LIMNOLOGY and OCEANOGRAPHY

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"

Geneviève J. Parent [ID],[1]* Stéphane Plourde,[2] Howard I. Browman,[3] Julie Turgeon,[4] Adam Petrusek [ID][5]
[1]Laboratory of Genomics, Maurice Lamontagne Institute, Fisheries and Oceans Canada, Mont-Joli, Quebec, Canada
[2]Maurice Lamontagne Institute, Fisheries and Oceans Canada, Mont-Joli, Quebec, Canada
[3]Institute of Marine Research, Ecosystem Acoustics Group, Austevoll Research Station, Storebø, Norway
[4]Département de Biologie, Université Laval, Quebec, Quebec, Canada
[5]Department of Ecology, Faculty of Science, Charles University in Prague, Prague, Czech Republic

Abstract

The article 'No evidence for hybridization between *Calanus finmarchicus* and *C. glacialis* in a subarctic area of sympatry' (Choquet et al. 2020) concludes that "no evidence supports a potential for hybridization between *C. finmarchicus* and *C. glacialis*". We argue that the InDel markers used by Choquet et al. (2020) may have limited capacity to detect admixed genotypes between *C. finmarchicus* and *C. glacialis* due to an inappropriate choice of reference sample for each species during the marker development. We first review terms and concepts used in genetic classification using reference samples and describe problems associated with the selection of genetic markers in the context of possible hybridization. We reanalyze InDel genotypes provided with Choquet et al. (2017) and identified an admixed individual. We then contrast methods used by Choquet et al. (2017) and Parent et al. (2012) and explain how Parent et al. (2012) developed microsatellite markers capable of discriminating admixed genotypes from parental species. In this comment, we have identified a major issue that must be considered when selecting reference samples in the context of testing for possible hybridization.

In their article "No evidence for hybridization between *Calanus finmarchicus* and *C. glacialis* in a subarctic area of sympatry," Choquet et al. (2020) conclude that "no evidence supports a potential for hybridization between *C. finmarchicus* and *C. glacialis*" and they suggest that previous evidence of hybridization between *Calanus* species (Parent et al. 2012) may have resulted from the use of inadequate molecular tools. They based their conclusions on genotyping of 1126 individuals of these *Calanus* species using six insertion–deletion (InDel) markers developed by Smolina et al. (2014). Those same markers have also been used in two previous studies that conclude that there is no evidence for hybridization between *Calanus* species (Nielsen et al. 2014; Choquet et al. 2017). We argue that the InDel markers used again by Choquet et al. (2020), although appropriate for detecting first-generation hybrids, may have limited capacity to detect admixed genotypes between *C. finmarchicus* and *C. glacialis*

due to an inappropriate choice of reference sample for each species during the marker development. Therefore, although their dataset indicates a general absence of F1 hybrids in the examined samples, their conclusion that hybridization, and consequently introgression, does not occur or has not occurred between these two *Calanus* species is not as definitive as they claim.

We will first review terms and concepts used in genetic classification of species based on reference samples, and then present the problems associated with the selection of genetic markers in the context of possible hybridization. We will also reanalyze InDel genotypes provided with Choquet et al. (2017) and contrast their results with those of Parent et al. (2012) using microsatellite markers to screen for admixed genotypes.

Species can be discriminated using allelic variants occurring at different frequencies in each species. Choosing nuclear genetic markers to discriminate species in the simplified case presented in Fig. 1 is straightforward since the genetic composition of each species is completely distinct in some chromosomal sections. In reality, allele composition is frequently not entirely different (i.e., alleles are not fixed) in related species.

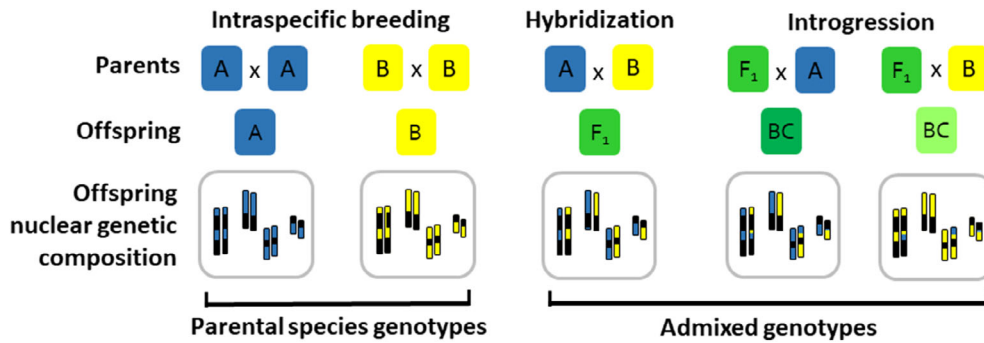*Correspondence: genevieve.parent@dfo-mpo.gc.ca

