RESOURCE ARTICLE

Proteomic fingerprinting enables quantitative biodiversity assessments of species and ontogenetic stages in *Calanus* congeners (Copepoda, Crustacea) from the Arctic Ocean

Sven Rossel[1] | Patricia Kaiser[2] | Maya Bode-Dalby[2] | Jasmin Renz[3] | Silke Laakmann[4,5] | Holger Auel[2] | Wilhelm Hagen[2] | Pedro Martínez Arbizu[1] | Janna Peters[3]

[1]German Centre for Marine Biodiversity Research (DZMB), Senckenberg Research Institute, Wilhelmshaven, Germany

[2]Universität Bremen, BreMarE – Bremen Marine Ecology, Marine Zoology, Bremen, Germany

[3]German Centre for Marine Biodiversity Research (DZMB), Senckenberg Research Institute, Hamburg, Germany

[4]Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), Oldenburg, Germany

[5]Alfred Wegener Institute, Helmholtz-Centre for Polar and Marine Research (AWI), Bremerhaven, Germany

Correspondence
Sven Rossel, German Centre for Marine Biodiversity Research (DZMB), Senckenberg Research Institute, 26382 Wilhelmshaven, Germany.
Email: sven.rossel@senckenberg.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: RE2808/3-1 and RE2808/3-2; Shiptime, Grant/Award Number: AWI_PS121; Volkswagen Foundation and Ministry for Science and Culture of Lower Saxony, Grant/Award Number: ZN3285

Handling Editor: Michael Hansen

Abstract

Species identification is pivotal in biodiversity assessments and proteomic fingerprinting by MALDI-TOF mass spectrometry has already been shown to reliably identify calanoid copepods to species level. However, MALDI-TOF data may contain more information beyond mere species identification. In this study, we investigated different ontogenetic stages (copepodids C1–C6 females) of three co-occurring *Calanus* species from the Arctic Fram Strait, which cannot be identified to species level based on morphological characters alone. Differentiation of the three species based on mass spectrometry data was without any error. In addition, a clear stage-specific signal was detected in all species, supported by clustering approaches as well as machine learning using Random Forest. More complex mass spectra in later ontogenetic stages as well as relative intensities of certain mass peaks were found as the main drivers of stage distinction in these species. Through a dilution series, we were able to show that this did not result from the higher amount of biomass that was used in tissue processing of the larger stages. Finally, the data were tested in a simulation for application in a real biodiversity assessment by using Random Forest for stage classification of specimens absent from the training data. This resulted in a successful stage-identification rate of almost 90%, making proteomic fingerprinting a promising tool to investigate polewards shifts of Atlantic *Calanus* species and, in general, to assess stage compositions in biodiversity assessments of Calanoida, which can be notoriously difficult using conventional identification methods.

KEYWORDS
Atlantification, *Calanus*, Fram strait, MALDI-TOF, ontogenetic stage, proteomic fingerprinting

1 | INTRODUCTION

Species identification is crucial for biodiversity assessments. Traditionally, these have been carried out by morphological identification of single specimens. However, taxonomical work requires expert knowledge and is time-consuming when for example, dissection of body parts is required for species identification. Especially in groups with enormous abundances such as planktonic calanoid copepods, species identification can consume many working hours. Usually, identification keys are based on adult characters only. As an alternative to this taxonomic work, methods such as COI barcoding have been introduced for species identification (Hebert et al., 2003). Using self-assessed or public reference libraries like BOLD (Ratnasingham & Hebert, 2007) or GenBank (Benson et al., 2012), genetic data can be used for identification without a comprehensive, taxonomic knowledge. However, quantitative genetic applications are costly. Moreover, it may be of interest to receive information beyond the level of mere species identification, for example, to investigate population dynamics and cohort analysis of natural communities (Laakmann et al., 2020). Genetic barcodes such as COI are invariant through the life of the organism, thus having the advantage of allowing species identification from the egg to adult stage. A caveat is that no differentiation between ontogenetic stages can be done.

Proteomic fingerprinting using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry facilitates species identification based mainly on peptide- and protein compositions of specimens (Singhal et al., 2015). Molecules are embedded in a matrix solution to protect these from destruction by radiation through a laser during analyte ionization (Singhal et al., 2015). Molecule mass is measured by time-of-flight through a TOF tube towards a detector, resulting in species-specific proteome fingerprints. The method originates from microbiology where it is applied for pathogen identification (Chen et al., 2021; Papagiannopoulou et al., 2020; Tan et al., 2012; Van Driessche et al., 2019). However, in metazoan identification it is still in its infancy. Nevertheless, successful species identification of specimens has been achieved from a large variety of taxonomic groups such as different crustaceans (Bode et al., 2017; Hynek et al., 2018; Kaiser et al., 2018; Kürzel et al., 2022; Laakmann et al., 2013; Paulus et al., 2022; Rossel et al., 2019; Rossel & Martínez Arbizu, 2018b, 2019), fish (Maász et al., 2017; Mazzeo et al., 2008; Rossel et al., 2020; Volta et al., 2012), cnidarians (Holst et al., 2019; Korfhage et al., 2022), molluscs (Hamlili, Thiam, et al., 2021; Wilke et al., 2020) and a large variety of insects, preferentially those which are potential disease vectors (Dieme et al., 2014; Hasnaoui et al., 2022; Loaiza et al., 2019; Mathis et al., 2015; Nabet et al., 2021; Nebbak et al., 2017; Yssouf et al., 2014).

Calanus species belong to the most important copepods with regard to biomass and food-web dynamics in many parts of the North Atlantic basins. However, identification to species level remains challenging despite its relevance. In Arctic Fram Strait, three species of *Calanus* co-occur, two "true-Arctic" species, *C. hyperboreus* Krøyer, 1838 and *C. glacialis* Jaschnov, 1955, and the boreal-Atlantic expatriate, *C. finmarchicus* (Gunnerus, 1770) (Choquet et al., 2017).

While in *C. hyperboreus* copepodids of stage C4 and older are easily identified by morphology and size (Brodskii et al., 1983), the morphological identification of the congeneric species *C. finmarchicus* and *C. glacialis* is problematic due to the absence of consistent differences. A common method to distinguish *C. finmarchicus* from *C. glacialis* is based on differences in prosome length. However, this approach is inaccurate because the species overlap in body size in areas, where they co-occur (Gabrielsen et al., 2012; Lindeque et al., 2006; Trudnowska, Stemmann, et al., 2020; Weydmann & Kwasniewski, 2008). They also show high size plasticity depending on environmental conditions (Trudnowska, Balazy, et al., 2020). Recent comparisons of genetic and morphological identification indicate that no morphological criterion can reliably distinguish between *C. glacialis* and *C. finmarchicus* (Choquet et al., 2018). Yet, accurate species identification is crucial considering the increasing northward shift of boreal species due to climate change (Beaugrand et al., 2009; Weydmann, Carstensen, et al., 2014), with a recent study indicating the successful recruitment of *C. finmarchicus* in Fram Strait (Tarling et al., 2022). A shift from generally larger and lipid-richer Arctic *C. hyperboreus* and *C. glacialis* to boreal-Atlantic, smaller and less lipid-rich *C. finmarchicus* has consequences for Arctic food-webs. For instance, little auks (*Alle alle* [Linnaeus, 1758]) actively select larger Arctic *Calanus* and even perform longer foraging trips to find their preferred prey (Kwasniewski et al., 2010). Thus, accurate, reliable and efficient species identification in order to monitor distribution developments is essential.

In spite of pronounced morphological similarities, genetic identification of *Calanus* congeners is possible at any of the six naupliar and six copepodid developmental stages with the use of a variety of different molecular tools (Choquet et al., 2018; Gabrielsen et al., 2012; Hill et al., 2001; Lindeque et al., 2006; Weydmann et al., 2017; Weydmann, Coelho, et al., 2014). Taking the above into account, the best current method for the effective species and developmental stage identification of this genus seems to be using conventional microscopy together with molecular markers, although it remains laborious.

Previously, studies have been carried out on species identification of Calanoida using MALDI-TOF MS and some authors mentioned specific clustering of different ontogenetic stages (Kaiser et al., 2018; Laakmann et al., 2013; Riccardi et al., 2012). However, these studies were carried out either as pilot studies to test the application of proteomic fingerprinting for identification or in actual biodiversity assessments. Thus, the existence of a stage-specific signal in calanoid identification has not been investigated in more detail. Because stage-level identification would add a further dimension of community composition in rapid biodiversity assessments, in this study on Arctic *Calanus* species we aimed at (i) testing for the existence of a stage-specific signal, and (ii) investigating the cause of this signal.

2 | MATERIALS AND METHODS

2.1 | Sampling

Specimens of *Calanus finmarchicus*, *Calanus glacialis* and *Calanus hyperboreus* were obtained from two sampling sites during the expedition PS121 in August/September 2019 with the research vessel *Polarstern* to Arctic Fram Strait. Station SV3 was located on the eastern side of Fram Strait close to Svalbard under strong Atlantic influence of the West Spitsbergen Current (78°59.898N; 8°14.930E). By contrast, station EG1 was on the western side of Fram Strait close to the East Greenland shelf and influenced by the polar East Greenland Current (78°58.854N; 5°21.639W). Both stations were sampled from the sea surface down to bottom depth (900 and 1000 m) using a multiple opening and closing net equipped with five nets (Hydrobios Multinet Midi, mouth opening 0.25 m^2, mesh size 150 μm). Immediately after the haul, the samples were preserved in undenatured ethanol (96%). Samples were stored for 2 years at 0°C before measurements.

