Potential of MALDI-TOF MS-based proteomic fingerprinting for species identification of Cnidaria across classes, species, regions and developmental stages

Sven Rossel[1], Janna Peters[2], Silke Laakmann[3], Pedro Martinez[1], and Sabine Holst[1]

[1]Senckenberg am Meer Deutsches Zentrum fur Marine Biodiversitatsforschung
[2]Senckenberg am Meer Deutsches Zentrum für Marine Biodiversitätsforschung
[3]Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg

March 11, 2023

Abstract

Morphological identification of cnidarian species can be difficult throughout all life stages due to the lack of distinct morphological characters. Moreover, in some cnidarian taxa genetic markers are not fully informative, and in these cases combinations of different markers or additional morphological verifications may be required. Proteomic fingerprinting based on MALDI-TOF mass spectra was previously shown to provide reliable species identification in different metazoans including some cnidarian taxa. For the first time, we tested the method across four cnidarian classes (Staurozoa, Scyphozoa, Anthozoa, Hydrozoa) and included different scyphozoan life-history stages (polyp, ephyra, medusa) into our dataset. Our results revealed reliable species identification based on MALDI-TOF mass spectra across all taxa with species-specific clusters for all 23 analyzed species. In addition, proteomic fingerprinting was successful for distinguishing developmental stages, still by retaining a species specific signal. Furthermore, we identified the impact of different salinities in different regions (North Sea and Baltic Sea) on proteomic fingerprints to be negligible. In conclusion, the effects of environmental factors and developmental stages on proteomic fingerprints seem to be low in cnidarians. This would allow using reference libraries built up entirely of adult or cultured cnidarian specimens for the identification of their juvenile stages or specimens from different geographic regions in future biodiversity assessment studies.

Potential of MALDI-TOF MS-based proteomic fingerprinting for species identification of Cnidaria across classes, species, regions and developmental stages

Sven Rossel*[1], Janna Peters[2], Silke Laakmann[3,4], Pedro Martínez Arbizu[1]& Sabine Holst[2]

[1] Senckenberg am Meer, German Centre for Marine Biodiversity Research (DZMB), 26382 Wilhelmshaven, Germany

[2] Senckenberg am Meer, German Centre for Marine Biodiversity Research (DZMB), 20146 Hamburg, Germany

[3]Helmholtz Institute for Functional Marine Biodiversity at the University of Oldenburg (HIFMB), 26129 Oldenburg, Germany

[4]Alfred Wegener Institute, Helmholtz-Centre for Polar and Marine Research (AWI), 27570 Bremerhaven, Germany

Corresponding author email: sven.rossel@senckenberg.de

Abstract

Morphological identification of cnidarian species can be difficult throughout all life stages due to the lack of distinct morphological characters. Moreover, in some cnidarian taxa genetic markers are not fully informative, and in these cases combinations of different markers or additional morphological verifications may be required. Proteomic fingerprinting based on MALDI-TOF mass spectra was previously shown to provide reliable species identification in different metazoans including some cnidarian taxa. For the first time, we tested the method across four cnidarian classes (Staurozoa, Scyphozoa, Anthozoa, Hydrozoa) and included different scyphozoan life-history stages (polyp, ephyra, medusa) into our dataset. Our results revealed reliable species identification based on MALDI-TOF mass spectra across all taxa with species-specific clusters for all 23 analyzed species. In addition, proteomic fingerprinting was successful for distinguishing developmental stages, still by retaining a species specific signal. Furthermore, we identified the impact of different salinities in different regions (North Sea and Baltic Sea) on proteomic fingerprints to be negligible. In conclusion, the effects of environmental factors and developmental stages on proteomic fingerprints seem to be low in cnidarians. This would allow using reference libraries built up entirely of adult or cultured cnidarian specimens for the identification of their juvenile stages or specimens from different geographic regions in future biodiversity assessment studies.

