLE 3 Information on *Calanus* spp. samples from which RNA-seq data was downloaded from NCBI.

	NCBI ref. ID	Developmental stage	Sampling site Lat./long.	Sampling date	Reference
C. helgolandicus	SRR17245869	CV	56.95/−2.12	04/2019	Lizano et al. (2022)
	SRR17245870	CV	56.95/−2.12	04/2019	
	SRR17245871	CV	56.95/−2.12	04/2019	
C. finmarchicus	SRR1153468	CV	44.03/−68.05	07/2011	Lenz et al. (2014)
	SRR1141107	CV	NA	05/2012	Tarrant et al. (2014)
	SRR1141110	CV	NA	05/2012	Tarrant et al. (2014)
C. glacialis	SRR17240410	CV	67.23/−14.73	06/2019	Lizano et al. (2022)
	SRR17240411	CV	67.23/−14.73	06/2019	
	SRR17240412	CV	67.23/−14.73	06/2019	
C. hyperboreus	SRR17307980	Ad. female	74.57/−11.3	09/2018	Lizano et al. (2022)
	SRR17307981	Ad. female	74.57/−11.3	09/2018	
	SRR17307982	Ad. female	74.57/−11.3	09/2018	

a 10-SNPs increment between windows and a r^2 maximum threshold value set at .8. The remaining SNPs were concatenated into one single nucleotide sequence per individual by randomly drawing one of the two alleles at each locus. Jukes-Cantor genetic distances were estimated among these using the R (v4.0.5) package *Phangorn* v2.11.1 (Schliep, 2011; R Core Team, 2021) and represented in a bootstrapped NJ tree.

Principal component analyses (PCA) and ancestry analyses were performed separately for the target-capture and RNA-seq SNPs. The two datasets were filtered to exclude SNPs with a minor-allele count of less than three (as recommended by Linck and Battey (2019) for population structure analyses as rare variants may confound model-based and multivariate analyses), and subsequently LD-pruned as detailed above. PCAs were performed with Plink (Chang et al., 2015), while ancestry analyses were performed with ADMIXTURE v1.3.0 (Alexander et al., 2009), ran with a five-fold cross-validation and a *K* number of clusters set at values comprised between two and six.

2.5 | Testing for hybrids between *C. finmarchicus* and *C. glacialis*

We further explored the variation in species trees inferred between the most recently diverged species, that is, *C. finmarchicus* and *C. glacialis*, with SNAPP (Bryant et al., 2012) using *C. hyperboreus* as an outgroup. Genotypes shared among *C. finmarchicus*, *C. glacialis* and *C. hyperboreus* were then extracted from the 4-species target-capture SNP dataset and only LD-pruned SNPs with a minor allele count of at least two were analysed. SNAPP relies on a coalescent model and uses Markov chain Monte Carlo (MCMC) to infer multiple species trees from a set of unlinked SNPs. SNAPP was run with 10,000 burn-in steps followed by 100,000 iterations. All the generated trees were represented graphically on top of each other with DensiTree 2 v2.2.7 (Bouckaert & Heled, 2014).

The 2-species SNP dataset obtained from target-capture on *C. finmarchicus* and *C. glacialis* was used to perform a PCA using Plink (Chang et al., 2015), after removing SNPs with a minor-allele count of less than three (Linck & Battey, 2019) and subsequent LD-pruning. The same dataset was also used in an ancestry analysis with ADMIXTURE 1.3.0 (Alexander et al., 2009), ran with a five-fold cross-validation, with *K* number of clusters set at values from two to three.

Then, we evaluated the power of admixture analyses to detect hybrids between *C. finmarchicus* and *C. glacialis* with our 2-species SNP dataset by simulations following Coyer et al. (2007). Individuals from two geographical areas of allopatry per species were used as references (Loch Etive and Gulf of Maine for *C. finmarchicus*, Chukchi Sea and North Pole for *C. glacialis*). We calculated pairwise F_{ST} (Weir & Cockerham, 1984) for each SNP between *C. finmarchicus* and *C. glacialis* reference individuals using vcftools v0.1.15 (Danecek et al., 2011). Then, we extracted the SNPs fixed between *C. finmarchicus* and *C. glacialis* (i.e. SNPs with a pairwise $F_{ST}=1$). This step left us with a set of 182 ancestry-informative SNPs suitable for simulating hybrids between *C. finmarchicus* and *C. glacialis*. Genotypes from reference individuals from both species, at these 182 loci, were used to simulate 40 first-generation (F1) hybrids using HYBRIDLAB v1.0 (Nielsen et al., 2006). Then, 20 backcross hybrids between F1 and *C. finmarchicus* reference individuals were simulated, as well as 20 backcross hybrids between F1 and *C. glacialis* reference individuals. In addition, 20 hybrids of second generation (F2) were simulated based on the 40 F1 genotypes (F1xF1). Genotypes at the 182 loci from the reference individuals, the 100 simulated hybrids and the other sequenced individuals were then used all together in an ancestry analysis with ADMIXTURE, ran with a five-fold cross-validation and *K* set at two. Numbers of heterozygous sites across the 182 loci were counted per individual, including reference individuals, simulated hybrids and sequenced individuals from areas of sympatry. We represented the admixture index against the individual heterozygosity in a triangle-plot with the R package *ggplot2* v3.3.5 (Wickham, 2016).

2.6 | Assessment of microsatellite power to detect putative hybrids

Following the recent public release of the microsatellite genotype datasets from Parent et al. (2012, 2015; dataset: Parent, 2021), from which hybridization between *C. finmarchicus* and *C. glacialis* in the East Canadian Arctic was inferred, we have re-examined these results. The power of admixture analyses to detect hybridization in *Calanus* with the 10 microsatellites was evaluated by simulations using HYBRIDLAB v1.0 (Nielsen et al., 2006), as described in Coyer et al. (2007) and Hoarau et al. (2015). We used the genotypes from the reference stations 'St. 1' (for *C. glacialis*) and 'St. 14' (for *C. finmarchicus*), as defined in Parent et al. (2012), in a Bayesian analysis with STRUCTURE v2.3.4 (Pritchard et al., 2000) inferring $K=2$ clusters, to evaluate the ability of the dataset to distinguish between the two pure species (following a strict assignment model with no admixture and independent allelic frequencies, initial burn-in of 50,000 MCMC steps followed by 250,000 iterations). The analysis revealed at least one admixed individual, which we excluded from our next analyses. The remaining 'pure' individuals were used to simulate four hybrid classes, using HYBRIDLAB v1.0 (Nielsen et al., 2006), as follows: 100 hybrids of first generation (F1), 100 hybrids of second generation (F1 × F1), 100 backcrosses between 'pure' *C. finmarchicus* and F1 and 100 backcrosses between 'pure' *C. glacialis* and F1. Admixture analyses of the four simulated hybrid classes were used to define the boundaries of the admixture coefficient for 'pure' *C. finmarchicus*, 'pure' *C. glacialis*, F1 hybrids and F2 hybrids in STRUCTURE with $K=2$. Then, all the genotypes from Parent (2021) were analysed in STRUCTURE with $K=2$. Four different evolutionary scenarios were tested per admixture analysis, following Henriques et al. (2016): (1) No interbreeding and two independently evolving species (no admixture and independent allelic frequencies); (2) No interbreeding but species sharing a recent common ancestor (no admixture and correlated allelic frequencies); (3) Interbreeding but two distantly related species (admixture and independent allelic frequencies); (4) Interbreeding between recently evolved species (admixture and correlated allelic frequencies).

2.7 | Assessment of microsatellite genotyping bias

The microsatellites used by Parent et al. (2012, 2015) were derived from a *C. finmarchicus* expressed sequence tag library (Parent et al., 2012; Provan et al., 2007). We took advantage of the recent publication of several *Calanus* transcriptomes (Lenz et al., 2014; Lizano et al., 2022) to re-evaluate the original microsatellite loci in a newly available extensive genomic landscape. Thus, we mapped the sequences of each of the 10 microsatellite primer pairs to both *C. finmarchicus* and *C. glacialis* transcriptomes. We used BLASTN (https://blast.ncbi.nlm.nih.gov/) against the transcriptome shotgun assembly BioProjects 236,528 (*C. finmarchicus*) and 744,376 (*C. glacialis*) and kept matches with *E*-values <2 (i.e. expected number of matches by chance; Altschul et al., 1990, 1997). For each primer pairs, matching sequences were retrieved, aligned with the Geneious Aligner (Geneious Prime v2023.1.2, https://www.geneious.com) and primer sequences were mapped with *Primer3* (Kõressaar et al., 2018) implemented in Geneious. Only retrieved sequences including both forward and reverse primers were considered. The resulting alignments were then each checked manually.

3 | RESULTS

3.1 | Target-capture performance in four species of *Calanus*

In total, 179 individual libraries were prepared with the target-capture protocol and sequenced, including 15 individuals of *C. helgolandicus*, 64 individuals of *C. finmarchicus*, 80 individuals of *C. glacialis* and 20 individuals of *C. hyperboreus* (Supplementary Material 1 in Data S1).

On average, nine million paired-end (PE) reads per individual were yielded for *C. finmarchicus*, with about 48% of these mapping back concordantly and in pair to the reference used (Supplementary Material 1 in Data S1), and thereby useful for subsequent analyses. In *C. glacialis*, an average of 19 million PE reads were yielded per individual, with an average of 23% of these mapping back with high quality to the reference used. For two *C. glacialis* individuals, CG_WGr_92 (from West Greenland) and Chuk_79 (from the Chukchi Sea), capture and / or sequencing failed as only very few reads were recovered (Supplementary Material 1 in Data S1). In *C. helgolandicus*, about 11 million PE reads per individual were yielded, with an average of nearly 14% of these mapping back to the reference (Supplementary Material 1 in Data S1). In *C. hyperboreus*, an average of nine million PE reads per individual were yielded and about 12% of these were mapped to the reference (Supplementary Material 1 in Data S1). The large difference in number of reads sequenced among species resulted from deeper sequencing depth allocated to libraries of *C. glacialis*, *C. hyperboreus* and *C. helgolandicus* to compensate for their larger genome size (compared to *C. finmarchicus*) and to account for the fact that the capture probe set was developed originally for *C. finmarchicus* (Choquet et al., 2019).

Variant calling from 90 target-capture libraries including four *Calanus* species yielded 788,407 'raw' SNPs in *C. finmarchicus*, the focal species of the target-capture probes, 1,168,120 SNPs in *C. glacialis*, 450,195 SNPs in *C. helgolandicus* and 195,147 SNPs in *C. hyperboreus*. We observed relatively high levels of missing data in our target-capture datasets (Supplementary Material 2 in Data S1), even for the focal species *C. finmarchicus*, which do not appear to result from low sequencing depth. One possible explanation is that there is a very high proportion of inter-individual structural variation, such as deletion events, within each species (as suggested by unpublished whole-genome preliminary data). Averaged pairwise identity score reported by NextGenMap between each species and the *C. finmarchicus*' genomic reference used for the capture probes was the

highest for *C. finmarchicus*, followed by *C. glacialis*, *C. helgolandicus* and *C. hyperboreus* (Figure S1), in line with the proportions of reads mapping to the *C. finmarchicus* genomic reference in each species (Supplementary Material 1 in Data S1).