Specimens of the three *Calanus* species were sorted from the ethanol-preserved samples by morphology using a dissecting microscope (Leica MZ12.5). For each sorted specimen, stage as well as prosome and urosome length were noted and, based on its size and sampling location, assigned to a species to allow a double check with MALDI-TOF identification results (see Table 1 for species- and stage-specific prosome lengths of sorted individuals). For *C. hyperboreus*, besides its much larger size, a protrusion at the prosome end from copepodite stage 4 on was used for morphological discrimination. Copepodite stages were determined based on number of swimming legs (P) and number of urosome (U) segments, that is, C1 has two P, U with two segments; C2 with three P, U with two segments; C3 with four P, U with two segments; C4 with five P, U with three segments; C5 with five P, U with four segments; adult female with five P, U with four segments, first segment of U with well-developed genital somite. Each specimen was cut in half, that is, in an anterior prosome and a posterior prosome including the urosome. Each half was stored separately in an Eppendorf tube with absolute ethanol for subsequent genetic and proteomic analyses.

2.2 | DNA barcoding

The posterior thoracic segments 2–5 in stages C4 to C6 females (C6f) including the urosome, for the younger stages C1–3 approximately the posterior body half, were used for DNA extraction and subsequent amplification of the COI barcode fragment of three randomly chosen specimens from each developmental stage (C1–C6f) and suspected species.

Samples were incubated in 30 μl chelex (InstaGene Matrix, Bio-Rad) for 50 min at 56°C, followed by a denaturation of the enzymes for 10 min at 96°C. Of the DNA extract, 2 μl were used as template for the PCR in a volume of 20 μl with 10 μl Accu Start (2× PCR master mix, Quantabio), 7.6 μl molecular grade water and 0.2 μl of each primer (20 pmol/μl). For amplification of the COI barcoding region, Folmer primers LCO1490 and HCO2198 (Folmer et al., 1994) were used for *C. hyperboreus*. Amplification was carried out with the following settings: initial step at 94°C for 5 min, denaturation step at 94°C for 45 s, annealing at 42°C for 45 s, elongation at 72°C for 80 s and a final elongation for 7 min. Denaturation, annealing and elongation were carried out in 38 cycles. Amplification of the COI fragment for *C. finmarchicus* and *C. glacialis* was done using Coxf and CoxR2 primers (Cheng et al., 2013). Cycler settings were: initial denaturation at 94°C for 5 min, denaturation at 94°C for 30 s, annealing at 44°C for 75 s, elongation at 72°C for 60 s and final elongation at 72°C for 3 min. Denaturation, annealing and elongation were repeated 42 times. Purification and sequencing of the resulting PCR products was carried out at Macrogen Europe, Amsterdam, Netherlands.

Resulting sequencing reads were assembled in Geneious R7 version 7.0.6. and checked for contamination (e.g., bacteria, fungi, noncrustacean taxa) as well as for correct species-level identification using the basic local alignment search tool (BLAST) (Altschul et al., 1997) against GenBank.

2.3 | MALDI-TOF MS measurements

The anterior prosome body parts of 179 specimens were used for MALDI-TOF MS measurements. Depending on size, these were incubated in 3–30 μl (in adults: 5 μl for *C. finmarchicus*, *C. glacialis* in 10 μl and *C. hyperboreus* in 30 μl) of a matrix solution covering the entire specimen with some supernatant. The matrix contained α-Cyano-4-hydroxycinnamic acid (HCCA) as a saturated solution in 50% acetonitrile, 47.5% molecular grade water, and 2.5% trifluoroacetic acid. After 5 min of incubation, 1.5 μl was transferred to a target plate for cocrystallization of matrix and molecules. Subsequently, measurements were carried out using a Microflex LT/SH System (Bruker Daltonics). Employing the flexControl 3.4. (Bruker Daltonics) software, molecule masses were measured from 2 to 20k Dalton (kDA). A centroid peak detection algorithm was carried out for peak evaluation by analysing the mass peak range from 2 to 20 kDa. Furthermore,

Species	Prosome length (mm)					
	C1	C2	C3	C4	C5	C6f
C. finmarchicus	0.5–0.7	0.9–1.0	1.1–1.4	1.6–1.9	2.1–2.8	2.4–3.0
C. glacialis	0.8–1.0	1.2–1.6	1.8–2.1	2.3–3.0	3.3–3.6	3.4–4.1
C. hyperboreus	–	1.5–1.7	2.3–2.8	3.5–3.9	4.4–5.4	5.9–6.7

TABLE 1 Prosome lengths [mm] of *Calanus* copepodid stages C1 to females (C6f) from Fram Strait included in this study

peak evaluation was carried out by a signal-to-noise threshold of two and a minimum intensity threshold of 600 with a peak resolution higher than 400. To validate fuzzy control, the proteins/oligonucleotide method was employed by maximal resolution of 10 times above the threshold. To create a sum spectrum, a total of 160 laser shots were applied to a spot. Each spot was measured three times.

2.4 | MALDI-TOF data processing

MALDI-TOF raw data were imported to R, version 4.1.0 (R Core Team, 2022) and processed using R packages MALDIquantForeign, version 0.12 (Gibb, 2015) and MALDIquant, version 1.20 (Gibb & Strimmer, 2012). Spectra were square-root transformed, smoothed using the Savitzky Golay method (Savitzky & Golay, 1964), baseline corrected using the SNIP method (Ryan et al., 1988) and spectra normalized using the TIC method. Repeated measurements were averaged by using mean intensities. Peak picking was carried out using a signal to noise ratio (SNR) of 12 and a half window size of 13. Mass peaks smaller than a SNR of 12 were, however, retained if they occurred in other mass spectra as long as these were larger than a SNR value of 1.75, which is assumed as a lower detection limit. Repeated peak binning was carried out to align homologous mass peaks. Resulting data was Hellinger transformed (Legendre & Gallagher, 2001) and used for further analyses.

Hierarchical clustering was carried out in R using Ward's D and Euclidean distances. Random Forest (RF) (Breimann, 2001) was carried out using the R package randomForest, version 4.6.14 (Liaw & Wiener, 2002). Settings were used according to Rossel and Martínez Arbizu (2018a) (ntree = 2000, mtry = 35, sampsize = number of specimens in the smallest class). Classifications were tested using the RF post hoc test (Rossel & Martínez Arbizu, 2018a), function rf.post.hoc in package RFtools version 0.0.3 (https://github.com/pmartinezarbizu/RFtools). Classifications were tested for correct class assignment based on empirical assignment probabilities of the RF model. Specimens with correct RF classification and assignment probabilities not deviating significantly ($p < .05$) from the empirical distribution were considered true positive (tp) assignments. Specimens with correct RF classification and significantly different assignment probability were recorded as false positives (fp). If RF classification was incorrect and the assignment probability for the class a specimen was assigned to did not differ significantly, the classification was recorded as true negative (tn). Considered as false negatives (fn) were specimens assigned to the incorrect class by RF but the assignment probability differed significantly from the empirical distribution of the respective class. To test for significant differences between stages, pairwise adonis of all species-stage combinations was carried out using the R-package pairwiseAdonis, version 0.4 (Martínez Arbizu, 2022) (https://github.com/pmartinezarbizu/pairwiseAdonis).

The respective RF models, using either species or species and stage as classes, were also used to find the most important variables for differentiation of the groups using the Gini index, which shows the degree of dissimilarity of the respective variables (Han et al., 2016). T-distributed stochastic neighbour embedding (t-SNE) plots were created using R package Rtsne version 0.15 (Krijthe, 2015) with the following settings: perplexity = 15, max_iter = 4000 and theta = 0.

2.5 | Dilution series

To test the effect of decreasing amounts of biomass on the stage-specific signal, a dilution series was carried out. The cephalosomes of five adult specimens per species were prepared and measured as described above. The amount of matrix used for the source specimens depended on the size of the respective species. *C. finmarchicus* cephalosome was incubated in 5 μl, *C. glacialis* in 10 μl and *C. hyperboreus* in 30 μl matrix solution. The remaining matrix solution was then used to set up a dilution series with dilutions of 1:1, 1:2, 1:4, 1:8, 1:16, 1:32, 1:64 and 1:128 with further HCCA matrix.

3 | RESULTS

From all 179 specimens a mass spectrum was obtained. Based on hierarchical clustering (Figure 1) and prior morphological staging of copepodites, specimens were randomly chosen for DNA barcoding to support species level identifications. In total, from 49 specimens a DNA barcode with up to 792 bp was obtained, supporting the identifications as *C. glacialis*, *C. hyperboreus* and *C. finmarchicus*. This is further supported by an RF classification model on species level without any misclassifications within the model.

Calanus finmarchicus was collected mainly from station SV3 at depths ranging from 0–50 m (C1 and C2) and 50–100 m (C3 to C6f). However, a single adult female was found between 200–500 m at station EG1. The identification of this specimen was supported based on proteomic fingerprinting and DNA barcoding. All *C. glacialis* specimens were obtained from station EG1. In depths from 0–50 m, all stages were found. From 50–200 m stages C2 and C3 and from 200–500 m only C6 females were identified. For *C. hyperboreus*, only stages from C2 to adult females were identified, which were all found in 0–50 m depth. Copepodids C2, C3 and C6 females were also retrieved from 50–200 m. The latter two were also collected at depths from 200–500 m.