Fig. 1. Schematic representation of terms and concepts associated with nuclear genetic admixture between two species. The classification of each individual is based on nuclear genomic markers and is summarized as colors and letters in the colored squares for parents and offspring. Blue and yellow indicate genetic composition of species or parental species genotypes (A or B), whereas green indicates admixed genetic composition or admixed genotypes. F1 hybrids are issued from the first generation of interspecific breeding or hybridization, whereas backcrossed (BC) individuals are results from breeding between F1 hybrids and a parental species leading to introgression in future generations. Note that the term hybrid may be used as a synonym for admixed individuals (Parent et al. 2012, Harrison and Larson 2014) but it will be used specifically for F1 hybrids in this comment to avoid confusion. The nuclear genetic composition of offspring is represented with four pairs of chromosomes (diploid). Colors in the chromosomes indicate regions of the genome that allow species discrimination (blue and yellow) and others that cannot due to sequence similarity (black). Discriminant and common chromosomal regions may be shorter and more numerous but for clarity they are represented as large bands on chromosomes. For some taxa, BC individuals may reproduce with parental species or admixed individuals generating offspring with nuclear genetic composition that is even more admixed than what is presented here.
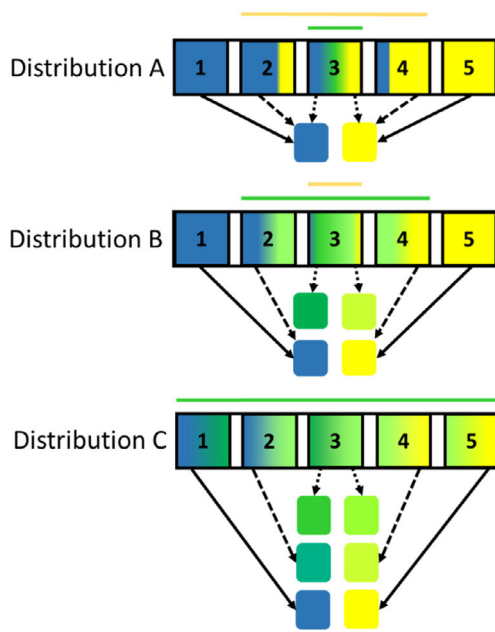

Fig. 2. Effect of sampling location on the genetic composition of reference samples in the presence of hybridization and introgression. We present three distributions (A, B, C) differing in the geographic extent of the hybrid zone (green line) and of the area where parental species are sympatric (orange line). Three sampling designs in five (1–5) regions are presented for each distribution; they are presented as full, dashed, and dotted lines. For each sampling design, the genetic composition of reference samples is presented as two squares using blue and yellow colors for parental species genotypes and a gradient of greens for admixed genotypes using possibly the most morphologically distinct specimens (Fig. 1). Genotyping of all individuals from a sample is done in a first step and selection of the most genetically distinct individuals as reference samples in a second step.

Even in cases of allele sharing between species (i.e., shared allele polymorphism), differences in allele frequencies at multiple loci make it possible to identify species.

If hybridization or introgression occurs, F1 hybrids and backcrossed individuals have admixed genetic composition compared to that of parental species and the level of this admixture is variable (see Fig. 1 and its caption for more definitions). Discriminating if allele sharing is due to incomplete lineage sorting (i.e., species share alleles jointly inherited from their last common ancestor) and introgression is a difficult task. To achieve it in the context of possible hybridization, the choice of representatives for parental species (i.e., reference samples) and of nuclear genetic markers is crucial. Those two factors, important to determine the capacity of a set of nuclear genomic markers to discriminate between incomplete lineage sorting and introgression, are explained in the following paragraphs.