3.2 | Two independent SNP datasets support distinct species boundaries among four *Calanus* species

For the first part of the study, including all four species, two independent SNP datasets were generated, one from target-capture (Table 2) and one from published RNA-seq data (Table 3). Our protocol of target-capture applied on 90 individuals including 15 individuals of *C. helgolandicus*, 20 individuals of *C. finmarchicus*, 35 individuals of *C. glacialis* and 20 individuals of *C. hyperboreus* resulted, after filtering, in 2138 independent (unlinked) SNPs located on 276 distinct contigs, in 89 individuals (one individual was removed). From the downloaded RNA-seq reads, variant calling in 12 individuals from four species resulted in 9199 independent SNPs, post-filtering, located along 1659 contigs.

The target-capture dataset was used to reconstruct a neighbour-joining tree (Figure 2). The NJ tree separated all four species without ambiguity. Bootstrap values of the branches supported by more than 60% of the trees are displayed and show a full support for the branches leading to *C. hyperboreus* and *C. helgolandicus*. The lowest bootstrap value of 64 was observed for the branch leading to *C. glacialis* (Figure 2).

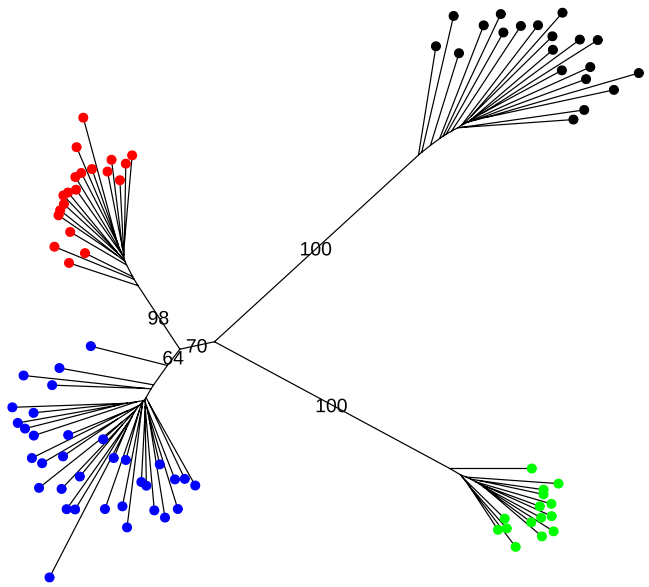

FIGURE 2 Unrooted neighbour-joining tree of four species of *Calanus* based on Jukes-Cantor genetic distances estimated from 2138 unlinked target-capture SNPs. Bootstrap values of the branches supported by more than 60% of the trees generated are displayed on each branch. Species are distinguished by colour: green for *C. helgolandicus*, red for *C. finmarchicus*, blue for *C. glacialis*, and black for *C. hyperboreus*.

A slightly different filtering of the target-capture SNP dataset, aimed for PCA and ancestry analyses, resulted in 1535 independent (unlinked) SNPs located on 256 contigs. The PCA performed with the target-capture dataset revealed a clear separation of *C. hyperboreus* from the other three species on the first axis, which represents 26.2% of the variability (Figure 3a). The second axis, explaining 18.5% of the variability, differentiates the remaining three species. Overall, three main clusters are distinguished from this analysis, corresponding to *C. hyperboreus*, *C. helgolandicus* and the pair *C. finmarchicus*–*C. glacialis* clustering more closely. *Calanus finmarchicus* and *C. glacialis* differentiate on the PC2 (Figure 3a), while one *C. glacialis* individual ('176–34' from Labrador Sea—see Table S1) appears intermediate between the two species. This specific individual had the lowest mean sequencing depth of all sequenced *C. glacialis* individuals (Supplementary Material 2 in Data S1), which likely explains its ambiguity. The PCA with the RNA-seq dataset shows a clear separation between the four species on the two first axes representing 36.3% and 33.2% of the variability respectively (Figure 3b).

Results from ancestry analyses of genetic admixture are presented as bar plots for both datasets: target-capture (Figure 3c) and RNA-seq (Figure 3d). Each individual genotype is represented as a vertical column filled with colour(s). Each colour corresponds to one genetically distinct cluster identified by ADMIXTURE and each vertical column represents the proportion of each cluster/species in the genotype of one individual. In case of gene flow among species, admixed individuals are expected to appear as columns containing a mixture of several colours (i.e. a mixture between different species). An ancestry coefficient of zero or one corresponds to a non-genetically admixed or 'pure' individual. An ancestry coefficient comprised between but not including 0 and 1 corresponds to an individual genetically admixed and potentially hybrid between two genetic clusters. The most strongly supported value for *K*, numbers of genetic clusters in our datasets, was identified as the value with the lowest cross-validation error allowed. For the target-capture SNPs, the most statistically supported number of *K* clusters was four (Figure S2a). These four clusters correspond to the four species of *Calanus* with no apparent genetic admixture among them (Figure 3c). Four individuals of *C. glacialis* show a very low level of genetic admixture with *C. helgolandicus* (less than 3%). For the RNA-seq SNPs, the most statistically supported number of *K* clusters was again four (Figure S2b). Here, absolutely no admixture was found (Figure 3d).

3.3 | No evidence for gene flow between *C. finmarchicus* and *C. glacialis*

The phylogenetic tree generated by SNAPP using 1346 independent SNPs from *C. finmarchicus*, *C. glacialis* and *C. hyperboreus* (used here as the outgroup) shows a perfect overlap of all trees with no apparent ambiguities in the phylogeny of the three species (Figure 4). A single split separates *C. finmarchicus* and *C. glacialis*, with no sign of gene flow occurring after that initial split. Here, in the scenario

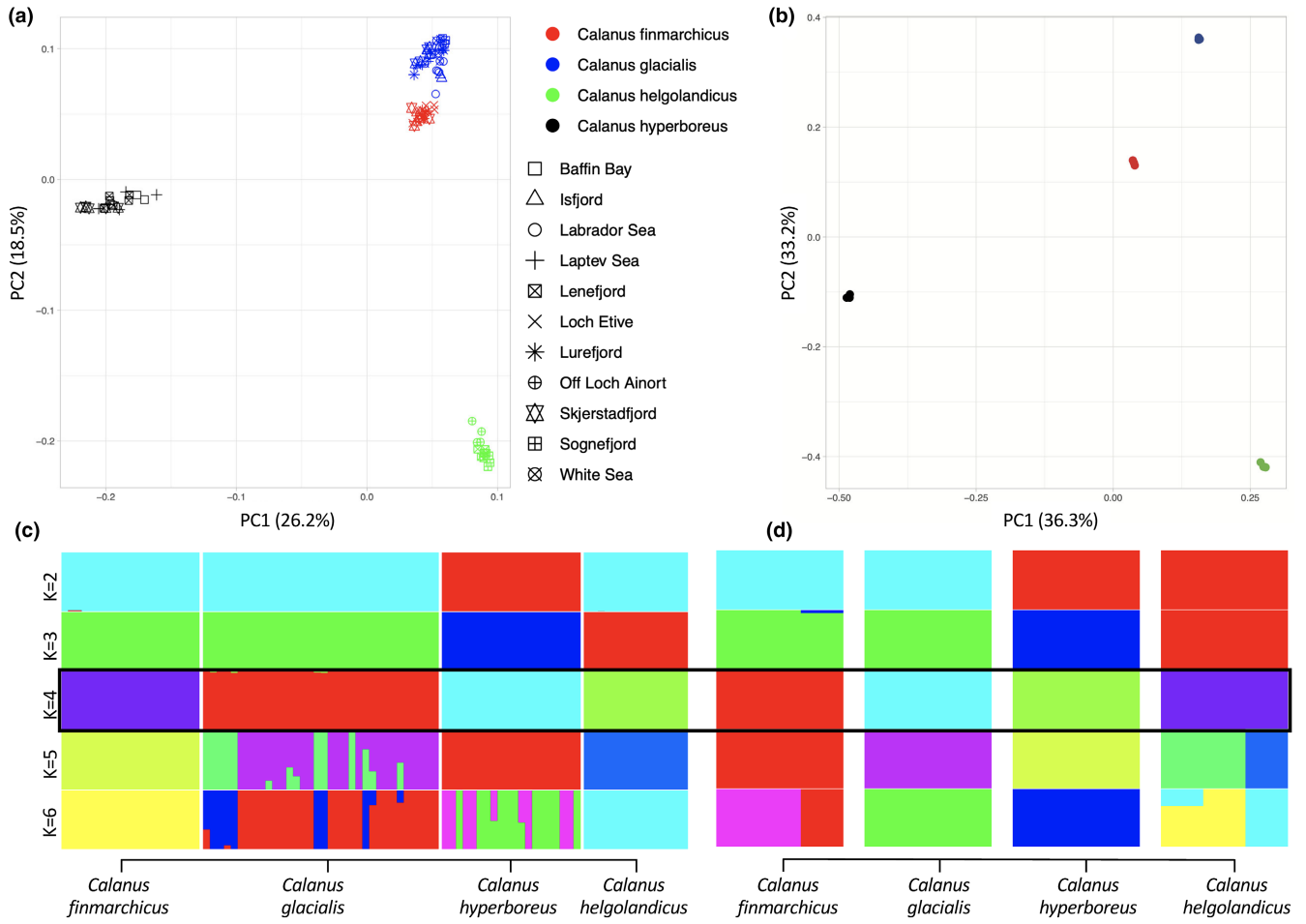

FIGURE 3 Interspecific genetic structure among *Calanus helgolandicus*, *C. finmarchicus*, *C.glacialis* and *C. hyperboreus* revealed by SNPs from target-capture (a, c) and RNA-seq (b, d). (a) Principal component analysis (plot of the two first components) performed with 1535 SNPs obtained from the target-capture protocol, for the four species of *Calanus*. Eighty-nine individuals from four species and from 11 locations were used. Each symbol represents the genotype of one individual, with its shape corresponding to the location and its colour corresponding to the species. (b) Principal component analysis (plot of the two first components) performed with 9199 SNPs mined from published RNA-seq data from 12 *Calanus* spp. individuals, with three individuals per species. (c) Admixture analysis performed with same dataset as (a). (d) Admixture analysis performed with same dataset as (b). For both (c) and (d), the number of inferred clusters *K* was set from two to six. Each individual genotype is represented as a vertical column. The most supported scenario, with the lowest value of cross-validation error, corresponds to $K=4$ for both datasets. The four distinct genetic clusters, in four colours, correspond to the four different species.

where the two species started diverging from a single event and did not hybridize later, we would expect all the trees to converge with the species tree, showing a single split between *C. finmarchicus* and *C.glacialis*. In a scenario where the two species would have diverged after a first event of speciation, and then shared gene flow later, we would expect some parts of the genome, and thus some trees, to show a connection between *C. finmarchicus* and *C.glacialis* branches after the first split of speciation.