Introduction

Cnidaria is a highly diverse phylum currently including six accepted classes (WoRMS; 2022). Two classes, Anthozoa and Staurozoa, comprise benthic species (Mills et al., 2007; Miranda et al., 2018). Three classes (Scyphozoa, Hydrozoa, Cubozoa) include species with metagenetic life cycles, usually with a sexually reproducing planktonic medusa generation and an asexually reproducing benthic polyp generation, however, there are many exceptions from these reproduction modes (Bouillon et al., 2006; Jarms and Morandini, 2019). Another class (Myxozoa) are endoparasitic species diverged from free-living cnidarian ancestors (Okamura et al., 2015). The free-living species of the first five classes are distributed in marine ecosystems across all depths and latitudes (Mills et al., 2007; Rodríguez et al., 2014; Miranda et al., 2018; Jarms and Morandini, 2019). However, distribution and ecology of cnidarians are largely unexplored and cnidarian species are often neglected in biodiversity studies because of difficulties in their correct species identification (Häussermann, 2004; Martell et al., 2022). In particular, juvenile stages often lack diagnostic characters making identification to the species level impossible in many cases (Holst, 2012; Schuchert, 2012). However, morphological species identification is challenging because of lacking distinct morphological features and high morphological variation in many cnidarians (Brugler et al., 2018; Pruski and Miglietta, 2019; Lawley et al., 2021). Moreover, fixation of the gelatinous medusae and the soft-bodied polyps can lead to shrinkage and deformation of fragile specimens leading to distortion of morphological diagnostic features (Häussermann, 2004; Schuchert, 2012; Holst and Laakmann, 2014; Holst et al., 2019).

Molecular genetic techniques have been widely tested and applied for identification and differentiation of cnidarian species (Holst and Laakmann, 2014; Moura et al., 2018; Holst et al., 2019; Bucklin et al., 2021). However, not in all groups, standard fragments such as the COI barcode regions are fully informative on species level (Brugler et al., 2018; Schuchert, 2020). In some taxa, a combination of several molecular genetic markers achieves a satisfying resolution (McFadden et al., 2011, 2014). As an alternative to the expensive and time-consuming molecular genetic analyses, proteomic fingerprinting was successfully used for species identification in Cnidaria (Holst et al., 2019; Park et al., 2021; Korfhage et al., 2022). Using matrix-assisted laser desorption/ionization–time of flight (MALDI-TOF) mass spectrometry, a subset of the specimen's peptides and proteins are measured resulting in a mass spectrum that can be used for differentiation of species. The method is already widely established for routine pathogen identification (Chen et al., 2021) and is more recently also used in pathogen-vector identification, of mainly insects and mites (Dieme et al., 2014; Lafri et al., 2016; Hamlili et al., 2021). The low costs and short hand-on times per specimen (Tran et al., 2015; Rossel et al., 2019) coupled with high identification success makes it a promising tool for rapid and reliable species identification also in biodiversity assessments. In marine science it was applied for identification of a variety of animal groups such as copepods (Laakmann et al., 2013; Bode et al., 2017; Kaiser et al., 2018; Rossel and Martínez Arbizu, 2019), isopods (Kürzel et al., 2022; Paulus et al., 2022) and fish (Mazzeo et al., 2008; Rossel et al., 2020). Recently, unsupervised methods for rapid delimitation in biodiversity assessments based on MALDI-TOF MS data were devised (Rossel and Martínez Arbizu, 2020; Renz et al., 2021) and

applied in investigations of deep-sea biodiversity (Rossel et al., 20232b). Further studies have shown that the resolution of proteomic fingerprinting can go beyond mere species identification such das differentiation of some species on sex level (Rossel and Martínez Arbizu, 2019), developmental stage (Rossel et al., 2022a) and host species of important pathogen-vector species (Niare et al., 2016, 2018).