Avoiding admixed individuals in reference samples is essential to study hybridization and introgression. In the presence of interspecific gene flow, the probability of selecting admixed individuals as reference samples for parental species may vary in a sympatric zone (Fig. 2; Table 1). If the hybrid zone is narrow, the probability is low that reference samples will contain admixed genotypes (distribution A, Fig. 2). In contrast, this probability is high if pure parental species are rare or absent from some regions of the taxon distribution (distributions B and C, Fig. 2). In the latter case, allele frequencies will be more similar on average over all loci between reference samples (i.e., two different greens in reference samples, Fig. 2). Discriminating between incomplete lineage sorting and introgression would be an impossible task using unbiased genome wide markers.

Table 1. Effect of genetic admixture on marker selection to discriminate parental species and admixed genotypes. Each line represents a reference individual that was morphologically identified and genotyped at four nuclear loci (L1–L4). For each locus, F or G indicates the alleles of the diploid species. The last two columns show the classification results (F or G for parental species individuals; A for admixed individuals) when reference samples were collected in regions with parental species genotypes (PSG, e.g., distribution A in Fig. 2) or admixed genotypes (ADG, e.g., distributions B or C, Fig. 2). The gray highlight indicates the nuclear loci (L2 and L3) that would not be selected for species discrimination in reference samples using admixed-types. Classification of individuals is based on either four or two loci for parental species genotypes (reference samples PSG) and admixed genotypes (reference samples ADG), respectively. Using admixed genotypes in reference samples impede or restrain the classification of individuals as admixed genotypes.

Morphological identification	Loci selection				Classification	
	L1	L2	L3	L4	PSG	ADG
G	GG	GG	GG	GG	G	G
G	GG	GG	GG	GG	G	G
G	GG	GG	GG	GG	G	G
G	GG	GG	FG	GG	A	G
G	GG	FG	GG	GG	A	G
F	FF	FF	FF	FF	F	F
F	FF	FG	FF	FF	A	F
F	FF	FF	FG	FF	A	F
F	FF	FF	FF	FF	F	F
F	FF	FF	FF	FF	F	F

The choice of genetic markers used to discriminate species will vary with the "purity" of reference samples for a small set or genome wide markers (Table 1). For example, in reference samples parental species genotype (Table 1), the nuclear loci used to discriminate between species are selected from parental species genotypes. In contrast, in reference samples admixed genotype (ADG), the nuclear loci used to discriminate between species are selected from the most different admixed genotypes (Table 1; Fig. 2). Some regions of the genome may be substantially more differentiated than others, as genetic drift or introgression tend to be variable across the genome (reviewed in Harrison and Larson 2014). Genetic markers in those differentiated regions could then be preferentially selected as the presumed species-specific ones since genomic regions affected by admixture would be considered uninformative. In the latter case, even strongly admixed genotypes (except F1 hybrids) might be identified as pure parental species genotypes since markers indicating admixture would not be used for genotyping. In the given example, real parental species as well as admixed genotypes would be identified as parental species genotypes if the reference samples ADG were used to select markers for species discrimination. To summarize, genotyping with markers issued from reference samples ADG will impede the ability to appropriately detect admixed genotypes.

To discriminate species, Smolina et al. (2014) selected 12 InDel markers from 6 individuals collected in a zone of sympatry between *C. finmarchicus* and *C. glacialis*. Reference samples were selected in two steps from a total of three regions. In a first step, massive parallel sequencing was performed on a single individual for each species sampled from
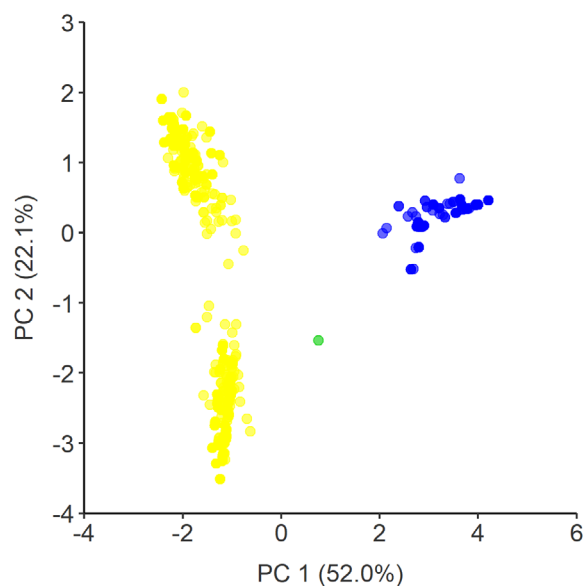