Hierarchical clustering including all specimens resulted in distinct clusters for the three species (Figure 1) supporting the species-specificity of the proteomic fingerprint. Using an RF model on species level, none of the analysed specimens was misclassified. However, five classifications were recognized by the *post hoc* test as false positives (Table 2a). In addition, all species show a more or less pronounced clustering by stages. In *C. finmarchicus* C1, C2 and C6f specimens show strong stage-specific clustering with only two C1 specimens clustering with C2 specimens (Figure 1). Copepodids C3 to C5 also cluster stage-specifically, but clustering of single specimens with specimens from other stages happens more frequently. The same is true for the different stages of *C. hyperboreus* and *C.*
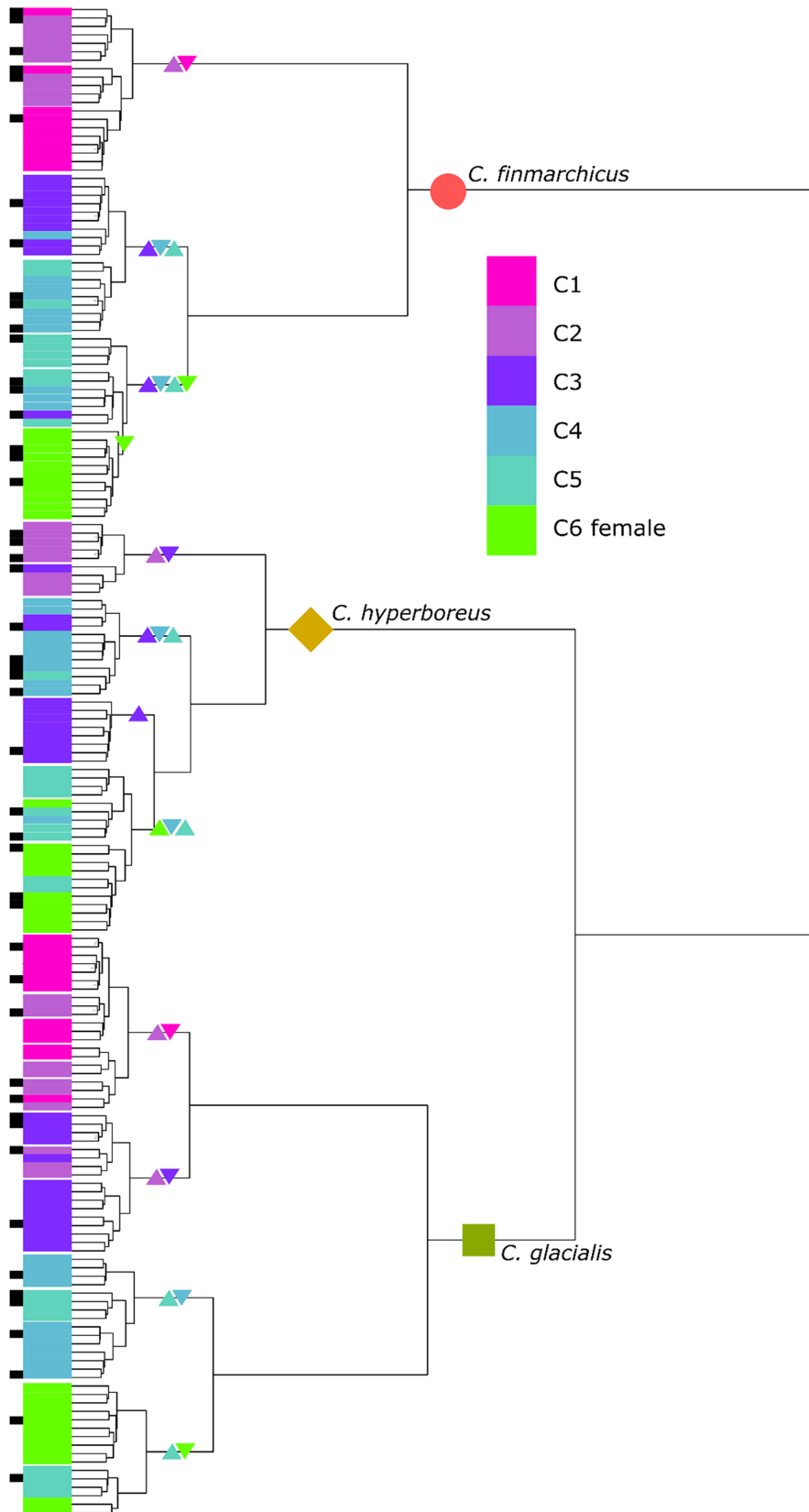
FIGURE 1 Hierarchical clustering of copepodite stages C1–C6f of the three *Calanus* species included in this study. Species names indicate branches including only specimens of the respective species. Triangles indicate copepodite stages included in a certain branch. Coloured tips indicate the stage of the respective specimen. The analysis supports a clear species-specific signal and illustrates similarity of certain copepodite stages within a species. Black bars indicate specimens analysed using DNA barcoding.

glacialis. In both species, stages tend to cluster into stage-specific clusters. However, variability is more pronounced compared to C1, C2 and C6f in *C. finmarchicus* (Figure 1). Species and stage differentiation is also emphasized by a t-SNE analysis, showing distinct groups for the species and, within these, clear separations for the respective copepodite stages (Figure 2a). This also depicts the more pronounced differentiation of *C. finmarchicus* C1 and C2 specimens compared to the remaining stages. Pairwise adonis between

TABLE 2 Classification errors within Random Forest (RF) models generated in this study and classification error of classification tests. In classification tests, one specimen was taken out of the model and classified using the remaining specimens. (a) Species level model without stages assigned within the classification test. (b) Species-stage level model and classification test. (c) Model containing only adult specimens and test of classification of juvenile specimens. (d) Model containing only juvenile specimens and test of classification of adult specimens

(a)

Species	Stage	Within RF model			RF classification		Post hoc test			
		n specimens	*n* misassigned	Class error	*n* correct	*n* incorrect	*n* tp	*n* fp	*n* fn	*n* tn
C. finmarchicus	Not assigned	61	0	0	61	0	57	4	0	0
C. glacialis	Not assigned	69	0	0	69	0	67	2	0	0
C. hyperboreus	Not assigned	49	0	0	49	0	45	4	0	0
Total		179	0		179	0	169	10	0	0

(b)

Species	Stage	Within RF model			RF classification		Post hoc test			
		n specimens	*n* misassigned	Class error	*n* correct	*n* incorrect	*n* tp	*n* fp	*n* fn	*n* tn
C. finmarchicus	C1	10	1	0.10	9	1	9	0	0	1
C. finmarchicus	C2	10	1	0.10	9	1	8	1	1	0
C. finmarchicus	C3	10	1	0.10	9	1	9	0	1	0
C. finmarchicus	C4	10	2	0.20	8	2	7	1	1	1
C. finmarchicus	C5	10	3	0.30	7	3	6	1	0	3
C. finmarchicus	C6f	11	0	0.00	11	0	10	1	0	0
C. glacialis	C1	13	0	0.00	13	0	13	0	0	0
C. glacialis	C2	11	1	0.09	10	1	9	1	0	1
C. glacialis	C3	14	1	0.01	13	1	12	1	0	1
C. glacialis	C4	11	0	0.00	11	0	11	0	0	0
C. glacialis	C5	8	1	0.13	7	1	6	1	1	0
C. glacialis	C6f	12	0	0.00	12	0	10	2	0	0
C. hyperboreus	C2	8	0	0.00	8	0	8	0	0	0
C. hyperboreus	C3	11	3	0.27	8	3	7	1	1	2
C. hyperboreus	C4	10	1	0.10	8	2	7	1	1	1
C. hyperboreus	C5	10	2	0.20	8	2	8	0	1	1
C. hyperboreus	C6f	10	2	0.20	9	1	8	1	0	1
Total		179	19		160	19	148	12	7	12

(c)

Model containing only adult specimens		Within RF model			Classification of juvenile specimens		Post-hoc test			
Species	Stage	*n* specimens	*n* misassigned	Class error	*n* correct	*n* incorrect	*n* tp	*n* fp	*n* fn	*n* tn
C. finmarchicus	Adult	11	0	0	50	0	30	20	0	0
C. glacialis	Adult	12	0	0	56	1	5	51	3	0
C. hyperboreus	Adult	10	0	0	36	3	21	15	0	0
Total		33	0		142	4	56	86	3	0

(d)