For the second part of the study, focusing exclusively on *C. finmarchicus* and *C.glacialis*, 89 additional specimens were added to the previous *C. finmarchicus* and *C.glacialis* for target-capture, resulting in 144 individuals (64 *C. finmarchicus* and 80 *C.glacialis*Table 2). A total of 3,309,961 SNPs were initially yielded from these individuals. After quality filtering, the dataset consisted of 141 individuals and 6030 independent SNPs located along 1266 contigs. The separation between *C. finmarchicus* and *C.glacialis* was evident with individual-based analyses (PCA and admixture), with a clear separation of the two species on the first axis of the PCA representing 17.1% of the variability, and no apparent admixture found between them (Figure 5; Figure S2c: best $K=2$). A greater genetic structure was found within *C.glacialis* compared to *C. finmarchicus*, as observed along PC2 explaining 1.49% of the variability (Figure 5a) and by the admixture analyses for $K=3$ (Figure 5b).

The simulated hybrids between *C. finmarchicus* and *C.glacialis* showed clear differences in hybrid index and observed heterozygosity (Figure 6), as expected from the clear divergence between the two species. Here, F1 and F2 hybrids are expected to show a coefficient of genetic admixture around 50%, with heterozygosity levels at 100% and 50%, respectively, while backcrosses between a given parental species and F1 or F2 hybrids should show around 75% of ancestry for the given parental species and a heterozygosity level of 50%. All sequenced individuals clustered together with the

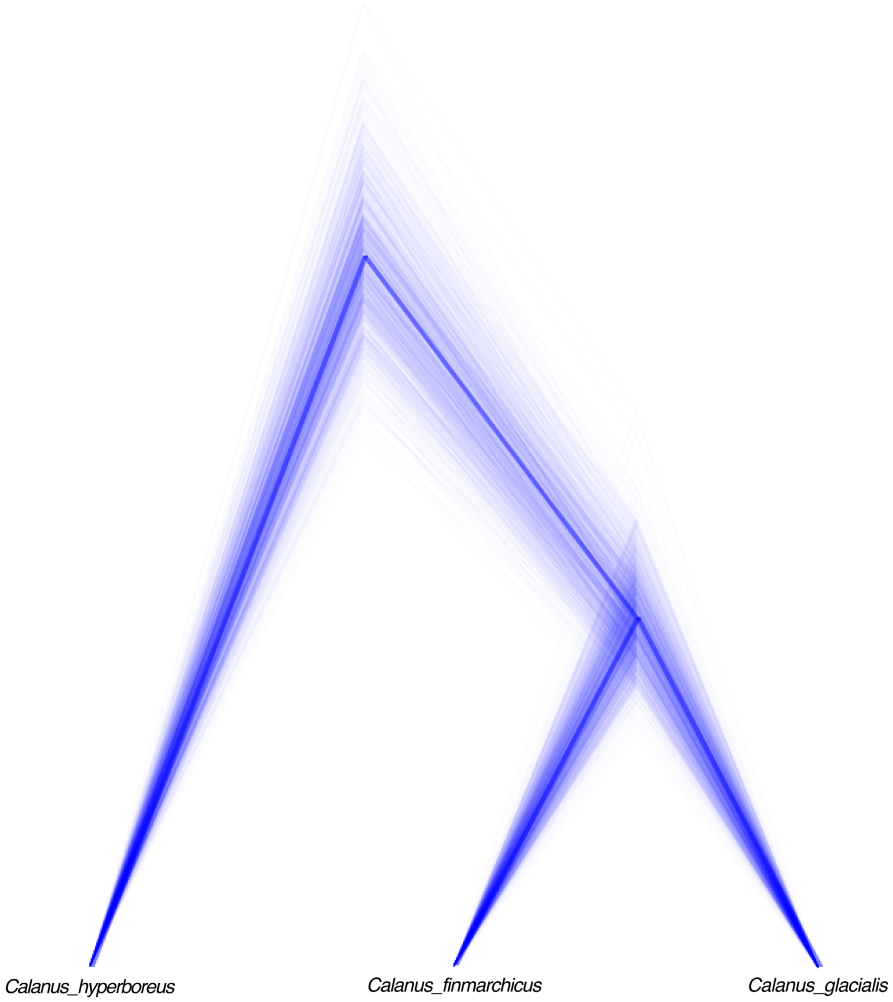

FIGURE 4 Overlap of species trees inferred from SNAPP for *Calanus finmarchicus*, *C. glacialis* and *C. hyperboreus* as the outgroup, represented with DensiTree. Calculations are based on 1346 unlinked SNPs obtained from the protocol of target-capture applied to 74 individuals. SNAPP was run with 10,000 initial burn-in steps followed by 100,000 iterations. The dark line represents the consensus tree.

allopatric reference individuals (Figure 6). The only notable outlier was found in one individual of *C. glacialis* ('176–34' from the Labrador Sea) showing relatively high value of hybrid index (~10%). This is the same individual that appeared intermediate between *C. finmarchicus* and *C. glacialis* in the first PCA (Figure 3a). This individual has, however, the lowest mean sequencing depth (Supplementary Material 2 in Data S1) and a very low heterozygosity, suggesting that its position on the triangle plot is not reliable. Altogether, this supports the absence of recent hybrids in our samples.

3.4 | Microsatellites cannot distinguish between 'pure' *C. finmarchicus*/*C. glacialis* and their putative hybrids

Admixture analyses of four simulated hybrid classes were used to define the boundaries of the admixture coefficient for pure *C. finmarchicus*, pure *C. glacialis*, F1 and F2 hybrids. Subsequent analysis of the dataset from Parent et al. (2021) revealed the inability of the microsatellites to accurately distinguish between the different hybrid classes simulated but also, and more importantly, in some cases between parental species and hybrids (overlapping 'boxes' in Figure S3).

3.5 | The majority of the microsatellite loci suffer from 'fatal' flaws

Mapping of the microsatellite primer sequences to available transcriptomes of *C. finmarchicus* and *C. glacialis* revealed major issues for at least seven of the 10 loci (Table 4, Figure 7). Locus FK868270, developed by Parent et al. (2012) appears to be targeting the exact same locus as locus EH666474 (Provan et al., 2007; Figure 7a) and is therefore redundant and should be completely discarded. As the transcriptome BioProjects used were based on several individuals of each species, we could investigate the possibility of size homoplasy. Indication of size homoplasy was found in four of the remaining nine loci (FG632811, FK867682, EH666870 and EH666474). Sequence variation in the primer binding area (potentially leading to null alleles) was found in three loci (FG632811, FK867682 and EH666870). Locus FK670364 mapped to two clearly different genes (paralogues) with similar expected PCR product sizes. Locus EL585922 mapped to several transcripts, with multiple binding sites for the reverse primer at 20 bp intervals, potentially leading to multiple PCR products in the size range of interest.

In addition, PCR amplification of locus EL773519 (for which mapping did not reveal any issues) clearly showed multiple peaks, thus challenging the validity of these genotypes (Figure S4).

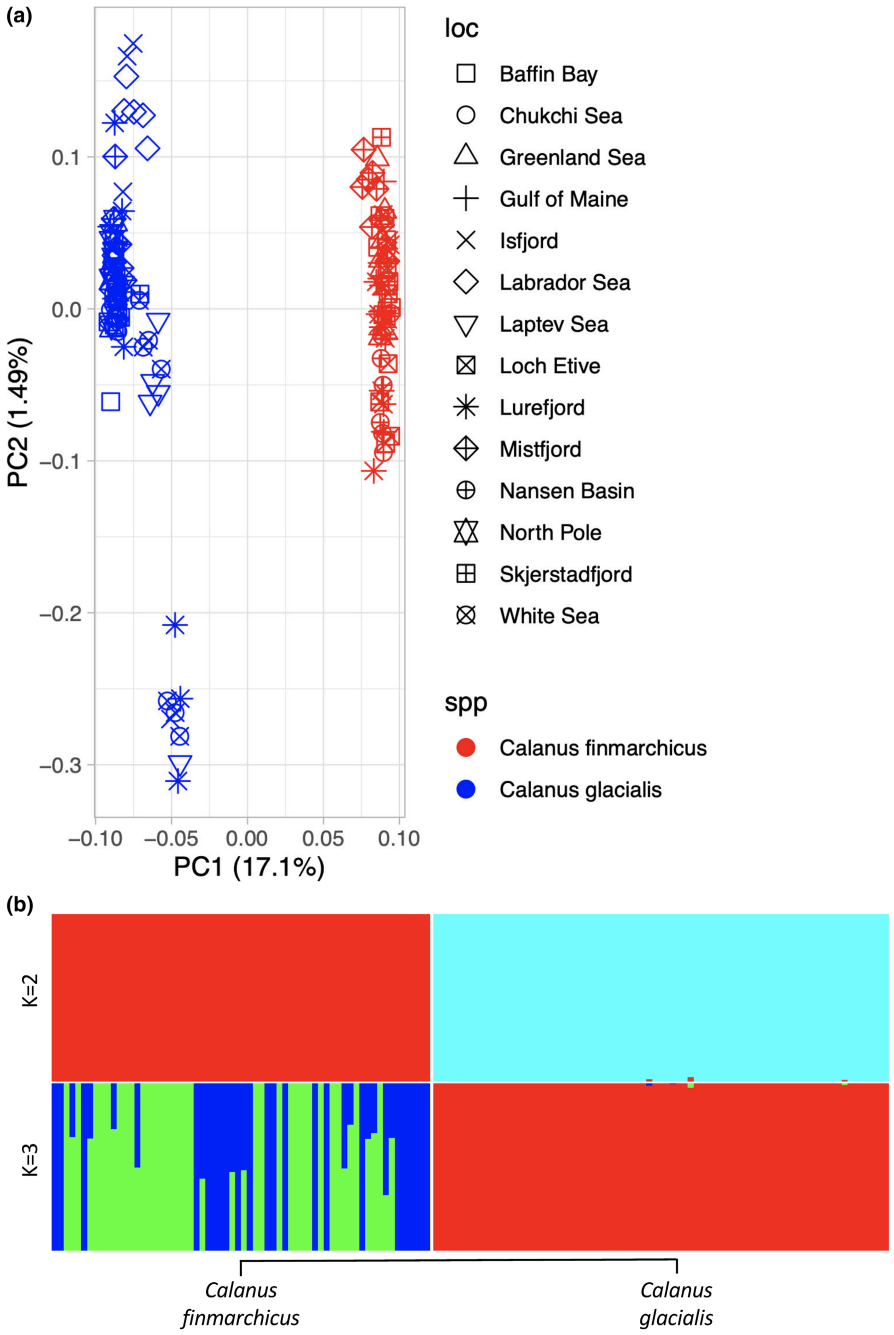

FIGURE 5 (a) Principal component analysis (plot of the two first components) performed with 6030 SNPs obtained with the protocol of target-capture, from 141 individuals of *Calanus finmarchicus* and *C. glacialis* (64 *C. finmarchicus* and 77 *C. glacialis*) from 14 locations. Each symbol represents the genotype of one individual, with its shape corresponding to the location and its colour corresponding to the species. (b) Admixture analysis performed on the same dataset. The number of clusters *K* was set from two to three. The plot corresponding to $K=2$ is the most supported scenario with the lowest value of cross-validation error. This plot shows two distinct genetic clusters, in two colours, corresponding to the two different species: in red *C. finmarchicus* and in blue *C. glacialis*.