Fig. 3. Principal component analysis (PCA) for 3807 individuals genotyped at six InDel markers (Choquet et al. 2017). *C. glacialis*, *C. finmarchicus*, and admixed-type individuals, are presented as blue, yellow, and green, respectively. Classification is based on Choquet et al. (2017) except for the admixed genotype individual, which was originally identified as *C. glacialis*. For the PCA analysis, we used the dudi.pca function of the *adegenet* package in R (Core Team 2019) with default parameters. Missing data were replaced using the tab function (mean method) and species classification was used to simulate mean alleles.

what they have identified as "an area of minimal sympatry," where one species is much more abundant (Svinoy Island for *C. finmarchicus* and Rijpfjorden for *C. glacialis*; Smolina

et al. 2014). That single individual was used for an in silico selection of InDel markers (involving bioinformatics). In a second step, InDel markers were selected after in vitro testing using two reference individuals for each species from an area of sympatry where both species are almost equally abundant (Disco Bay, Greenland). This last step selected for 12 InDel markers that were able to discriminate species, which was the main objective of that study. In selecting for markers that discriminate *Calanus* species from reference samples collected from an area of sympatry, Smolina et al. (2014) could not eliminate the possibility that they inadvertently excluded markers that affected by admixture between the species (reference samples ADG, Table 1). Loci less introgressed would have been preferentially selected among the 12 InDel markers used by Smolina et al. (2014) and, if this were the case, admixed genotypes would be then classified as parental species genotypes using these InDel markers.

Nonetheless, the InDel markers selected by Smolina et al. (2014) should identify F1 hybrids, and possibly admixed genotypes that possess a combination of both parental species alleles at most loci. To assess whether there is any evidence for such individuals, we reran a principal component analysis (methods provided in the caption of Fig. 3) on 3807 individuals genotyped at six InDel markers from Choquet et al. (2017). While the InDel markers clearly split the dataset into two clusters representing the two species, there is also an individual identified in Choquet et al. (2017) as a *C. glacialis* that has an admixed genotype (i.e., heterozygous for presumably parental species-specific alleles at two loci, homozygous for different parental species alleles at two loci). No F1 hybrid, however, was detected among these individuals, indicating that contemporary hybridization was not detectable in this sample. However, even these InDel markers may indicate that some introgression between *C. finmarchicus* and *C. glacialis* occurs or has occurred in the past. Past introgression with limited—or even absent—contemporary hybridization might be a plausible alternative explanation of the discordance between Parent et al. (2012) and the subsequent publications based on InDel markers.

Choquet et al. (2020) claim that microsatellites from Parent et al. (2012) "were initially developed for *C. finmarchicus* only, for studying genetic differentiation among populations and were therefore not reliable tools to be applied to *C. glacialis* or to characterize hybrids." They argue that "microsatellites are generally not the most suited molecular markers for species identification and hybrid detection because of frequent occurrences of null alleles (Dakin and Avise 2004), possible homoplasy when comparing two species (Chambers and MacAvoy 2000), high mutation rate, and difficulties to score alleles (Pompanon et al. 2005; Selkoe and Toonen 2006)." We argue that the 10 microsatellite markers used by Parent et al. (2012) could legitimately discriminate parental species and admixed genotypes. Six of these 10 microsatellite markers were indeed originally developed for *C. finmarchicus* (Provan et al. 2007). Cross-amplification in other species may result in

null alleles, impeding the identification of hybridization or introgression. However, null alleles would not lead to the identification of false admixed genotypes, but rather underestimate the admixture rates. Further, Parent et al. (2012) used 46 and 48 reference individuals from allopatric zones to estimate allelic frequency for each parental species. Shared polymorphism was identified at all loci but some of them had private alleles that should allow discrimination of incomplete lineage sorting from introgression (Parent et al. 2012). Estimates of allele frequencies for microsatellite markers based on samples of at least 25 individuals are usually similar to that obtained from larger sample sizes (Hale et al. 2012). Parent et al. (2012) then simulated F1 hybrids using reference samples from allopatric zones and performed assignment to parental and admixed genotypes using two methods to validate the reliability of their classification as suggested by Manel et al. (2005). Thus, Parent et al. (2012) used adequate reference samples (i.e., sourced from allopatric zones and large sample sizes) to select genetic markers and provided robust analyses to discriminate not only parental species genotypes but also parental species genotypes from admixed genotypes.