This makes MALDI-TOF MS a promising tool for species identification in Cnidaria but it has only been applied on selected cnidarian taxa so far including benthic stalked jellyfish (Staurozoa) (Holst et al., 2019), siphonophores (colonial pelagic Hydrozoa) (Park et al., 2021), and cold water corals (Anthozoa) (Korfhage et al., 2022). However, MALDI-TOF MS has never been tested for species identification by using the gelatinous tissues of scyphozoan and hydrozoan medusae before. In addition to these previous studies applying MALDI-TOF MS on selected cnidarian taxa, the present study aims to investigate mass spectra variability across various taxa using the so far largest data set of cnidarians including four classes (Scyphozoa, Hydrozoa, Staurozoa and Anthozoa). Beyond species classification, we furthermore investigate the influence of factors such as developmental stages and spatial origin on classification success at species level to evaluate the applicability of the method in future classification approaches.

Material and Methods

Samples

In total, 278 specimens of Cnidaria belonging to 23 different species from four classes were analyzed (Table 1). For four species (*Aurelia aurita* Linnaeus, 1758, *Chrysaora hysoscella*(Linnaeus, 1766), *Cyanea capillata* (Linnaeus, 1758) and *C. lamarckii* Péron & Lesueur, 1810) specimens of different ontogenetic stages (polyp, ephyra and medusa) were included (Table 1). Specimens were either sampled in the field or obtained from cultures (see supplementary table 1). Raw data of measurements from staurozoan species were obtained from (Holst et al., 2019). Field specimens were morphologically identified to species level by taxonomic experts immediately after collection, before complete specimens or subsamples were preserved in undenatured ethanol (80 - 96%). Cultures of scyphozoan polyps were reared from planulae which were obtained from sampled mature medusae; ephyrae from laboratory cultures were produced by strobilating polyps (supplement table 2).

Table 1: Number, developmental stage, and sampling region of analyzed specimens of species from four cnidarian classes. **NS** = North Sea, **BS** = Baltic Sea.

Species	**Class**	**n specimens**	**n polyps (NS/BS)**	**n ephyrae (NS/BS**
Adamsia palliata (Fabricius, 1779)	Anthozoa	15	15 (15/0)	-
Alcyonium digitatum Linnaeus, 1758	Anthozoa	6	6 (6/0)	-
Epizoanthus incrustatus Düben & Koren, 1847	Anthozoa	8	8 (8/0)	-
Funiculina quadrangularis (Pallas, 1766)	Anthozoa	4	4 (4/0)	-
Hormathia digitata (O.F.Müller, 1776)	Anthozoa	13	13 (13/0)	-
Metridium senile (Linnaeus, 1761)	Anthozoa	3	3 (3/0)	-
Pennatula phosphorea Linnaeus, 1758	Anthozoa	7	7 (7/0)	-
Urticina felina Linnaeus, 1767	Anthozoa	4	4 (4/0)	-
Aglantha digitale O. F. Müller, 1766	Hydrozoa	3	NA	-
Corymorpha nutans M. Sars, 1835	Hydrozoa	2	NA	-
Eirene viridula (Péron & Lesueur, 1810)	Hydrozoa	2	NA	-
Eucheilota maculata Hartlaub, 1894	Hydrozoa	3	NA	-
Hydractinia echinata Fleming, 1823	Hydrozoa	4	NA	-
Leuckartiara octona (Fleming, 1823)	Hydrozoa	3	NA	-
Rathkea octopunctata Russell, 1953	Hydrozoa	3	NA	-
Aurelia aurita Linnaeus, 1758	Scyphozoa	66	6 (6/0)	42 (19/13)
Chrysaora hysoscella (Linnaeus, 1766)	Scyphozoa	22	2 (2/0)	10 (10/0)
Cyanea capillata (Linnaeus, 1758)	Scyphozoa	36	5 (0/5)	23 (10/13)
Cyanea lamarckii Péron & Lesueur, 1810	Scyphozoa	45	5 (5/0)	26 (26/0)