These results impair the validity of the microsatellite genotypes from the majority of the loci previously used and indicate that these genotypes should be discarded or at least treated with extreme caution.

4 | DISCUSSION

All our analyses converge towards the same conclusion, that the four *Calanus* species investigated here do not hybridize, nor do they appear to have in the recent past. This confirms previously published results of InDel markers that never detected any hybrids across thousands of individuals (Choquet et al., 2017, 2020), and contrasts substantially with microsatellite results, which suggested hybridization between *C. finmarchicus* and *C. glacialis* (Parent et al., 2012, 2015).

4.1 | Microsatellites: 99 problems but a hybrid ain't one

Putman and Carbone (2014) emphasize the critical need to test the power of microsatellites to detect hybrids by analysing simulated datasets, both prior to and following data collection and analyses. Our simulations based on the microsatellite data newly released by Parent et al. (2021) revealed that, even if we assumed their

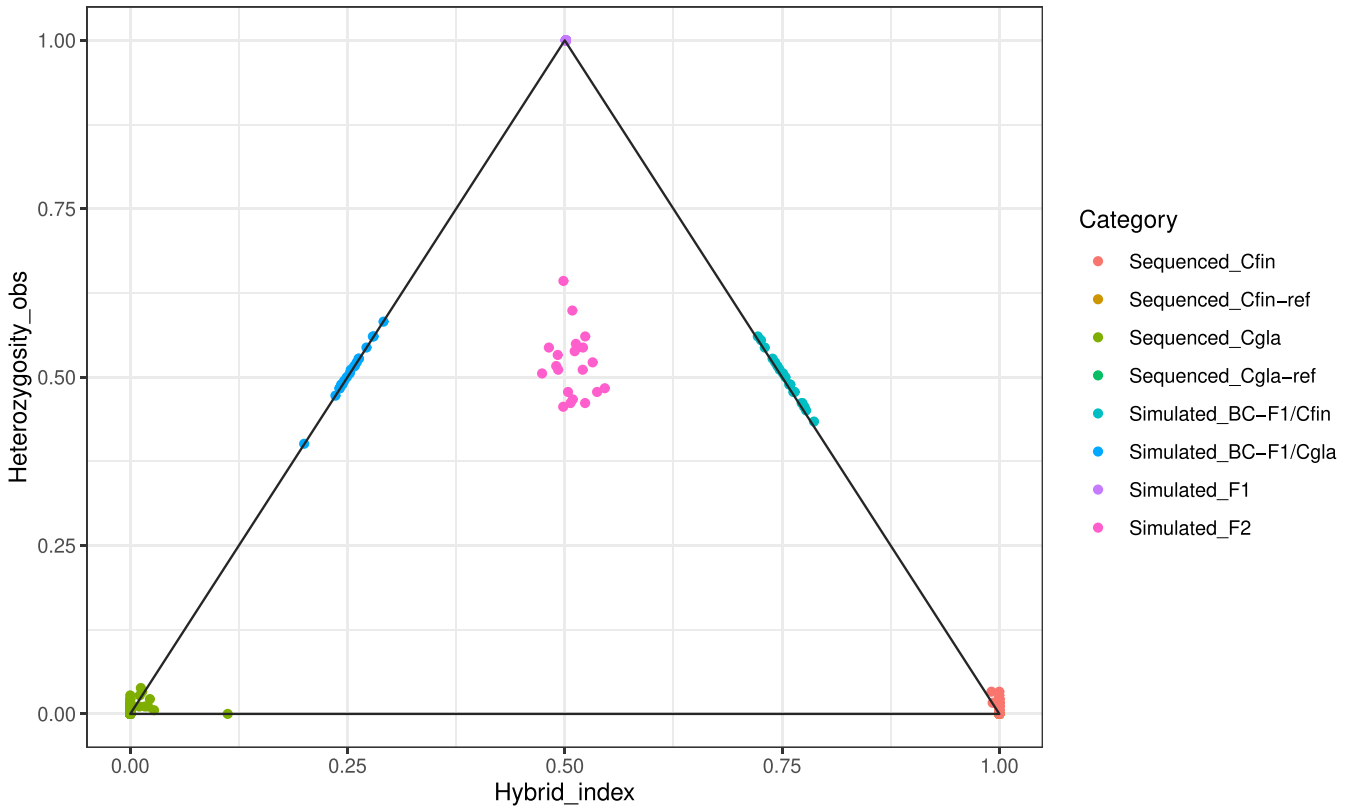

FIGURE 6 Triangle plot representing the level of genetic admixture ('Hybrid_index') and observed heterozygosity ('Heterozygosity_obs') per individual, including sequenced and simulated genotypes, at 182 loci diagnostic between *Calanus finmarchicus* and *C. glacialis*. Genotypes of individuals from allopatric areas were selected for each species and used as reference to simulate 100 hybrid genotypes with HYBRIDLAB. These included: 40 first-generation hybrids ('Simulated_F1'), 20 backcrosses between F1 and *C. finmarchicus* reference individuals ('Simulated_BC_F1/Cfin'), 20 backcrosses between F1 and *C. glacialis* reference individuals ('Simulated_BC_F1/Cgla') and 20 second generation hybrids ('Simulated_F2') generated by crossing F1 with F1. Simulated hybrid genotypes and sequenced genotypes, including reference individuals ('Sequenced_Cfin_ref'; 'Sequenced_Cgla_ref') plus the rest of the dataset ('Sequenced_Cfin'; 'Sequenced_Cgla') are represented.

genotypes were not biased by technical issues, these 10 microsatellite loci were unable to accurately distinguish between different hybrid classes simulated but also between parental species and hybrids. However, and perhaps more concerning were the potential genotyping artefacts for these loci, which we detected in the light of recently published transcriptomes (Table 4). The artefacts we identified in eight out of 10 *Calanus* microsatellite loci include size homoplasy, null alleles, paralogous loci, locus duplication and multiple reverse primer sites within the same locus (Figure 7). Therefore, previous conclusions related to hybridization in *Calanus* spp. based solely on these microsatellites must be dismissed. Such genotyping artefacts inherent to the nature of microsatellites are well documented (e.g. Chambers & MacAvoy, 2000; Dakin & Avise, 2004; Pompanon et al., 2005; Šarhanová et al., 2018; Selkoe & Toonen, 2006). Regrettably, although well known, these important limitations are often ignored and can lead to false reports of hybridization.

Erroneous reports of hybridization from microsatellites have been found in a wide range of organisms including newts (Miralles et al., 2023), lemurs (Poelstra et al., 2022), dingoes (Cairns et al., 2023), polecats (Szatmári et al., 2021) and Cape hakes (Henriques et al., 2016). These false reports of hybridization were mainly attributed to the lack of power resulting from the use of a (very) small number of hyper-variable loci. Unfortunately, only a few studies have investigated the effect of genotyping artefacts by either cloning and sequencing alleles or high-throughput sequencing (see, e.g. Germain-Aubrey et al., 2016). Size homoplasy appears nonetheless to be very common. In plants, for example, considering only the amplicon size can lead to the misidentification of ½ to 2/3 of the alleles (Šarhanová et al., 2018; Viruel et al., 2018). Similar levels of homoplasy were also found in fish (Shirai et al., 2009; Vartia et al., 2016) and in *Calanus* (4/9 loci, present study – Table 4). In their study of Cape hakes, Henriques et al. (2016) sequenced the flanking regions of the most divergent microsatellite markers and subsequently performed phylogenetic analyses, showing that the reported 'hybrids' were in fact the result of microsatellite homoplasy mimicking the effect of hybridization. As another example on the challenges of using microsatellites for studying hybridization, Hoffman and Amos (2005) reported the high error rate in their results due to the difficulty of interpreting microsatellite data, with situations where homozygous and heterozygous individuals were confounded by the reader.

TABLE 4 Results from the investigation of the 10 microsatellite loci from which hybrids between *Calanus finmarchicus* and *C. glacialis* were detected.

Microsatellite locus	Matching sequences accession number	Issues
FK868270	GAXK01014110.1 GJQS01464451.1	Primers target the same locus as EH666474; Size homoplasy (Figure 7a)
FG632811	GJQS01052751.1 GJQS01431533.1 GJQS01052751.1	Sequence variation included in the primer binding areas: potential for null alleles; Size homoplasy (Figure 7b)
FK670364	GAXK01066661.1 GJQS01261638.1 GJQS01058844.1 GJQS01396414.1	Both forward and reverse primers match two completely distinct transcripts with similar expected PCR product size (i.e. paralogues) (Figure 7c)
FK867682	GAXK01062764.1 GJQS01537761.1 GJQS01368927.1 GJQS01123245.1	Sequence variation in the primer binding areas: potential for null alleles; Size homoplasy (Figure 7d)
EL696609	–	None detected
EL585922	GAXK01187596.1 GJQS01521200.1 GJQS01223321.1 GJQS01027808.1	Multiple binding sites for the reverse primers within the same transcript. Multiple expected PCR products with 20 bp intervals (Figure 7e)
EH666870	GJQS01109021.1 GJQS01324418.1 GJQS01531887.1 GJQS01109021.1	Sequence variation included in the primer binding areas: potential for null alleles; Size homoplasy (Figure 7f)
EL773519	GAXK01097658.1 GJQS01424031.1 GJQS01345254.1 GJQS01050050.1	Multiple peaks on the trace file (Figure S4)
EH666474	GAXK01014110.1 GJQS01464451.1	Primers target the same locus as FK868270; Size homoplasy
EL773359	GAXK01109269.1	None detected

Note: Matching sequences with accession numbers starting with 'GA' correspond to *C. finmarchicus* transcript sequences, while the ones starting with 'GJ' correspond to *C. glacialis*.

Altogether, these studies, including ours, indicate that without proper assessment of power, homoplasy and the possibility of null alleles, extreme caution should be used with microsatellites when reporting hybridization.