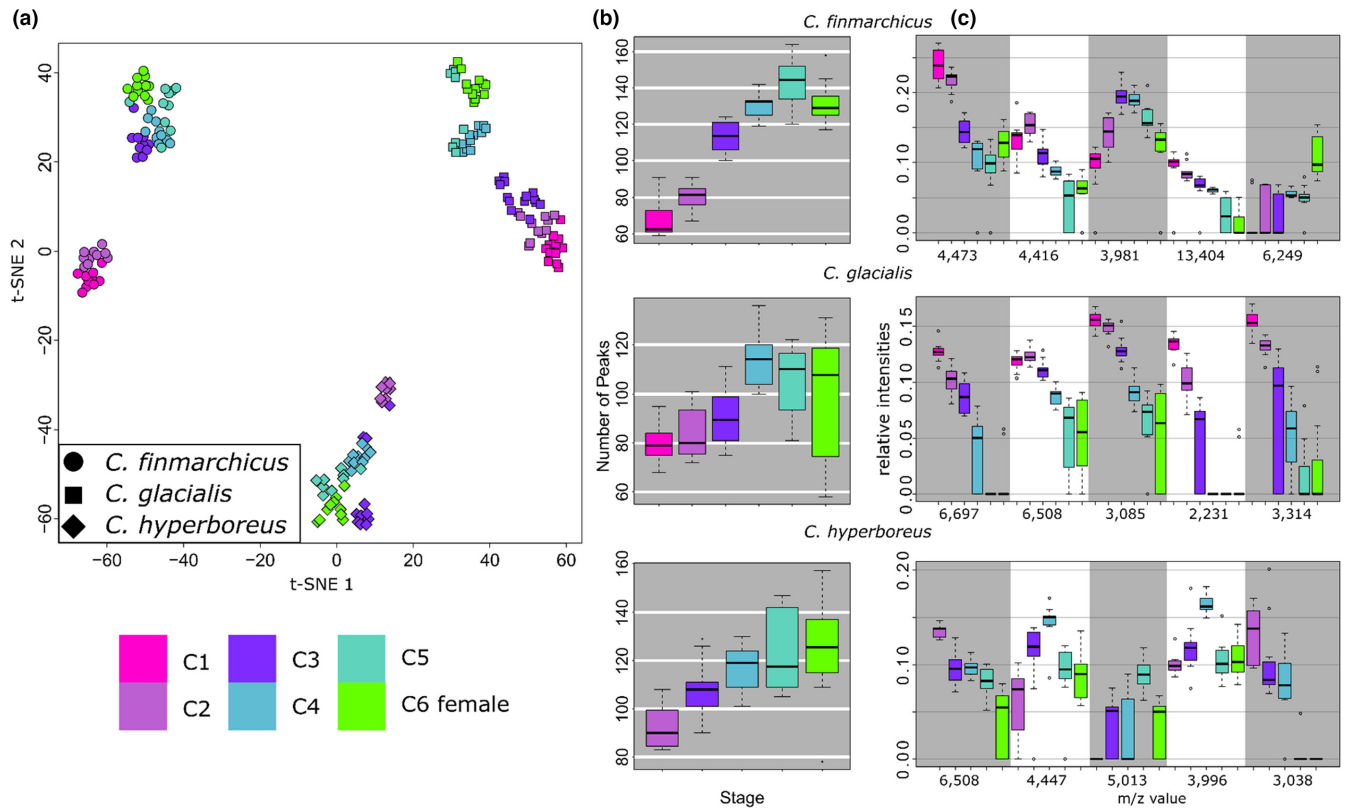| Model containing only juvenile specimens		Within RF model			Classification of adult specimens		Post-hoc test			
Species	Stage	*n* specimens	*n* misassigned	Class error	*n* correct	*n* incorrect	*n* tp	*n* fp	*n* fn	*n* tn
C. finmarchicus	Juvenile	50	0	0	11	0	10	1	0	0
C. glacialis	Juvenile	57	0	0	12	0	8	4	0	0
C. hyperboreus	Juvenile	39	0	0	10	0	10	0	0	0
Total		146	0		33	0	33	0	0	0

FIGURE 2 (a) t-SNE plot of the *Calanus* specimens included in this study, shape-coded by species and colour-coded by stage. (b) Number of peaks per species and stage. (c) Relative intensities of the five most important mass peaks for classification using Rrandom Forest displayed by species and stage. Colour coding in all images according to the legend in the bottom left side.

all species-stage combinations shows significant differences (*p*-value < .01) for all pairs, supporting differentiation of ontogenetic stages based on protein mass spectra.

A RF classification model using stages by species as classes results in model-class errors ranging from 0 to 0.3 (Table 2b). In *C. glacialis* all specimens were classified as the correct copepodite stage within the model except for one individual of stages C2, C3 and C5 each. In *C. finmarchicus*, no misclassification was found for the adult specimens; in C1 to C3 one specimen each was misclassified. Moreover, two specimens in C4 and three in C5 showed ambiguous signals. In *C. hyperboreus*, no misclassifications were found for C2 specimens. However, for the remaining stages between one to three specimens each showed ambiguous classifications within the model.

In a classification test, single specimens were removed from the RF training data set and subsequently classified using the classification model. On stage level, the classification test resulted in correct stage classifications for the vast majority. In total, of 179 specimens, 160 (89.4%) were staged correctly. Of the 19 misclassified specimens, seven are considered false negatives. Thus, their misclassification would have been recognized. The classifications of the remaining 12 specimens were considered correct by the *post hoc* test, making these unrecognized misclassifications (Table 2b). Twelve correctly classified specimens were considered false positives by the *post hoc* test. None of the specimens were assigned to a different species but only to a different stage. The majority of

these stage misclassifications occurred for neighbouring classes. Only one *C. finmarchicus* C3 was classified as a C5 specimen. Of 61 *C. finmarchicus* specimens eight (13.1%) were misassigned, of 49 *C. hyperboreus* eight (16.3%) were misassigned and for *C. glacialis*, of 69 specimens only two (2.9%) were misclassified. It must be noted that the training dataset was smallest for *C. hyperboreus*.

The classification approach mentioned above with a complete reference library showed high correct classification rates. However, comprehensive libraries are not always available. Thus, we tested the classification success in the absence of either adult or juvenile specimens.

Using a reference library including only adult specimens, the success rate was still very high. Of 145 classified specimens, four (2.8%) were classified as another species. However, only 56 were recognized by the *post hoc* test as true positives (Table 2c). The smaller the developmental stage of the classified specimen, the more likely it was to be recognized as false positive by the *post hoc* test with lower *p*-values. Conversely, including only C1–C5 specimens in the RF model resulted in successful classification of all adult specimens in the data set. Of these, five specimens were assigned by the *post hoc* test as false positives with *p* < .01 (Table 2d).

First differences among mass spectra of the different copepodite stages are apparent, when looking at the raw spectra of the different stages (Figure 3). With increasing ontogenetic stages, dominant mass peaks of sizes larger than *m/z* 10,000 are becoming
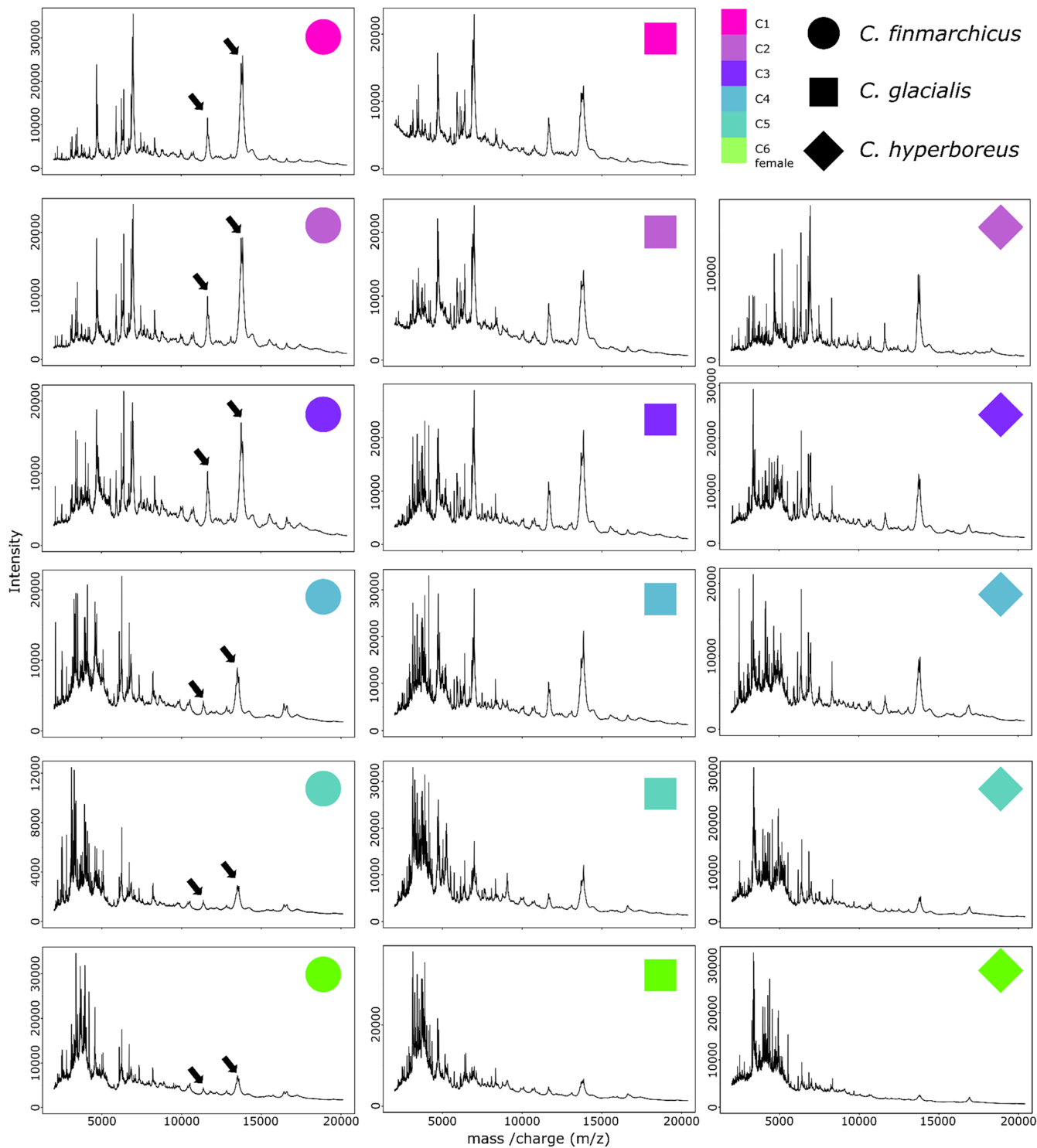
FIGURE 3 Raw mass spectra from all copepodite stages of the investigated three *Calanus* species. Displayed mass spectra are averaged from all specimens of the respective species and stage. In all species dominant peaks of larger molecules (>*m/z* 10,000) show decreasing relative intensities with increasing ontogenetic stages (arrowed in *C. finmarchicus*).

smaller in relation to the remaining signal (Figure 3 arrowed). This remains true for all three investigated species. However, regarding the most important mass peaks for stage differentiation (Figure 2c), as retained from RF analyses, among the five most important peaks for stage classification almost no peak larger than *m/z* 10,000 was found. Only in *C. finmarchicus* the 13,404 peak was identified as the

decisive peak, which showed decreasing relative intensity with increasing developmental stages (Figure 2c). From the RF model, important peaks for stage classification were present in most stages rather than being present in a single stage. Older stages of all species generally showed more peaks (Figure 2b) but at the same time decreasing intensities of the important classification peaks (Figure 2c).