We recommend that hybridization and introgression in *Calanus* species be studied using not only adequate reference samples for genetic marker selection but also new sets of genome wide makers. The markers so far used do not have the power to precisely detect gene flow or variable levels of introgression among genomic regions. Ancestral introgression may affect a limited portion of the genome, which would be difficult to detect with a small number of molecular markers, especially in case their selection might have been biased towards the unaffected genomic regions. Markers used by Parent et al. (2012) could not reliably discriminate F1 hybrids and backcrossed individuals using only the 10 microsatellites due to poor assignment results (Parent et al. 2012). To characterize the extent of gene flow between *C. finmarchicus* and *C. glacialis*, new sets of large numbers of markers (single nucleotide polymorphisms, SNPs) using appropriate reference individuals for each species should be used.

Future studies using genomic markers in *Calanus* species should also consider that hybridization and/or introgression may be spatially heterogeneous, being more frequent in some areas than others, and that admixed genotypes may be more abundant at some specific depths and during some periods of the year. There is little geographic overlap among Choquet et al. (2017), Choquet et al. (2020), and Parent et al. (2012), which may also explain the inconsistencies in the abundance of admixed genotypes between the two studies. Admixed individuals may also have an intermediate phenotype to that of parental species (reviewed in Goulet et al. 2017). Parent et al. (2015) observed that the reproductive phenology of admixed individuals was intermediate to that of *C. finmarchicus* and *C. glacialis*. Sampling active reproductive females between the peaks of species-specific spawning may help to target admixed genotypes.

Recent articles that used InDel markers and concluded that hybridization does not occur in *Calanus* species risk reducing

the already very limited interest in hybridization in marine pelagic zooplankton. We argue that we should not "throw the baby out with the bathwater": despite the apparent absence of F1 hybrids in these samples, admixture between *Calanus* species cannot be yet dismissed and the phenomenon should be studied further. Interspecific hybridization occurs across a wide taxonomic range and at relatively high frequencies between some taxonomic groups (Schwenk et al. 2008). Hybridization and introgression are observed in some intertidal copepods (Pritchard et al. 2011) and freshwater zooplankton (Petrusek et al. 2008; Vergilino et al. 2011; Xu et al. 2011; Liu et al. 2018). Importantly, hybridization may play a role in influencing the evolution of a species' range (Lewontin and Birch 1966; Pfennig et al. 2016). Marine pelagic zooplankton would be interesting species to test this hypothesis due to their wide distribution, high potential for dispersal and potentially limited zones of genetic admixture. Admixed genetic composition may underlie the resilience of the species to past and recent changes in environmental conditions and may contribute to the success of *Calanus* species.

Conflict of Interest

None declared.

Data Availability Statement

Genotypes from Parent et al. (2012) and (2015) as well as R script and data used in this comment are available on github. com/gjparent/2021-LO_Calanus_Comment.

References

Chambers, G. K., and E. S. MacAvoy. 2000. Microsatellites: Consensus and controversy. Comp. Biochem. Physiol. B Biochem. Mol. Biol. **126**: 455–476. doi:10.1016/S0305-0491(00)00233-9

Choquet, M., and others. 2017. Genetics redraws pelagic biogeography of *Calanus*. Biol. Lett. **13**: 20170588. doi:10.1098/rsbl.2017.0588

Choquet, M., G. Burckard, S. Skreslet, G. Hoarau, and J. E. Søreide. 2020. No evidence for hybridization between *Calanus finmarchicus* and *Calanus glacialis* in a subarctic area of sympatry. Limnol. Oceanogr. **66**: S314–S325. doi:10.1002/lno.11583

Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from https://www.R-project.org/

Dakin, E. E., and J. C. Avise. 2004. Microsatellite null alleles in parentage analysis. Heredity **93**: 504–509. doi:10.1038/sj.hdy.6800545

Goulet, B. E., F. Roda, and R. Hopkins. 2017. Hybridization in plants: Old ideas, new techniques. Plant Physiol. **173**: 65–78. doi:10.1104/pp.16.01340

Hale, M. L., T. M. Burg, and T. E. Steeves. 2012. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PLoS One **7**: e45170. doi:10.1371/journal.pone.0045170

Harrison, R. G., and E. L. Larson. 2014. Hybridization, introgression, and the nature of species boundaries. J. Hered. **105**: 795–809. doi:10.1093/jhered/esu033, S1

Lewontin, R. C., and L. C. Birch. 1966. Hybridization as a source of variation for adaptation to new environments. Evolution **20**: 315–336.