4.2 | Scaling up the exploration of hybridization in *Calanus* spp. with SNPs

Genome-wide markers such as SNPs are powerful tools in the field of population genomics to assess species boundaries (Arias et al., 2016; Loureiro et al., 2020; Wagner et al., 2013). They allow much higher resolution compared to traditional molecular markers (such as microsatellites and InDels) to address questions related to hybridization (Harrison & Larson, 2014; Melville et al., 2017). However, generating a sufficient number of SNPs randomly spread across a genome may be a challenge in non-model organisms (da Fonseca et al., 2016; Davey et al., 2011; Helyar et al., 2011), especially when the genome is large and complex (Deagle et al., 2015). Thus, population genomics studies have long been prohibitive in *Calanus* species as the small amount of DNA available per individual, the large genome size ranging from 6.3 Gb in *C. finmarchicus* to 12.2 Gb in *C. glacialis* and *C. hyperboreus* (McLaren et al., 1988) and the potential complexity of the genome (e.g. repeats and duplications) limit the success of certain genome-reduced representation protocols such as RAD-seq (Choquet et al., 2019). In the present study, we made use of the latest techniques and data available for the genus *Calanus* to generate two powerful SNP datasets.

Our protocol of target-capture was successfully applied to all four species of *Calanus*. As expected, the success of target-capture (assessed from on-target mapping rates) decreased with increasing genetic distance from the reference species *C. finmarchicus* (Choquet et al., 2019), with *C. hyperboreus* (presumably the most genetically divergent species) having the lowest number of sequences captured, behind *C. helgolandicus*. This trend is typical of target-capture-based methods (Bartoš et al., 2023) and was also reported in Choo et al. (2020) using a similar protocol, with capture probes designed for the pteropod *Limacina bulimoides* and applied to congeneric species. We were able to yield 1535 high-quality SNPs across four *Calanus* species, and 6030 SNPs for the investigation focused

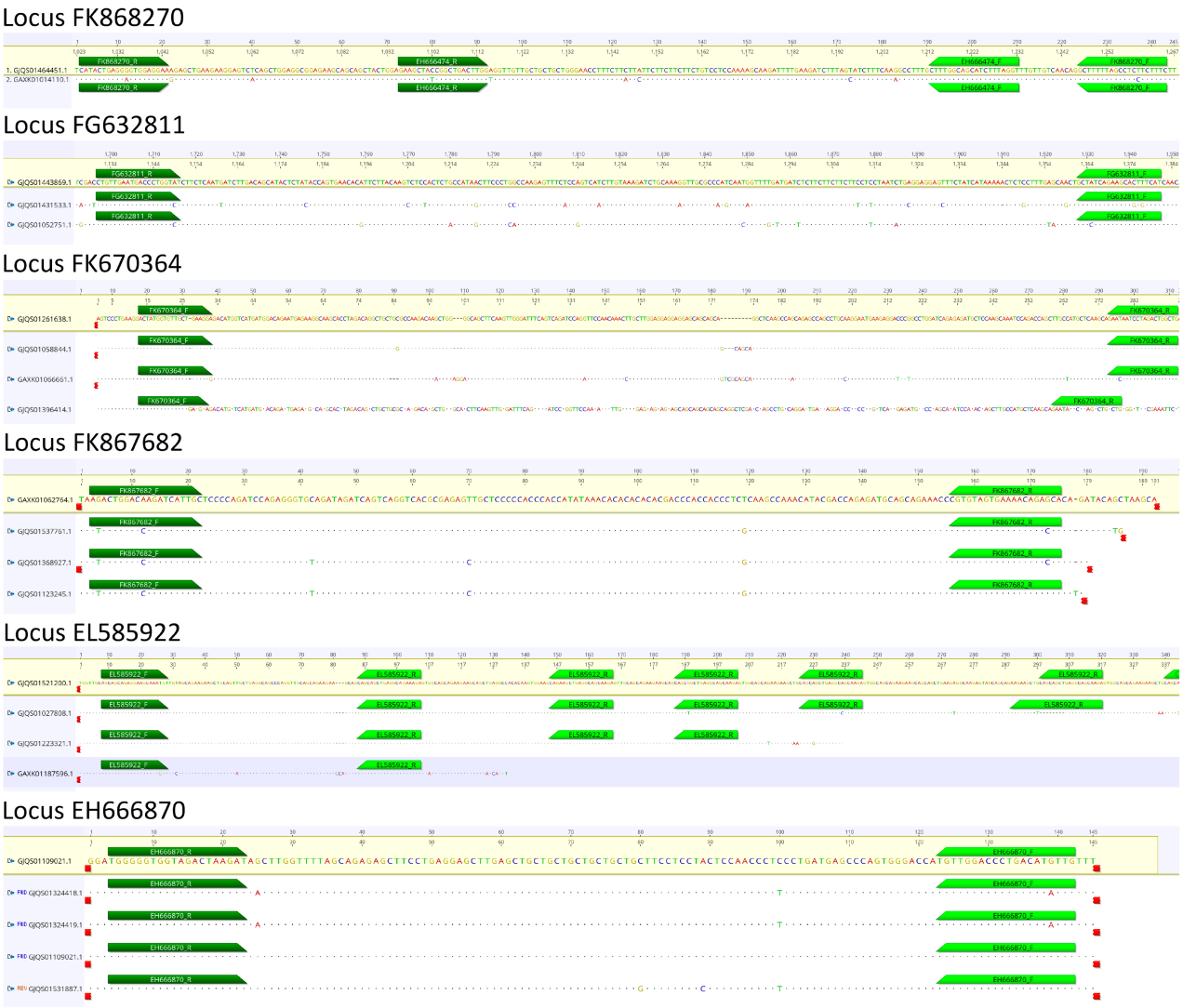

FIGURE 7 Overview of the different technical issues revealed by mappings of microsatellite primer pairs to transcripts of *Calanus finmarchicus* and *C. glacialis*. Primer pairs are displayed in green, dark green for the Forward primer sequences and light green for the Reverse primer sequences. Sequences starting with 'GA' correspond to *C. finmarchicus* transcripts, and with 'GJ' correspond to *C. glacialis* transcripts. (a) Same locus targeted by primer pairs for loci FK868270 and EH666474, plus size homoplasy; (b) size homoplasy and sequence variation in the primer binding area, potentially leading to null alleles, in locus FG632811; (c) paralogues identified for locus FK670364; (d) same problems as (b) for locus FK867682; (e) multiple binding sites for the reverse primer of locus EL585922 at 20 bp intervals, which may lead to multiple amplified fragments of different sizes; (f) same problems as (b) and (d).

on *C. finmarchicus* and *C. glacialis*. In complement to the target-capture approach, the recently released transcriptome data from *C. helgolandicus*, *C. finmarchicus*, *C. glacialis* and *C. hyperboreus* (Lizano et al., 2022) allowed us to yield a second independent dataset of 9199 SNPs.

4.3 | Zero hybridization detected: Harnessing the power of target-capture and RNA-seq

Analyses of the target-capture and RNA-seq SNP datasets revealed the lack of genetic introgression (i.e. incorporation of genetic material from one species in the genome of another) among the four species, including *C. finmarchicus* and *C. glacialis*. Species boundaries analyses performed on both SNP datasets are very clear in clustering *C. hyperboreus* and *C. helgolandicus* well apart from the two sister species *C. finmarchicus* and *C. glacialis*, which appear closer to each other, as expected considering the recently published phylogeny of the genus *Calanus* based on whole transcriptomes (Lizano et al., 2022). Regardless of whether they came from areas of sympatry or allopatry, individuals clustered per species in very similar patterns in both PCAs with greater distinction obtained from the RNA-seq-based analysis. Simulations revealed that the few 'noisy' individuals identified in our target-capture SNP results were not real introgressed individuals as their heterozygosity level did not match such hypothesis, but better explained by missing data linked to the nature of the target-capture protocol. Likewise, the very low level of admixture (<3%) detected between *C. glacialis* and

C. helgolandicus was an indication of 'noise' in the target-capture SNP dataset and reflected the higher levels of missing data and overall lower sequencing depth compared to the RNA-seq-based dataset (Supplementary Material 2 in Data S1), as the *C. glacialis* individuals showing these signs of admixture (from the Labrador Sea and Isfjord, Svalbard) do not occur anywhere near *C. helgolandicus* distribution range. The higher number of SNPs yielded from RNA-seq data, as well as overall higher sequencing depth and lesser proportion of missing data (Supplementary Material 2 in Data S1) are expected to provide higher resolution in clustering analyses, which we clearly observed here.

Our phylogenetic analysis supports a clear speciation history among *C. finmarchicus*, *C. glacialis* and *C. hyperboreus* without any sign of secondary gene flow between *C. finmarchicus* and *C. glacialis*. All trees calculated for that analysis support the same scenario of a single split between *C. finmarchicus* and *C. glacialis*, indicating that signatures of divergence are homogeneously distributed across the sampled regions of the genomes. Our second target-capture dataset focusing with more power on *C. finmarchicus* and *C. glacialis* showed a clearer picture of the separation between the two species. Interestingly, we detected a greater intra-specific genetic diversity in *C. glacialis* compared to *C. finmarchicus*, with a higher number of SNPs detected despite a slightly lower mapping rate. These results are in line with earlier findings based on microsatellites and SNPs suggesting higher levels of genetic diversity in *C. glacialis* compared to *C. finmarchicus* (Choquet et al., 2017, 2019). It is noteworthy that the variant datasets used in our analyses are not suitable for investigating patterns of population structure within species as SNPs were called on several species together at once and thus only the loci conserved across species were retained.

Simulation analyses revealed the power of our target-capture SNPs to identify putative hybrids between *C. finmarchicus* and *C. glacialis*. While we can confidently reject the presence of hybrids in our samples, we can never exclude to have missed rare hybridization that would require much more samples to be detected. However, Parent et al. (2012) estimated an average of 21% of hybrids in the East-Canadian Arctic region. If we extrapolate this rate to our data where we analysed 96 specimens from sympatric areas between *C. finmarchicus* and *C. glacialis*, we would expect to find at least 20 hybrids. Thus, it is unlikely that we would have missed hybrids if they would occur in similar proportions as suggested by Parent et al. (2012). Furthermore, our previous extensive spatio-temporal investigations using InDels, which should at the very least be able to detect F1 hybrids (as endorsed by Parent et al., 2021), of nearly 6000 specimens including nauplii, never revealed any hybrid (Choquet et al., 2017, 2020).

This study represents the first attempt to use a protocol of target-capture to explore putative introgression in non-model organisms (Bartoš et al., 2023). Beyond simulations, the power of our target-capture SNPs was further demonstrated by our parallel analyses of an independent SNP dataset generated from RNA-seq, which displayed the exact same patterns of clustering and differentiation among species, only with better resolution, which is expected from a higher number of markers. The high resolution achieved using transcriptomic SNPs in our analyses also highlights the power of transcriptomic datasets to solve evolutionary questions as proposed by Lenz et al. (2021). The value of using the transcriptome as an alternative when genomes are too challenging to sequence is particularly well illustrated by Choquet et al. (2023), where transcriptomes were used to explore genetic variation and local adaptation across 20 species of krill.