Thus, relative intensities seem to be important for discrimination and the differences among copepodite stages may be attributable to increasing mass spectra complexity with the ontogenetic stage and partly to the relative intensities of some larger molecules.

Even though the results rather point at relative intensities being responsible for stage discrimination in RF classification, still an effect of less biomass resulting in less overall mass peaks and subsequently in stage distinction needs to be excluded. Thus, using a dilution series of material taken from additional adult specimens, the influence of the amount of biomass on mass spectra was tested. However, dilutions of up to 1:128 of the original concentration did not result in patterns, as they were found in mass spectra from smaller ontogenetic stages (Figure 4). Also in clustering approaches, t-SNE plots (Figure 5a) or RF models and classification tests, these were not more likely to be assigned to smaller stages. At least for *C. finmarchicus* and *C. hyperboreus*, dilution of matrix:tissue ratio resulted in a decreasing number of peaks (Figure 5b) making the amount of peaks comparable to smaller ontogenetic stages. However, this did not result in making C6f mass spectra more similar to those obtained from smaller ontogenetic stages. This is also emphasized by Euclidean

distances of the different stages and dilutions to C1 (C2 in *C. hyperboreus*) (Figure 5c). In all species, Euclidean distances of mass spectra from different stages increased with each stage. The distances further increased to mass spectra obtained from diluted samples, even though the number of peaks became more similar to mass spectra obtained from smaller stages (compare Figure 2b and Figure 5b).

4 | DISCUSSION

Applying three *Calanus* congeners as model cases, we showed that species identification by proteomic fingerprinting is stable in Calanoida over all ontogenetic stages despite increasing complexity of proteomic profiles during development. Species identification accuracy was 100%, similar to high success rates shown by other studies on Calanoida (Bode et al., 2017; Kaiser et al., 2018; Laakmann et al., 2013). With a success rate of almost 90%, the RF model assigned specimens even to their ontogenetic stages with misassignments mainly occurring between neighbouring ontogenetic stages. In an actual application, based on *post hoc* test results, classifications
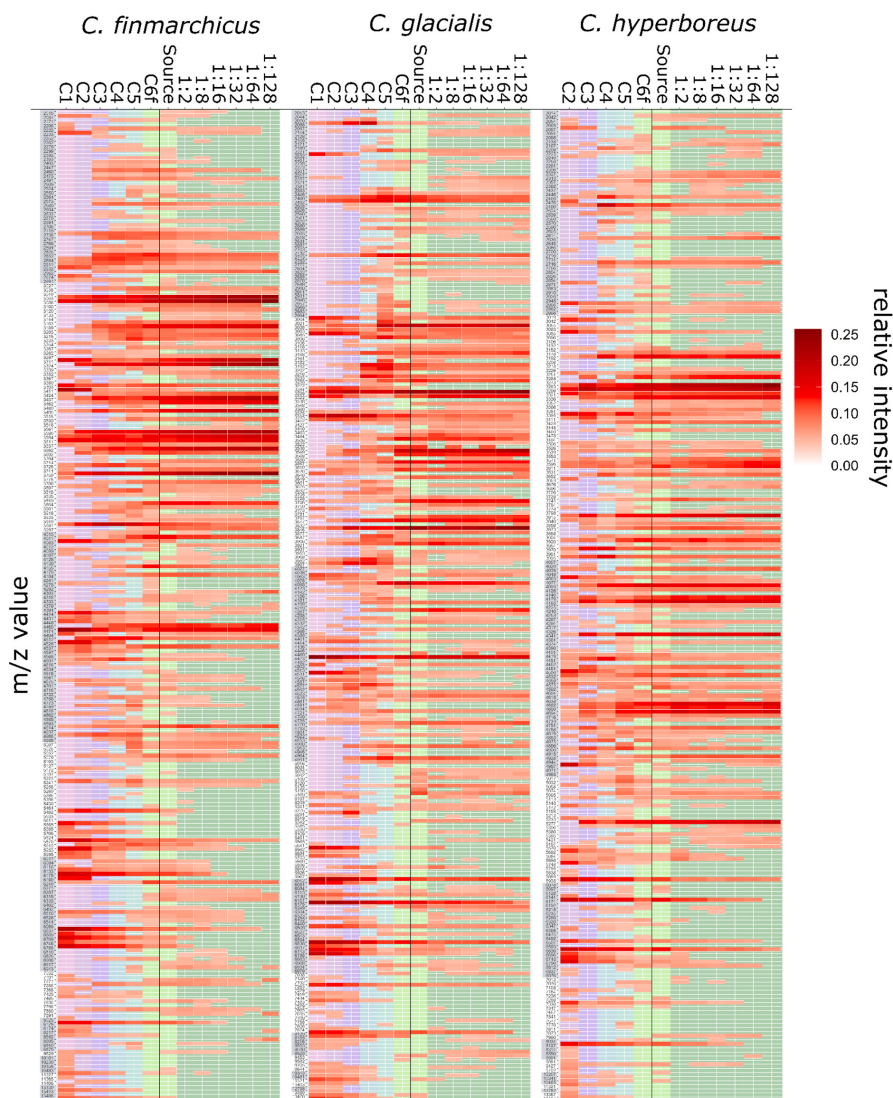

FIGURE 4 Peaks and the respective relative intensities shown for the three *Calanus* species separately by stage and dilution. Sources are mass spectra from protein extracts that have not been diluted. From these, dilutions were produced. In all species, the number of peaks is increasing from younger to older stages. In diluted measurements, the number of peaks is reduced, but the spectra do not resemble the younger stages. As evidenced in Figure 5a spectra derived from dilution of C6f are still clearly separated from other stages in ordination space
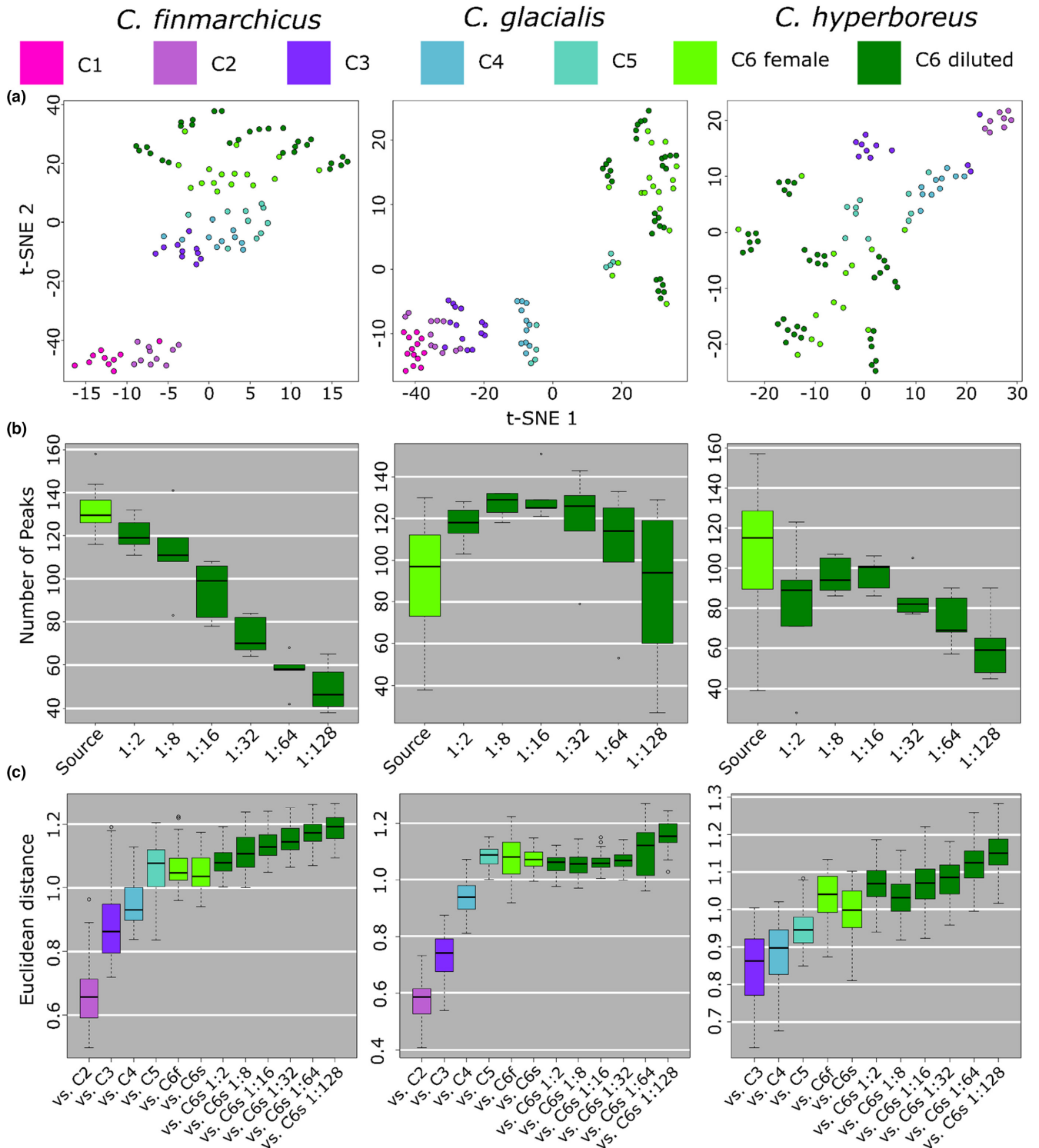
FIGURE 5 (a) tSNE-plots of the dilution experiments with the three *Calanus* species. Displayed in dark green are mass spectra retained from different dilutions of material taken from C6 female specimens. These mass spectra still are more similar to other adult specimens rather than to juvenile stages. (b) Number of peaks of mass spectra from diluted samples. Undiluted samples are displayed light green. Diluted samples are represented in dark green. (c) Euclidean distances of the mass spectra from different stages and dilutions to mass spectra from C1 specimens for *C. finmarchicus* and *C. glacialis* and from C2 individuals for *C. hyperboreus*.