Liu, P., and others. 2018. Species and hybrids in the genus *Diaphanosoma* Fischer, 1850 (Crustacea: Branchiopoda: Cladocera). Mol. Phylogenet. Evol. **118**: 369–378. doi:10.1016/j.ympev.2017.10.016

Manel, S., O. E. Gaggiotti, and R. S. Waples. 2005. Assignment methods: Matching biological questions with appropriate techniques. Trends Ecol. Evol. **20**: 136–142. doi:10.1016/j.tree.2004.12.004

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

Parent, G. J., S. Plourde, P. Joly, and J. Turgeon. 2015. Phenology and fitness of *Calanus glacialis*, *C. finmarchicus* (Copepoda), and their hybrids in the St. Lawrence Estuary. Mar. Ecol. Prog. Ser. **524**: 1–9. doi:10.3354/meps11240

Parent, Geneviève J., Plourde, Stéphane, and Turgeon, Julie 2012. Natural hybridization between Calanus finmarchicus and C. glacialis (Copepoda) in the Arctic and Northwest Atlantic. Limnol. Oceanogr. **57**: 1057–1066. doi:10.4319/lo.2012.57.4.1057

Petrusek, A., J. Seda, J. Macháček, S. Ruthová, and P. Šmilauer. 2008. *Daphnia* hybridization along ecological gradients in pelagic environments: The potential for the presence of hybrid zones in plankton. Phil. Trans. R. Soc. B **363**: 2931–2941. doi:10.1098/rstb.2008.0026

Pfennig, K. S., A. L. Kelly, and A. A. Pierce. 2016. Hybridization as a facilitator of species range expansion. Proc. R. Soc. B. **283**: 20161329. doi:10.1098/rspb.2016.1329

Pompanon, F., A. Bonin, E. Bellemain, and P. Taberlet. 2005. Genotyping errors: Causes, consequences and solutions. Nat. Rev. Genet. **6**: 847–859. doi:10.1038/nrg1707

Pritchard, V. L., L. Dimond, J. S. Harrison, C. Cristina, S. Velázquez, J. T. Zieba, R. S. Burton, and S. Edmands. 2011. Interpopulation hybridization results in widespread viability selection across the genome in *Tigriopus californicus*. BMC Gen. **12**: 54. doi:10.1186/1471-2156-12-54

Provan, J., G. E. Beatty, C. A. Maggs, and G. Savidge. 2007. Expressed sequence tag-derived microsatellites for the cool-water marine copepod *Calanus finmarchicus*. Mol. Ecol. Notes **7**: 1369–1371. doi:10.1111/j.1471-8286.2007.01889.x

Schwenk, K., N. Brede, and B. Streit. 2008. Introduction. Extent, processes and evolutionary impact of interspecific

hybridization in animals. Philos. Trans. R. Soc. B Biol. Sci. **363**: 2805–2811. doi:10.1098/rstb.2008.0055

Selkoe, K. A., and R. J. Toonen. 2006. Microsatellites for ecologists: A practical guide to using and evaluating microsatellite markers. Ecol. Lett. **9**: 615–629. doi:10.1111/j.1461-0248.2006.00889.x

Smolina, I., and others. 2014. Genome- and transcriptome-assisted development of nuclear insertion/deletion markers for *Calanus* species (Copepoda: Calanoida) identification. Mol. Ecol. Resour. **14**: 1072–1079. doi:10.1111/1755-0998.12241

Vergilino, R., S. Markova, M. Ventura, M. Manca, and F. Dufresne. 2011. Reticulate evolution of the *Daphnia pulex* complex as revealed by nuclear markers. Mol. Ecol. **20**: 1191–1207. doi:10.1111/j.1365-294X.2011.05004.x

Xu, L., B.-P. Han, K. Van Damme, A. Vierstraete, J. R. Vanfleteren, and H. J. Dumont. 2011. Biogeography and evolution of the Holarctic zooplankton genus Leptodora (Crustacea: Branchiopoda: Haplopoda). J. Biogeogr. **38**: 359–370. doi:10.1111/j.1365-2699.2010.02409.x