4.4 | Sympatric and morphologically similar, but reproductively isolated

Interestingly, the four species of *Calanus* analysed here are morphologically very similar (Choquet et al., 2018; Conover, 1988; Gabrielsen et al., 2012; Nielsen et al., 2014), share multiple life-history traits (Conover, 1988; Falk-Petersen et al., 2009) and often occur in sympatry (Arnkværn et al., 2005; Bucklin et al., 2000; Choquet et al., 2017; Lindeque et al., 2004; Schultz et al., 2023) where they may even mate in synchronicity (Choquet et al., 2020). Nonetheless, they seem to have remained genetically isolated after long periods of contact.

In fact, there has not been any report of natural inter-specific hybridization in marine pelagic zooplankton (to our knowledge) despite numerous documented cases of sympatry among morphologically alike or cryptic species. The most likely explanation for this apparent lack of hybridization in marine zooplankton is that very few studies have applied powerful molecular methods to investigate this question compared to, for instance, insects, often seen as the terrestrial counterparts of marine copepods (Schminke, 2007) and for which hybridization reports are on the rise due to climate change (Sánchez-Guillén et al., 2016). Nonetheless, among the scarce studies on marine zooplankton where genome-wide datasets were used, instances of reproductive isolation despite a lack of clear morphological differentiation have been reported. For example, in the Indian Ocean, the cryptic krill *Euphausia similis* and its variant form *armata* occur together in sympatry and are so morphologically alike that their taxonomic status has remained unresolved for long, but a recent assessment with genome-wide data did not detect any sign of recent gene flow between the two, and sufficient evidence was revealed to elevate *E. armata* as a distinct species (Choquet et al., 2023). In pteropods, a target-capture approach recently revealed the existence of three genetically distinct lineages within *Limacina bulimoides*, originally considered as a single circumglobal species. Among these, two genetically distinct lineages, morphologically undistinguishable live together in sympatry in the North Pacific Ocean without hybridizing (Choo et al., 2023).

Our understanding of which factors play a role in keeping morphologically alike zooplankton reproductively isolated despite living in sympatry is limited by a severe lack of data. Only a limited number of studies have explored the mechanisms underlying reproductive isolation in zooplankton. In the estuarine

copepod *Acartia tonsa*, observations suggested that pre-zygotic mechanisms, occurring pre- or during mating, may be responsible for reproductive isolation between cryptic lineages (Plough et al., 2018). The authors recommended performing further analyses of mating behaviour to advance the understanding of this pre-zygotic isolation mechanism. In the hybridizing freshwater copepods *Daphnia galeata* and *D. hyalina*, evidence of a reduced sexual fitness of hybrids was inferred to be the main post-zygotic determinant of reproductive isolation between the two parental species (Keller et al., 2007). For the genus *Calanus*, potential mechanisms involved in maintaining reproductive isolation between species have been discussed in Choquet et al. (2020) but data are missing to determine which pre-zygotic or post-zygotic processes are at work to keep species boundaries non-permeable. However, the recent phylogeny of the genus *Calanus* based on whole transcriptome analyses (Lizano et al., 2022), which displays the same topology as an earlier morphology-based phylogeny reported in Bucklin et al. (1995), (based on Frost, 1974 and unpubl. work of A. Fleminger) brings a new light to this matter. The similarity between the two phylogenies (morphology-based and transcriptome-based) suggests that some of the morphological characters selected may reflect the evolutionary history of the genus, including variations in the structure of secondary sexual organs in *Calanus* spp., supposedly playing a role of barrier to inter-species copulation (Bucklin et al., 1995; Fleminger & Hulsemann, 1977). Further investigations of reproductive isolation mechanisms, including but not limited to morphological characters, are needed to shed light on the history of speciation in the genus *Calanus* and should be based on genetically identified specimens.

Finally, as the extent of sympatry among *Calanus* species is expected to increase in the future due to climate change-induced shifts (e.g. Freer et al., 2022), especially in the Arctic, the absence of hybridization mechanisms raises a critical question. Will *Calanus* species coexist in new areas of sympatry by occupying distinct ecological niches, or will they be competing until one is excluded? Recent studies suggest that the contrasting life-history traits of *C. finmarchicus* and *C. glacialis* maintain their current cohabitation in Arctic fjords where they co-occur, with differences in reproductive timing reducing the competition (Hatlebakk et al., 2022). However, as the Arctic continues to change rapidly, further monitoring will be necessary to keep track of the emerging ecological dynamics.

5 | CONCLUSIONS

We have identified a variety of technical issues associated with microsatellites that led to the initial claim of hybridization between *Calanus* species. Furthermore, our comprehensive analysis of two independent powerful genome-wide datasets revealed no evidence of hybridization among four *Calanus* species, resolving a long-standing debate. Our results confirm the validity of the nuclear InDel markers previously used, and in the absence of hybridization suggested by our data, species identification based on mitochondrial DNA can be safely resumed in *Calanus* (e.g. with 16S as described in Lindeque et al., 1999). Furthermore, the absence of genetic admixture among *Calanus* species in our data provides a solid foundation for conducting physiological and ecological studies, especially relevant in light of climate changes currently affecting their distributions (Freer et al., 2022; Tarling et al., 2022). Our findings strongly advocate for caution when relying solely on microsatellites and we therefore recommend the integration of genome-wide data to reassess previous claims and thoroughly investigate the possibility of interspecific hybridization. This is particularly crucial for key species such as *Calanus* spp., which serve as indicators of climate change.

AUTHOR CONTRIBUTIONS

GH and MC designed the study. MC was involved in sampling. MC and AKSD carried out the molecular laboratory work. ALM and MR provided technical guidance for data analyses. MC, AML and ALM performed the bioinformatic analyses with SNPs. MC and GH analysed the microsatellite data and wrote the first draft. All authors thoroughly discussed the manuscript, contributed significantly to the writing and approved the submitted version.

ACKNOWLEDGEMENTS

This study was supported by Nord University (Bodø, Norway). We are very grateful to all the people who provided us with samples, including Ann Bucklin, Kim Last, Seona Wells, Irina Smolina, Ketil Eiane, Torbjørn Dale, Ksenia Kosobokova, Morten Krogstad, Gauthier Burckard, Ekaterina Abramova, Maja Hatlebakk and Janne Søreide. We warmly thank Kanchana Bandara for his help with the sampling map.

CONFLICT OF INTEREST STATEMENT

The authors of this manuscript declare having no conflict of interest to report.

OPEN RESEARCH BADGES

This article has earned Open Data, Open Materials and Preregistered Research Design badges. Data, materials and the preregistered design and analysis plan are available at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA996530/.

DATA AVAILABILITY STATEMENT

Raw sequencing data from the target-capture have been submitted to NCBI and are available under the BioProject ID: PRJNA996530. Individual BioSample accessions are listed in Supplementary Material 1 in Data S1 (Table S1).

ORCID

Marvin Choquet https://orcid.org/0000-0001-6719-2332
Alan Le Moan https://orcid.org/0000-0002-9124-6844
Mark Ravinet https://orcid.org/0000-0002-2841-1798

REFERENCES

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Arias, C. F., Van Belleghem, S., & McMillan, W. O. (2016). Genomics at the evolving species boundary. *Current Opinion in Insect Science*, *13*, 7–15.

Arnkværn, G., Daase, M., & Eiane, K. (2005). Dynamics of coexisting *Calanus finmarchicus*, *Calanus glacialis* and *Calanus hyperboreus* populations in a high-Arctic fjord. *Polar Biology*, *28*(7), 528–538. https://doi.org/10.1007/s00300-005-0715-8

Bartoš, O., Bohlen, J., Šlechtová, V. B., Kočí, J., Röslein, J., & Janko, K. (2023). Sequence capture: Obsolete or irreplaceable? A thorough validation across phylogenetic distances and its applicability to hybrids and allopolyploids. *Molecular Ecology Resources*, *23*(6), 1348–1360.

Beaugrand, G., Brander, K. M., Lindley, J. A., Souissi, S., & Reid, P. C. (2003). Plankton effect on cod recruitment in the North Sea. *Nature*, *426*(6967), 661–664.

Beaugrand, G., Reid, P. C., Ibanez, F., Lindley, J. A., & Edwards, M. (2002). Reorganization of North Atlantic marine copepod biodiversity and climate. *Science*, *296*(5573), 1692–1694.

Bonnet, D., Richardson, A., Harris, R., Hirst, A., Beaugrand, G., Edwards, M., Ceballos, S., Diekman, R., López-Urrutia, A., Carlotti, F., & Valdes, L. (2005). An overview of *Calanus helgolandicus* ecology in European waters. *Progress in Oceanography*, *65*(1), 1–53.

Bouckaert, R. R., & Heled, J. (2014). DensiTree 2: Seeing trees through the forest. *BioRxiv*, 012401. https://doi.org/10.1101/012401

Broad Institute. (2019). *Picard toolkit*. https://broadinstitute.github.io/picard/

Broms, C., Melle, W., & Kaartvedt, S. (2009). Oceanic distribution and life cycle of *Calanus* species in the Norwegian Sea and adjacent waters. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(21–22), 1910–1921. https://doi.org/10.1016/j.dsr2.2008.11.005

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N. A., & RoyChoudhury, A. (2012). Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution*, *29*(8), 1917–1932.

Bucklin, A., Frost, B., & Kocher, T. (1995). Molecular systematics of six *Calanus* and three *Metridia* species (Calanoida: Copepoda). *Marine Biology*, *121*(4), 655–664.

Bucklin, A., Kaartvedt, S., Guarnieri, M., & Goswami, U. (2000). Population genetics of drifting (*Calanus* spp.) and resident (*Acartia clausi*) plankton in Norwegian fjords. *Journal of Plankton Research*, *22*(7), 1237–1251.

Cairns, K. M., Crowther, M. S., Parker, H. G., Ostrander, E. A., & Letnic, M. (2023). Genome-wide variant analyses reveal new patterns of admixture and population structure in Australian dingoes. *Molecular Ecology.*, *32*, 4133–4150.

Canestrelli, D., Bisconti, R., Chiocchio, A., Maiorano, L., Zampiglia, M., & Nascetti, G. (2017). Climate change promotes hybridisation between deeply divergent species. *PeerJ*, *5*, e3072.

Chambers, G. K., & MacAvoy, E. S. (2000). Microsatellites: Consensus and controversy. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *126*(4), 455–476.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7.

Choo, L., Spagliardi, G., Malinsky, M., Choquet, M., Goetze, E., Hoarau, G., & Peijnenburg, K. (2023). Genome-wide phylogeography reveals cryptic speciation in the circumglobal planktonic calcifier *Limacina bulimoides*. *Molecular Ecology.*, *32*, 3200–3219.

Choo, L. Q., Bal, T. M., Choquet, M., Smolina, I., Ramos-Silva, P., Marlétaz, F., Kopp, M., Hoarau, G., & Peijnenburg, K. T. (2020). Novel genomic resources for shelled pteropods: A draft genome and target capture probes for *Limacina bulimoides*, tested for cross-species relevance. *BMC Genomics*, *21*(1), 1–14.

Choquet, M., Burckard, G., Skreslet, S., Hoarau, G., & Søreide, J. E. (2020). No evidence for hybridization between *Calanus finmarchicus* and *Calanus glacialis* in a subarctic area of sympatry. *Limnology and Oceanography*, *66*, S314–S325.