of 148 specimens (82.68%) would be accepted (tp); for 19 specimens (fp and fn) the test would recommend a (morphological) reinvestigation of the specimens. Correct morphological staging of these specimens would then add up the total correct stage classification

of 167 specimens (93.30%). Only 12 (6.70%) misclassifications would remain undiscovered.

Reliable morphological species identification is, specifically in zooplankton, often restricted to adult stages. For application of

proteomic fingerprinting in monitoring or field studies it would be highly beneficial, if species identification of juveniles could be done based on an adult reference library only. For all three *Calanus* species it was possible to identify most juvenile specimens using the adult reference library; however, with lower success regarding the *post hoc* test for the RF classification. Conversely, adult specimens were reliably identified using a juvenile-based library. This, as well as the increasing number of peaks in later developmental stages, implies that major components of the proteomic fingerprint are acquired already in early stages, but some peaks are only found in later developmental stages. Observed differences in the proteome among ontogenetic stages in other calanoid copepod species were, similar to our findings, always smaller than interspecific differences (Kaiser et al., 2018; Laakmann et al., 2013).

Differences among stages were found in proteomic spectra of different arthropod species, albeit to varying degrees. While nymphal and adult stages of the insect Heteroptera *Cimex hirundinis* (Lamarck, 1816) differed distinctly, still several major homologous peaks were detected in both stages with more complexity in adult patterns (Hamlili, Bérenger, et al., 2021). Similar differences between nymphal and adult stages were observed for the acarid *Ixodes ricinus* (Linnaeus, 1758) (Karger et al., 2019). Proteomic spectra of phlebotomine sand fly larvae were distinguishable at the species level and quite stable (Halada et al., 2018), contrasting the distinct differences we observed among copepodite stages in *Calanus*. However, after metamorphosis, spectra of adult sand flies strongly changed with more peaks being expressed. This fundamental change has also been observed for other holometabolous insects, for example, ceratopogonid and culicid larvae, which showed only a small number of identical peaks between adult and larvae (Steinmann et al., 2013). Therefore, larvae were not identified in their approach as the respective adults of the species from the reference library. Overall changes in the proteomic fingerprint observed in *Calanus* species were less pronounced than in these insects. Because juvenile and adult pelagic copepods share the same habitat and do not undergo such drastic morphological changes, a more uniform proteome composition can be expected. Although nauplii stages of *Calanus* were not considered in the present study, Laakmann et al. (2013) demonstrated the stability of species identification based on proteomic spectra even when nauplii were included.

In contrast to genetic markers, proteomic spectra are more variable and influenced by the environment, reflecting physiological responses (Karger et al., 2019). We observed several well-expressed peaks in the later copepodite stages (which remained abundant also in diluted samples) that are completely missing in younger stages, for example, in copepodids C1 and C2. This is most likely a consequence of severe physiological changes with a certain stage. Drastic changes in the expression of genes with changing copepodite stages have been observed for several processes such as lipid biosynthesis (Lenz et al., 2014). This may be also the reason for the observed misclassification between stages. These misclassified specimens may have been in an intermoult condition, in which the specimens would be physiologically advanced, but still showing the exoskeleton of a previous stage. Based on the m/z values only, it remains,

however, pure speculation, which processes are behind the observed patterns. It cannot be ruled out with certainty that external factors could also have caused changes in the composition and patterns of proteomic spectra. To date, there is no information about mass spectra variability of marine species based on their distribution range and environmental conditions. Mass spectra variability due to different ontogenetic stages may also be affected by abiotic or ecological differences such as varying temperatures, salinities or feeding regimes. This may not only be relevant for stage separation but also for the general composition of the library. Caution needs to be taken when applying proteomic fingerprinting for the identification of widely distributed species. Reference libraries may need updates from different investigated regions of the species´ distribution range. Otherwise, similar to the identification of juveniles based on an adult-specimen reference library, problems may occur either with classification or at least with classification supported by a *post hoc* test, because the reference library may not cover the full mass spectra variability of a widely distributed species.

However, if stage specificity of proteomic profiles as observed here is confirmed to be universal also; for example, in controlled experiments, this would be an extreme advantage of MALDI-TOF analyses in monitoring applications. While in a morphological assessment, staging of *Calanus* specimens would be possible, species identification would hardly be possible due to the lack of taxonomic features throughout the younger developmental stages (Laakmann et al., 2020). Thus, a rapid biodiversity assessment using proteome fingerprinting would be superior to a mere morphological or molecular-genetic approach, since the identification of the developmental stage as well as the species could be performed in only one analytical step and provide supplementary information on the recruitment processes and generations of investigated copepod populations. Compared to the MALDI-TOF MS, the alternative approach on morphological staging and genetic barcoding is less cost-effective and more time consuming (Kaiser et al., 2018; Rossel et al., 2019).

Costs of consumables for proteomic fingerprinting range between €0.02 to €0.75 per specimen (depending on amount of matrix used), compared to at least €5.65 for DNA barcoding (Rossel et al., 2019; see the Supporting Information material of the study for a detailed list of the costs). Those calculations did not include labour costs and general laboratory equipment such as thermocyclers for barcoding or the mass spectrometer for proteomic fingerprinting. In a study on expenses for microbiological identifications including labour, costs were summed up to $61.39 for molecular sequencing and $6.44 for MALDI-TOF MS identifications (Tran et al., 2015). The major bottleneck of MALDI-TOF MS is instrument accessibility. Prices are high and therefore instruments are not acquirable for everyone. Still, due to its low costs for consumables it is said that instrument costs offset after a few years of usage (Tran et al., 2015). Costs for application of MALDI-TOF MS are mainly influenced by the amount of HCCA matrix for specimen preparation. Specimens in the present study were rather large, and thus a comparably higher volume of matrix was used, increasing costs. These could be lowered by using smaller body parts and less HCCA matrix. In the present study, conditions for all species

were kept as similar as possible by using the complete prosome body part for measurements. This may not be necessary as smaller parts can be suitable as well. For instance, using only legs of analysed specimens for instance was shown to work well in other species such as deep-sea isopods to differentiate between specimens of a cryptic-species complex (Kürzel et al., 2022; Paulus et al., 2022).

In a further approach it could be investigated, if even more power lies in proteomic fingerprinting for community assessments. Rossel and Martínez Arbizu (2019) reported about the discrimination of adult harpacticoid copepods between sexes in some species. Including this in future assessments would increase the resolution accomplished using mass spectrometry even more.

AUTHOR CONTRIBUTIONS

Sven Rossel carried out MALDI-TOF MS measurements and carried out DNA barcoding. Sven Rossel and Janna Peters analysed the data and drafted the manuscript. Holger Auel and Patricia Kaiser carried out the fieldwork and collected the samples. Patricia Kaiser carried out initial staging and species assignment of specimens. All authors discussed the results and contributed to the final version of the manuscript. All authors significantly contributed to the design of the research.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

MALDI-TOF MS raw data and respective meta data have been stored at the Dryad data repository (https://doi.org/10.5061/dryad.3r228 0gkc). Additionally, R-scripts for processing of data on species and stage level are stored here. This includes a script for parallel computing for classification approaches. COI data can be found in the BOLD project *Calanus* stage identification using proteome fingerprinting (CSIPF) and on GenBank (accession nos ON480204–ON48025).

ORCID

Sven Rossel https://orcid.org/0000-0002-1187-346X
Patricia Kaiser https://orcid.org/0000-0002-9517-3515
Maya Bode-Dalby https://orcid.org/0000-0001-8002-0746
Jasmin Renz https://orcid.org/0000-0002-2658-445X
Silke Laakmann https://orcid.org/0000-0003-3273-7907
Pedro Martínez Arbizu https://orcid.org/0000-0002-0891-1154
Janna Peters https://orcid.org/0000-0002-0941-8257

REFERENCES

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.