Choquet, M., Hatlebakk, M., Dhanasiri, A. K., Kosobokova, K., Smolina, I., Søreide, J. E., Svensen, C., Melle, W., Kwaśniewski, S., Daase, M., Tverberg, V., Skreslet, S., Bucklin, A., Hoarau, G., & Eiane, K. (2017). Genetics redraws pelagic biogeography of *Calanus*. *Biology Letters*, *13*(12), 20170588.

Choquet, M., Kosobokova, K., Kwaśniewski, S., Hatlebakk, M., Dhanasiri, A. K., Melle, W., Svensen, C., Melle, W., Kwaśniewski, S., Eiane, K., Daase, M., Tverberg, V., Skreslet, S., Bucklin, A., & Hoarau, G. (2018). Can morphology reliably distinguish between the copepods *Calanus finmarchicus* and *C. glacialis*, or is DNA the only way? *Limnology and Oceanography: Methods*, *16*(4), 237–252.

Choquet, M., Lenner, F., Cocco, A., Toullec, G., Corre, E., Toullec, J.-Y., & Wallberg, A. (2023). Comparative population genomics provide new insight into the evolutionary history and adaptive potential of World Ocean krill. *Molecular Biology and Evolution*, msad225. https://doi.org/10.1093/molbev/msad225

Choquet, M., Smolina, I., Dhanasiri, A. K., Blanco-Bercial, L., Kopp, M., Jueterbock, A., Sundaram, A. Y. M., & Hoarau, G. (2019). Towards population genomics in non-model species with large genomes: A case study of the marine zooplankton *Calanus finmarchicus*. *Royal Society Open Science*, *6*(2), 180608.

Choquet, M., Smolina, I., Hoarau, G., & Hoarau, G. (2021). No evidence for hybridization in *Calanus*: Reply to the comment by Parent et al. *Limnology and Oceanography*, *66*(10), 3603–3606.

Chunco, A. J. (2014). Hybridization in a warmer world. *Ecology and Evolution*, *4*(10), 2019–2031.

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*, *71*(2), 241–253. https://doi.org/10.1093/icesjms/fst147

Conover, R. J. (1988). Comparative life histories in the genera *Calanus* and *Neocalanus* in high latitudes of the northern hemisphere. *Hydrobiologia*, *167*, 127–142. https://doi.org/10.1007/BF00026299

Coyer, J., Hoarau, G., Stam, W., & Olsen, J. (2007). Hybridization and introgression in a mixed population of the intertidal seaweeds *Fucus evanescens* and *F. Serratus*. *Journal of Evolutionary Biology*, *20*(6), 2322–2333.

Coyne, J. A., & Allen Orr, H. (1998). The evolutionary genetics of speciation. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *353*(1366), 287–305.

da Fonseca, R. R., Albrechtsen, A., Themudo, G. E., Ramos-Madrigal, J., Sibbesen, J. A., Maretty, L., Zepeda-Mendoza, M. L., Campos, P. F., Heller, R., & Pereira, R. J. (2016). Next-generation biology: Sequencing and data analysis approaches for non-model organisms. *Marine Genomics*, *30*, 3–13.

Dakin, E., & Avise, J. (2004). Microsatellite null alleles in parentage analysis. *Heredity*, *93*(5), 504–509.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., & McVean, G. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics*, *12*(7), 499–510.

Deagle, B. E., Faux, C., Kawaguchi, S., Meyer, B., & Jarman, S. N. (2015). Antarctic krill population genomics: Apparent panmixia, but genome complexity and large population size muddy the water. *Molecular Ecology*, *24*(19), 4943–4959.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Falk-Petersen, S., Mayzaud, P., Kattner, G., & Sargent, J. R. (2009). Lipids and life strategy of Arctic *Calanus*. *Marine Biology Research*, *5*(1), 18–39.

Fleminger, A., & Hulsemann, K. (1977). Geographical range and taxonomic divergence in North Atlantic *Calanus* (*C. helgolandicus*, *C. finmarchicus* and *C. glacialis*). *Marine Biology*, *40*(3), 233–248.

Fraïsse, C., Le Moan, A., Roux, C., Dubois, G., Daguin-Thiebaut, C., Gagnaire, P.-A., Viard, F., & Bierne, N. (2022). Introgression between highly divergent sea squirt genomes: An adaptive breakthrough? *Peer Community Journal*, *2*, e54. https://doi.org/10.24072/pcjournal.172

Freer, J. J., Daase, M., & Tarling, G. A. (2022). Modelling the biogeographic boundary shift of *Calanus finmarchicus* reveals drivers of Arctic Atlantification by subarctic zooplankton. *Global Change Biology*, *28*(2), 429–440.

Frost, B. (1974). *Calanus marshallae*, a new species of calanoid copepod closely allied to the sibling species *C. finmarchicus* and *C. glacialis*. *Marine Biology*, *26*(1), 77–99.

Gabrielsen, T. M., Merkel, B., Søreide, J. E., Johansson-Karlsson, E., Bailey, A., Vogedes, D., Nygård, H., Varpe, Ø., & Berge, J. (2012). Potential misidentifications of two climate indicator species of the marine arctic ecosystem: *Calanus glacialis* and *C. finmarchicus*. *Polar Biology*, *35*(11), 1621–1628. https://doi.org/10.1007/s00300-012-1202-7

Germain-Aubrey, C. C., Nelson, C., Soltis, D. E., Soltis, P. S., & Gitzendanner, M. A. (2016). Are microsatellite fragment lengths useful for population-level studies? The case of *Polygala lewtonii* (Polygalaceae). *Applications in Plant Sciences*, *4*(2), 1500115.

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*(S1), 795–809. https://doi.org/10.1093/jhered/esu033

Hatlebakk, M., Kosobokova, K., Daase, M., & Søreide, J. E. (2022). Contrasting life traits of sympatric *Calanus glacialis* and *C. finmarchicus* in a warming Arctic revealed by a year-round study in Isfjorden, Svalbard. *Frontiers in Marine Science*, *9*, 877910.

Hays, G. C., Richardson, A. J., & Robinson, C. (2005). Climate change and marine plankton. *Trends in Ecology & Evolution*, *20*(6), 337–344.

Helaouët, P., & Beaugrand, G. (2007). Macroecology of *Calanus finmarchicus* and *C. helgolandicus* in the North Atlantic Ocean and adjacent seas. *Marine Ecology Progress Series*, *345*, 147–165.

Helyar, S. J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M., Ogden, R., Limborg, M., Cariani, A., Maes, G. E., Diopere, E., Nielsen, E. E., & Carvalho, G. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, *11*(s1), 123–136.

Henriques, R., Von der Heyden, S., & Matthee, C. A. (2016). When homoplasy mimics hybridization: A case study of cape hakes (*Merluccius capensis* and *M. Paradoxus*). *PeerJ*, *4*, e1827.

Hoarau, G., Coyer, J. A., Giesbers, M., Jueterbock, A., & Olsen, J. L. (2015). Pre-zygotic isolation in the macroalgal genus *Fucus* from four contact zones spanning 100–10 000 years: A tale of reinforcement? *Royal Society Open Science*, *2*(2), 140538.

Hoffman, J., & Amos, W. (2005). Microsatellite genotyping errors: Detection approaches, common sources and consequences for paternal exclusion. *Molecular Ecology*, *14*(2), 599–612.

Jaschnov, W. (1972). On the systematic status of *Calanus glacialis*, *Calanus finmarchicus* and *Calanus helgolandicus*. *Crustaceana*, *22*(3), 279–284.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., & Madden, T. L. (2008). NCBI BLAST: A better web interface. *Nucleic Acids Research*, *36*(suppl_2), W5–W9.

Keller, B., Wolinska, J., Tellenbach, C., & Spaak, P. (2007). Reproductive isolation keeps hybridizing *daphnia* species distinct. *Limnology and Oceanography*, *52*(3), 984–991.

Kõressaar, T., Lepamets, M., Kaplinski, L., Raime, K., Andreson, R., & Remm, M. (2018). Primer3_masker: Integrating masking of template sequence with primer design software. *Bioinformatics*, *34*(11), 1937–1938.

Kosobokova, K. N., Hopcroft, R. R., & Hirche, H.-J. (2011). Patterns of zooplankton diversity through the depths of the Arctic's central basins. *Marine Biodiversity*, *41*(1), 29–50.

Larson, E. L., Tinghitella, R. M., & Taylor, S. A. (2019). Insect hybridization and climate change. *Frontiers in Ecology and Evolution*, *7*, 348.

Le Moan, A., Roby, C., Fraisse, C., Daguin-Thiébaut, C., Bierne, N., & Viard, F. (2021). An introgression breakthrough left by an anthropogenic contact between two ascidians. *Molecular Ecology*, *30*(24), 6718–6732.

Lenz, P. H., Lieberman, B., Cieslak, M. C., Roncalli, V., & Hartline, D. K. (2021). Transcriptomics and metatranscriptomics in zooplankton: Wave of the future? *Journal of Plankton Research*, *43*(1), 3–9.

Lenz, P. H., Roncalli, V., Hassett, R. P., Wu, L.-S., Cieslak, M. C., Hartline, D. K., & Christie, A. E. (2014). *De novo* assembly of a transcriptome for *Calanus finmarchicus* (Crustacea, Copepoda)–the dominant zooplankter of the North Atlantic Ocean. *PLoS One*, *9*(2), e88589.

Li, H. (2013). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. arXiv preprint arXiv:1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639–647.

Lindeque, P., Harris, R., Jones, M., & Smerdon, G. (1999). Simple molecular method to distinguish the identity of *Calanus* species (Copepoda: Calanoida) at any developmental stage. *Marine Biology*, *133*(1), 91–96.

Lindeque, P. K., Harris, R. P., Jones, M. B., & Smerdon, G. R. (2004). Distribution of *Calanus* spp. as determined using a genetic identification system. *Scientia Marina*, *68*(S1), 121–128.

Lindeque, P. K., Hay, S. J., Heath, M. R., Ingvarsdottir, A., Rasmussen, J., Smerdon, G. R., & Waniek, J. J. (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*(2), 221–238. https://doi.org/10.1093/plankt/fbi115

Lizano, A. M., Smolina, I., Choquet, M., Kopp, M., & Hoarau, G. (2022). Insights into the species evolution of *Calanus* copepods in the northern seas revealed by de novo transcriptome sequencing. *Ecology and Evolution*, *12*(2), e8606.

Loureiro, L. O., Engstrom, M. D., & Lim, B. K. (2020). Single nucleotide polymorphisms (SNPs) provide unprecedented resolution of species boundaries, phylogenetic relationships, and genetic diversity in the mastiff bats (molossus). *Molecular Phylogenetics and Evolution*, *143*, 106690.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17*(1), 10–12.