Beaugrand, G., Christophe, L., & Martin, E. (2009). Rapid biogeographical plankton shifts in the North Atlantic Ocean. *Global Change Biology*, 15(7), 1790–1803. https://doi.org/10.1111/j.1365-2486.2009.01848.x

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, 41(D1), D36–D42.

Bode, M., Laakmann, S., Kaiser, P., Hagen, W., Auel, H., & Cornils, A. (2017). Unravelling diversity of deep-sea copepods using integrated morphological and molecular techniques. *Journal of Plankton Research*, 39(4), 600–617.

Breimann, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

Brodskii, K., Vyshkvartseva, N., Kos, M., & Markhatseva, E. (1983). Copepods (Copepoda: Calanoida) of the seas of the USSR and adjacent waters. *Keys to the Fauna of the USSR*, 1(135), 358.

Chen, X.-F., Hou, X., Xiao, M., Zhang, L., Cheng, J.-W., Zhou, M.-L., Huang, J.-J., Zhang, J.-J., Xu, Y.-C., & Hsueh, P.-R. (2021). Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) analysis for the identification of pathogenic microorganisms: A review. *Microorganisms*, 9(7), 1536.

Cheng, F., Wang, M., Sun, S., Li, C., & Zhang, Y. (2013). DNA barcoding of Antarctic marine zooplankton for species identification and recognition. *Advances in Polar Science*, 24(2), 119–127.

Choquet, M., Hatlebakk, M., Dhanasiri, A. K. S., Kosobokova, K., Smolina, I., Søreide, J. E., Svensen, C., Melle, W., Kwaśniewski, S., Eiane, K., Daase, M., Tverberg, V., Skreslet, S., Bucklin, A., & Hoarau, G. (2017). Genetics redraws pelagic biogeography of *Calanus*. *Biology Letters*, 13(12), 20170588. https://doi.org/10.1098/rsbl.2017.0588

Choquet, M., Kosobokova, K., Kwaśniewski, S., Hatlebakk, M., Dhanasiri, A. K. S., Melle, W., Daase, M., Svensen, C., Søreide, J. E., & 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. https://doi.org/10.1002/lom3.10240

Dieme, C., Yssouf, A., Vega-Rúa, A., Berenger, J.-M., Failloux, A.-B., Raoult, D., Parola, P., & Almeras, L. (2014). Accurate identification of Culicidae at aquatic developmental stages by MALDI-TOF MS profiling. *Parasites & Vectors*, 7(1), 544.

Folmer, O., Black, M. B., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3(5), 294–299.

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

Gibb, S. (2015). *MALDIquantForeign: Import/Export routines for MALDIquant*. A package for R. https://CRAN.R-project.org/package=MALDIquantForeign

Gibb, S., & Strimmer, K. (2012). MALDIquant: A versatile R package for the analysis of mass spectrometry data. *Bioinformatics*, 28(17), 2270–2271.

Halada, P., Hlavackova, K., Dvorak, V., & Volf, P. (2018). Identification of immature stages of phlebotomine sand flies using MALDI-TOF MS and mapping of mass spectra during sand fly life cycle. *Insect Biochemistry and Molecular Biology*, *93*, 47–56. https://doi.org/10.1016/j.ibmb.2017.12.005

Hamlili, F. Z., Bérenger, J.-M., Diarra, A. Z., & Parola, P. (2021). Molecular and MALDI-TOF MS identification of swallow bugs *Cimex hirundinis* (Heteroptera: Cimicidae) and endosymbionts in France. *Parasites & Vectors*, *14*(1), 587. https://doi.org/10.1186/s13071-021-05073-x

Hamlili, F. Z., Thiam, F., Laroche, M., Diarra, A. Z., Doucouré, S., Gaye, P. M., Fall, C. B., Faye, B., Sokhna, C., Sow, D., & Parola, P. (2021). MALDI-TOF mass spectrometry for the identification of freshwater snails from Senegal, including intermediate hosts of schistosomes. *PLoS Neglected Tropical Diseases*, *15*(9), e0009725. https://doi.org/10.1371/journal.pntd.0009725

Han, H., Guo, X., & Yu, H. (2016). Variable selection using mean decrease accuracy and mean decrease gini based on random forest. *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 219–224.

Hasnaoui, B., Diarra, A. Z., Berenger, J.-M., Medkour, H., Benakhla, A., Mediannikov, O., & Parola, P. (2022). Use of the proteomic tool MALDI-TOF MS in termite identification. *Scientific Reports*, *12*(1), 718. https://doi.org/10.1038/s41598-021-04574-0

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1512), 313–321.

Hill, R., Allen, L., & Bucklin, A. (2001). Multiplexed species-specific PCR protocol to discriminate four N. Atlantic *Calanus* species, with an mtCOI gene tree for ten *Calanus* species. *Marine Biology*, *139*(2), 279–287. https://doi.org/10.1007/s002270100548

Holst, S., Heins, A., & Laakmann, S. (2019). Morphological and molecular diagnostic species characters of Staurozoa (Cnidaria) collected on the coast of Helgoland (German Bight, North Sea). *Marine Biodiversity*, *49*, 1775–1797. https://doi.org/10.1007/s12526-019-00943-1

Hynek, R., Kuckova, S., Cejnar, P., Junková, P., Přikryl, I., & Říhová Ambrožová, J. (2018). Identification of freshwater zooplankton species using protein profiling and principal component analysis. *Limnology and Oceanography: Methods*, *16*(3), 199–204. https://doi.org/10.1002/lom3.10238

Kaiser, P., Bode, M., Cornils, A., Hagen, W., Martínez Arbizu, P., Auel, H., & Laakmann, S. (2018). High-resolution community analysis of deep-sea copepods using MALDI-TOF protein fingerprinting. *Deep-Sea Research Part I: Oceanographic Research Papers*, *138*, 122–130.

Karger, A., Bettin, B., Gethmann, J. M., & Klaus, C. (2019). Whole animal matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry of ticks – Are spectra of *Ixodes ricinus* nymphs influenced by environmental, spatial, and temporal factors? *PLoS One*, *14*(1), e0210590. https://doi.org/10.1371/journal.pone.0210590

Korfhage, S. A., Rossel, S., Brix, S., McFadden, C. S., Ólafsdóttir, S. H., & Martínez Arbizu, P. (2022). Species delimitation of Hexacorallia and Octocorallia around Iceland using nuclear and mitochondrial DNA and proteome fingerprinting. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.838201

Krijthe, J. (2015). *R wrapper for Van der Maaten's Barnes-Hut implementation of t-distributed stochastic neighbor embedding* [C++]. https://github.com/jkrijthe/Rtsne

Kürzel, K., Kaiser, S., Lörz, A.-N., Rossel, S., Paulus, E., Peters, J., Schwentner, M., Martínez Arbizu, P., Coleman, C. O., Svavarsson, J., & Brix, S. (2022). Correct species identification and its implications for conservation using Haploniscidae (Crustacea, Isopoda) in Icelandic waters as a proxy. *Frontiers in Marine Science*, *8*, 795196. https://doi.org/10.3389/fmars.2021.795196

Kwasniewski, S., Gluchowska, M., Jakubas, D., Wojczulanis-Jakubas, K., Walkusz, W., Karnovsky, N., Blachowiak-Samolyk, K., Cisek, M., & Stempniewicz, L. (2010). The impact of different hydrographic conditions and zooplankton communities on provisioning Little Auks along the West coast of Spitsbergen. *Progress in Oceanography*, *87*(1), 72–82. https://doi.org/10.1016/j.pocean.2010.06.004

Laakmann, S., Blanco-Bercial, L., & Cornils, A. (2020). The crossover from microscopy to genes in marine diversity: From species to assemblages in marine pelagic copepods. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1814), 20190446. https://doi.org/10.1098/rstb.2019.0446

Laakmann, S., Gerdts, G., Erler, R., Knebelsberger, T., Martínez Arbizu, P., & Raupach, M. J. (2013). Comparison of molecular species identification for North Sea calanoid copepods (Crustacea) using proteome fingerprints and DNA sequences. *Molecular Ecology Resources*, *13*(5), 862–876. https://doi.org/10.1111/1755-0998.12139

Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, *129*(2), 271–280.

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. https://doi.org/10.1371/journal.pone.0088589

Liaw, A., & Wiener, M. (2002). Classification and regression by random-Forest. *R News*, *2*(3), 18–22.

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

Loaiza, J. R., Almanza, A., Rojas, J. C., Mejia, L., Cervantes, N. D., Sanchez-Galan, J. E., Merchán, F., Grillet, A., Miller, M. J., De León, L. F., & Gittens, R. A. (2019). Application of matrix-assisted laser desorption/ionization mass spectrometry to identify species of Neotropical *Anopheles* vectors of malaria. *Malaria Journal*, *18*(1), 95.

Maász, G., Takács, P., Boda, P., Várbíró, G., & Pirger, Z. (2017). Mayfly and fish species identification and sex determination in bleak (*Alburnus alburnus*) by MALDI-TOF mass spectrometry. *Science of the Total Environment*, *601*, 317–325.