McLaren, I. A., Sevigny, J.-M., & Corkett, C. J. (1988). Body sizes, development rates, and genome sizes among *Calanus* species. In G. A. Boxshall & H. K. Schminke (Eds.), *Biology of Copepods. Developments*

in Hydrobiology (Vol. *47*, pp. 275–284). Springer. https://doi.org/10.1007/978-94-009-3103-9_27

Melville, J., Haines, M. L., Boysen, K., Hodkinson, L., Kilian, A., Smith Date, K. L., Potvin, D. A., & Parris, K. M. (2017). Identifying hybridization and admixture using SNPs: Application of the DArTseq platform in phylogeographic research on vertebrates. *Royal Society Open Science*, *4*(7), 161061.

Miralles, A., Secondi, J., Pabijan, M., Babik, W., Lemaire, C., & Crochet, P.-A. (2023). Inconsistent estimates of hybridization frequency in newts revealed by SNPs and microsatellites. *Conservation Genetics*, 1–11. https://doi.org/10.1007/s10592-023-01556-9

Nielsen, E. E., Bach, L. A., & Kotlicki, P. (2006). HYBRIDLAB (version 1.0): A program for generating simulated hybrids from population samples. *Molecular Ecology Notes*, *6*(4), 971–973.

Nielsen, T. G., Kjellerup, S., Smolina, I., Hoarau, G., & Lindeque, P. (2014). Live discrimination of *Calanus glacialis* and *C. finmarchicus* females: Can we trust phenological differences? *Marine Biology*, *161*(6), 1299–1306. https://doi.org/10.1007/s00227-014-2419-5

Parejo, M., Henriques, D., Pinto, M. A., Soland-Reckeweg, G., & Neuditschko, M. (2018). Empirical comparison of microsatellite and SNP markers to estimate introgression in *Apis mellifera mellifera*. *Journal of Apicultural Research*, *57*(4), 504–506.

Parent, G. J. (2021). *Comment on hybridization in Calanus spp*. https://github.com/gjparent/2021-LO_Calanus_Comment

Parent, G. J., Plourde, S., Browman, H. I., Turgeon, J., & Petrusek, A. (2021). Defining what constitutes a reliable dataset to test for hybridization and introgression in marine zooplankton: Comment on Choquet et al. 2020 "No evidence for hybridization between *Calanus finmarchicus* and *C. glacialis* in a subarctic area of sympatry". *Limnology and Oceanography*, *66*(10), 3597–3602.

Parent, G. J., Plourde, S., Joly, P., & Turgeon, J. (2015). Phenology and fitness of *Calanus glacialis*, *C. finmarchicus* (Copepoda), and their hybrids in the St. Lawrence Estuary. *Marine Ecology Progress Series*, *524*, 1–9. https://doi.org/10.3354/meps11240

Parent, G. J., 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*(11), 1654–1665. https://doi.org/10.1093/plankt/fbr072

Parent, G. J., Plourde, S., & Turgeon, J. (2012). Natural hybridization between *Calanus finmarchicus* and *C. glacialis* (Copepoda) in the Arctic and Northwest Atlantic. *Limnology and Oceanography*, *57*(4), 1057–1066. https://doi.org/10.4319/lo.2012.57.4.1057

Plough, L., Fitzgerald, C., Plummer, A., & Pierson, J. (2018). Reproductive isolation and morphological divergence between cryptic lineages of the copepod *Acartia tonsa* in Chesapeake Bay. *Marine Ecology Progress Series*, *597*, 99–113.

Poelstra, J. W., Montero, B. K., Lüdemann, J., Yang, Z., Rakotondranary, S. J., Hohenlohe, P., Stetter, N., Ganzhorn, J. U., & Yoder, A. D. (2022). RADseq data reveal a lack of admixture in a mouse lemur contact zone contrary to previous microsatellite results. *Proceedings of the Royal Society B Biological Sciences*, *289*(1980), 20220596.

Poloczanska, E. S., Burrows, M. T., Brown, C. J., Garcia Molinos, J., Halpern, B. S., Hoegh-Guldberg, O., Kappel, C. V., Moore, P. J., Richardson, A. J., Sydeman, W. J., & Schoeman, D. S. (2016). Responses of marine organisms to climate change across oceans. *Frontiers in Marine Science*, *3*, 62.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: Causes, consequences and solutions. *Nature Reviews Genetics*, *6*(11), 847–859.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

Provan, J., Beatty, G. E., Maggs, C. A., & Savidge, G. (2007). Expressed sequence tag-derived microsatellites for the cool-water marine copepod *Calanus finmarchicus*. *Molecular Ecology Notes*, *7*(6), 1369–1371.

Putman, A. I., & Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution*, *4*(22), 4399–4428.

Rees, C. (1957). *Continuous plankton records: The distribution of Calanus finmarchicus (Gunn.) in the North Sea and north-eastern Atlantic, 1946-1953*.

Reid, P. C., Edwards, M., Beaugrand, G., Skogen, M., & Stevens, D. (2003). Periodic changes in the zooplankton of the North Sea during the twentieth century linked to oceanic inflow. *Fisheries Oceanography*, *12*(4–5), 260–269.

Sánchez-Guillén, R. A., Córdoba-Aguilar, A., Hansson, B., Ott, J., & Wellenreuther, M. (2016). Evolutionary consequences of climate-induced range shifts in insects. *Biological Reviews*, *91*(4), 1050–1064.

Šarhanová, P., Pfanzelt, S., Brandt, R., Himmelbach, A., & Blattner, F. R. (2018). SSR-seq: Genotyping of microsatellites using next-generation sequencing reveals higher level of polymorphism as compared to traditional fragment size scoring. *Ecology and Evolution*, *8*(22), 10817–10833.

Schliep, K. P. (2011). Phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593.

Schminke, H. K. (2007). Entomology for the copepodologist. *Journal of Plankton Research*, *29*, I149–I162. https://doi.org/10.1093/plankt/fbl073

Schultz, M., Choquet, M., Tverberg, V., & Hoarau, G. (2023). *Calanus helgolandicus* - more than a guest in the north? *Journal of Plankton Research*, *45*(1), 33–36.

Sedlazeck, F. J., Rescheneder, P., & Von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*, *29*(21), 2790–2791.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. *Ecology Letters*, *9*(5), 615–629.

Shirai, Y., Ikeda, S., & Tajima, S. (2009). Size homoplasy at microsatellite loci in two rose bittering subspecies, *Rhodeus ocellatus kurumeus* and *R. o. ocellatus*. *Japanese Journal of Ichthyology*, *56*, 165–169.

Smolina, I., Kollias, S., Poortvliet, M., Nielsen, T. G., Lindeque, P., Castellani, C., Møller, E. F., Blanco-Bercial, L., & Hoarau, G. (2014). Genome- and transcriptome-assisted development of nuclear insertion/deletion markers for *Calanus* species (Copepoda: Calanoida) identification. *Molecular Ecology Resources*, *14*(5), 1072–1079. https://doi.org/10.1111/1755-0998.12241

Stankowski, S., Zagrodzka, Z. B., Garlovsky, M. D., Pal, A., Shipilina, D., Castillo, D. G., Le Moan, A., Leder, E., Reeve, J., Johannesson, K., Westram, A. M., & Johannesson, K. (2023). Selection on many loci drove the origin and spread of a key innovation. *BioRxiv*, 2023.02.13.528213. https://doi.org/10.1101/2023.02.13.528213

Szatmári, L., Cserkész, T., Laczkó, L., Lanszki, J., Pertoldi, C., Abramov, A. V., Elmeros, M., Ottlecz, B., Hegyeli, Z., & Sramkó, G. (2021). A comparison of microsatellites and genome-wide SNPs for the detection of admixture brings the first molecular evidence for hybridization between *Mustela eversmanii* and *M. putorius* (Mustelidae, Carnivora). *Evolutionary Applications*, *14*(9), 2286–2304.

Tarling, G. A., Freer, J. J., Banas, N. S., Belcher, A., Blackwell, M., Castellani, C., Cook, K. B., Cottier, F. R., Daase, M., Last, K. S., Lindeque, P. K., Mayor, D. J., Mitchell, E., Parry, H. E., Speirs, D. C., Stowasser, G., Wootton, M., & Johnson, M. L. (2022). Can a key boreal *Calanus* copepod species now complete its life-cycle in the Arctic? Evidence and implications for Arctic food-webs. *Ambio*, *51*, 333–344.

Tarrant, A. M., Baumgartner, M. F., Hansen, B. H., Altin, D., Nordtug, T., & Olsen, A. J. (2014). Transcriptional profiling of reproductive development, lipid storage and molting throughout the last juvenile stage of the marine copepod *Calanus finmarchicus*. *Frontiers in Zoology*, *11*(1), 1–15.

R Core Team. (2021). *R: A language and environment for statistical computing*. R foundation for Statistical Computing. https://www.R-project.org/.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., DePristo, M. A., & Thibault, J. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(1110), 11.10.1–11.10.33.

Vartia, S., Villanueva-Cañas, J. L., Finarelli, J., Farrell, E. D., Collins, P. C., Hughes, G. M., Carlsson, J. E., Gauthier, D. T., McGinnity, P., FitzGerald, R. D., Mirimin, L., Crispie, F., Cotter, P. D., Carlsson, J., & Cross, T. F. (2016). A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. *Royal Society Open Science*, *3*(1), 150565.

Villarino, E., Chust, G., Licandro, P., Butenschön, M., Ibaibarriaga, L., Larrañaga, A., & Irigoien, X. (2015). Modelling the future biogeography of North Atlantic zooplankton communities in response to climate change. *Marine Ecology Progress Series*, *531*, 121–142. https://doi.org/10.3354/meps11299

Viruel, J., Haguenauer, A., Juin, M., Mirleau, F., Bouteiller, D., Boudagher-Kharrat, M., Ouahmane, L., La Malfa, S., Médail, F., Nieto Feliner, G., Baumel, A., & Sanguin, H. (2018). Advances in genotyping microsatellite markers through sequencing and consequences of scoring methods for *Ceratonia siliqua* (Leguminosae). *Applications in Plant Sciences*, *6*(12), e01201.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A., & Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, *22*(3), 787–798.

Wassmann, P., Kosobokova, K., Slagstad, D., Drinkwater, K., Hopcroft, R., Moore, S., Ellingsen, I. H., Nelson, R. J., Carmack, E., Berge, J., & Popova, E. (2015). The contiguous domains of Arctic Ocean advection: Trails of life and death. *Progress in Oceanography*, *139*, 42–65.

Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370. https://doi.org/10.2307/2408641

Wickham, H. (2016). Data Analysis. In *ggplot2. Use R!* Springer. https://doi.org/10.1007/978-3-319-24277-4_9

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Choquet, M., Lizano, A. M., Le Moan, A., Ravinet, M., Dhanasiri, A. K. S., & Hoarau, G. (2023). Unmasking microsatellite deceptiveness and debunking hybridization with SNPs in four marine copepod species of *Calanus*. *Molecular Ecology*, *00*, 1–20. https://doi.org/10.1111/mec.17183