Martínez Arbizu, P. (2022). *PairwiseAdonis* [R]. https://github.com/pmartinezarbizu/pairwiseAdonis

Mathis, A., Depaquit, J., Dvovrák, V., Tuten, H., Bañuls, A.-L., Halada, P., Zapata, S., Lehrter, V., Hlavavckóvá, K., Prudhomme, J., Volf, P., Sereno, D., Kaufmann, C., Pflüger, V., & Schaffner, F. (2015). Identification of phlebotomine sand flies using one MALDI-TOF MS reference database and two mass spectrometer systems. *Parasites & Vectors*, *8*(1), 266.

Mazzeo, M. F., Giulio, B. D., Guerriero, G., Ciarcia, G., Malorni, A., Russo, G. L., & Siciliano, R. A. (2008). Fish authentication by MALDI-TOF mass spectrometry. *Journal of Agricultural and Food Chemistry*, *56*(23), 11071–11076.

Nabet, C., Kone, A. K., Dia, A. K., Sylla, M., Gautier, M., Yattara, M., Thera, M. A., Faye, O., Braack, L., Manguin, S., Beavogui, A. H., Doumbo, O., Gay, F., & Piarroux, R. (2021). New assessment of anopheles vector species identification using MALDI-TOF MS. *Malaria Journal*, *20*(1), 1–16.

Nebbak, A., El Hamzaoui, B., Berenger, J.-M., Bitam, I., Raoult, D., Almeras, L., & Parola, P. (2017). Comparative analysis of storage conditions and homogenization methods for tick and flea species for identification by MALDI-TOF MS. *Medical and Veterinary Entomology*, *31*(4), 438–448.

Papagiannopoulou, C., Parchen, R., Rubbens, P., & Waegeman, W. (2020). Fast pathogen identification using single-cell matrix-assisted laser desorption/ionization-aerosol time-of-flight mass spectrometry data and deep learning methods. *Analytical Chemistry*, *92*(11), 7523–7531.

Paulus, E., Brix, S., Siebert, A., Martínez Arbizu, P., Rossel, S., Peters, J., Svavarsson, J., & Schwentner, M. (2022). Recent speciation and hybridization in Icelandic deep-sea isopods: An integrative approach

using genomics and proteomics. *Molecular Ecology*, 31(1), 313–330. https://doi.org/10.1111/mec.16234

R Core Team. (2022). *R: A language and environment for statistical computing (4.1.0)* [computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The barcode of life data system (http://www.barcodinglife.org). *Molecular Ecology Resources*, 7(3), 355–364.

Riccardi, N., Lucini, L., Benagli, C., Welker, M., Wicht, B., & Tonolla, M. (2012). Potential of matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) for the identification of freshwater zooplankton: A pilot study with three *Eudiaptomus* (Copepoda: Diaptomidae) species. *Journal of Plankton Research*, 34(6), 484–492.

Rossel, S., Barco, A., Kloppmann, M., Martínez Arbizu, P., Huwer, B., & Knebelsberger, T. (2020). Rapid species level identification of fish eggs by proteome fingerprinting using MALDI-TOF MS. *Journal of Proteomics*, 231, 103993.

Rossel, S., Khodami, S., & Martínez Arbizu, P. (2019). Comparison of rapid biodiversity assessment of meiobenthos using MALDI-TOF MS and metabarcoding. *Frontiers in Marine Science*, 6, 659. https://doi.org/10.3389/fmars.2019.00659

Rossel, S., & Martínez Arbizu, P. (2018a). Automatic specimen identification of Harpacticoids (Crustacea:Copepoda) using Random Forest and MALDI-TOF mass spectra, including a *post hoc* test for false positive discovery. *Methods in Ecology and Evolution*, 9(6), 1421–1434. https://doi.org/10.1111/2041-210X.13000

Rossel, S., & Martínez Arbizu, P. (2018b). Effects of sample fixation on specimen identification in biodiversity assemblies based on proteomic data (MALDI-TOF). *Frontiers in Marine Science*, 5, 149. https://doi.org/10.3389/fmars.2018.00149

Rossel, S., & Martínez Arbizu, P. (2019). Revealing higher than expected diversity of Harpacticoida (Crustacea: Copepoda) in the North Sea using MALDI-TOF MS and molecular barcoding. *Scientific Reports*, 9(1), 9182. https://doi.org/10.1038/s41598-019-45718-7

Ryan, C., Clayton, E., Griffin, W., Sie, S., & Cousens, D. (1988). SNIP, a statistics-sensitive background treatment for the quantitative analysis of PIXE spectra in geoscience applications. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 34(3), 396–402.

Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.

Singhal, N., Kumar, M., Kanaujia, P. K., & Virdi, J. S. (2015). MALDI-TOF mass spectrometry: An emerging technology for microbial identification and diagnosis. *Frontiers in Microbiology*, 6(791). https://doi.org/10.3389/fmicb.2015.00791

Steinmann, I. C., Pflüger, V., Schaffner, F., Mathis, A., & Kaufmann, C. (2013). Evaluation of matrix-assisted laser desorption/ionization time of flight mass spectrometry for the identification of ceratopogonid and culicid larvae. *Parasitology*, 140(3), 318–327.

Tan, K. E., Ellis, B. C., Lee, R., Stamper, P. D., Zhang, S. X., & Carroll, K. C. (2012). Prospective evaluation of a matrix-assisted laser desorption ionization–time of flight mass spectrometry system in a hospital clinical microbiology laboratory for identification of bacteria and yeasts: A bench-by-bench study for assessing the impact on time to identification and cost-effectiveness. *Journal of Clinical Microbiology*, 50(10), 3301–3308. https://doi.org/10.1128/JCM.01405-12

Tarling, G. A., Freer, J. J., Banas, N. S., Belcher, A., Blackwell, M., Castellani, C., Cook, K. B., Cottier, F. R., Daase, M., Johnson, M. L., Last, K. S., Lindeque, P. K., Mayor, D. J., Mitchell, E., Parry, H. E., Speirs, D. C., Stowasser, G., & Wootton, M. (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(2), 333–344. https://doi.org/10.1007/s13280-021-01667-y

Tran, A., Alby, K., Kerr, A., Jones, M., & Gilligan, P. H. (2015). Cost savings realized by implementation of routine microbiological identification by matrix-assisted laser desorption ionization–time of flight mass spectrometry. *Journal of Clinical Microbiology*, 53(8), 2473–2479. https://doi.org/10.1128/JCM.00833-15

Trudnowska, E., Balazy, K., Stoń-Egiert, J., Smolina, I., Brown, T., & Gluchowska, M. (2020). In a comfort zone and beyond—Ecological plasticity of key marine mediators. *Ecology and Evolution*, 10(24), 14067–14081. https://doi.org/10.1002/ece3.6997

Trudnowska, E., Stemmann, L., Błachowiak-Samołyk, K., & Kwasniewski, S. (2020). Taxonomic and size structures of zooplankton communities in the fjords along the Atlantic water passage to the Arctic. *Journal of Marine Systems*, 204, 103306. https://doi.org/10.1016/j.jmarsys.2020.103306

Van Driessche, L., Bokma, J., Deprez, P., Haesebrouck, F., Boyen, F., & Pardon, B. (2019). Rapid identification of respiratory bacterial pathogens from bronchoalveolar lavage fluid in cattle by MALDI-TOF MS. *Scientific Reports*, 9(1), 18381. https://doi.org/10.1038/s41598-019-54599-9

Volta, P., Riccardi, N., Lauceri, R., & Tonolla, M. (2012). Discrimination of freshwater fish species by Matrix-Assisted Laser Desorption/Ionization-Time Of Flight Mass Spectrometry (MALDI-TOF MS): A pilot study. *Journal of Limnology*, 71(1), e17.

Weydmann, A., Carstensen, J., Goszczko, I., Dmoch, K., Olszewska, A., & Kwasniewski, S. (2014). Shift towards the dominance of boreal species in the Arctic: Inter-annual and spatial zooplankton variability in the West Spitsbergen current. *Marine Ecology Progress Series*, 501, 41–52. https://doi.org/10.3354/meps10694

Weydmann, A., Coelho, N. C., Ramos, A. A., Serrão, E. A., & Pearson, G. A. (2014). Microsatellite markers for the Arctic copepod *Calanus glacialis* and cross-amplification with *C. finmarchicus*. *Conservation Genetics Resources*, 6(4), 1003–1005. https://doi.org/10.1007/s12686-014-0269-6

Weydmann, A., & Kwasniewski, S. (2008). Distribution of *Calanus* populations in a glaciated fjord in the Arctic (Hornsund, Spitsbergen)—The interplay between biological and physical factors. *Polar Biology*, 31(9), 1023–1035. https://doi.org/10.1007/s00300-008-0441-0

Weydmann, A., Przyłucka, A., Lubośny, M., Walczyńska, K. S., Serrão, E. A., Pearson, G. A., & Burzyński, A. (2017). Mitochondrial genomes of the key zooplankton copepods Arctic *Calanus glacialis* and North Atlantic *Calanus finmarchicus* with the longest crustacean non-coding regions. *Scientific Reports*, 7(1), 13702. https://doi.org/10.1038/s41598-017-13807-0

Wilke, T., Renz, J., Hauffe, T., Delicado, D., & Peters, J. (2020). Proteomic fingerprinting discriminates cryptic gastropod species. *Malacologia*, 63(1), 131–137.

Yssouf, A., Parola, P., Lindström, A., Lilja, T., L'Ambert, G., Bondesson, U., Berenger, J.-M., Raoult, D., & Almeras, L. (2014). Identification of European mosquito species by MALDI-TOF MS. *Parasitology Research*, 113(6), 2375–2378. https://doi.org/10.1007/s00436-014-3876-y