Rhizostoma octopus Mayer, 1910	Scyphozoa	11	0 (0/0)	0 (0/0)
Craterolophus convolvulus (Johnston, 1835)	Staurozoa	12	NA	-
Haliclystus auricula James-Clark, 1863	Staurozoa	3	NA	-
Haliclystus tenuis Kishinouye, 1910	Staurozoa	3	NA	-
n = 23	n = 4	n = 284		

MALDI-TOF MS measurements

From each specimen, a small tissue fragment (max. 1 mm^3) was incubated for 5 minutes with 5 µl of alpha-cyano-4-hydroxycinnamic acid (HCCA) matrix. Of this incubated solution, 1 to 1.5 µl were transferred to a target plate on one to nine spots for co-crystallization of matrix and analytes. Each spot was measured one to three times 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 analyzing the mass peak range from 2 to 20 kDa. Furthermore, 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 ten times above the threshold. To create a sum spectrum, a total of at least 120 laser shots were applied to a spot. Measurements were carried out using the same instrument at different occasions between 2013 and 2019.

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 and Strimmer, 2012). Spectra were square-root transformed, smoothed using the Savitzky Golay method (Savitzky and 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 and Gallagher, 2001) and used for further analyses.

Hierarchical clustering was carried out in R using average linkage and Euclidean distances and visualized as a circular dendrogram using the R-package dendextend, Version 1.15.1 (Galili, 2015). Random Forest (RF) (Breiman, 2001) was carried out using the R package randomForest, Version 4.6.14 (Liaw and Wiener, 2002). Settings were used according to (Rossel and Martínez Arbizu, 2018) (ntree=2,000, mtry=35, sampsize=number of specimens in the smallest class). Classifications were tested using the RF *post-hoc* test (Rossel and Martínez Arbizu, 2018), 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 ($p < 0.05$) 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). RF classification was applied to identification of the different scyphozoan developmental stages to species level excluding all specimens from the respective stage from the RF model. Also, classification of both juvenile stages (polyp and ephyra) was tested with only adult specimens (medusae) retained in the RF model. All RF models for classification always contained all 23 species included in this study.

RF models, using developmental stages as classes, were 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 Neighbor Embedding (t-SNE) plots based on RF model-votes were created using R package Rtsne Version 0.15 (Krijthe, 2015) with the following settings: perplexity=10,

max_iter=4,000 and theta=0. Principal Coordinate Analysis was applied to the Hellinger-transformed data using the R package ape, Version 5.5 (Paradis and Schliep, 2019).

Results

The hierarchical clustering carried out on the complete dataset of 278 specimens resulted in distinct, species-specific clusters for all 23 analyzed species (Fig.1) irrespective of sample storage time (<1 to 111 months) and measuring campaign (supplementary table 1). A RF model based on all analyzed species resulted in an OOB error of 0 thus supporting species-specificity of mass spectra for all analyzed species. Class-specific clusters (i.e. Anthozoa, Hydrozoa, Scyphozoa and Staurozoa) have only been recovered for Scyphozoa (Fig. 1, branch colors). Clustering of congeneric species was found in*Haliclystus* but not in *Cyanea* .

Interspecific Euclidean distances ranged from 0.99 to 1.38 (Fig. 2A), while intraspecific differences reached a maximum of 1.35 in the scyphozoan *A. aurita* when comparing a North Sea polyp and a North Sea medusa. Lowest interspecific distances were recorded between two scyphozoan species, a *C. hysoscella* medusa and an *A. aurita* ephyra.

Although the factor species explained the majority of variance found in the data, also different stages played an important role. According to the tested factors (stage, region, number of peaks) the stages in a PERMANOVA (Anderson, 2001) indicate a major percentage of variance explained (17.0% in *A. aurita* , 27.1% in *C. capillata* , 22.3% in *C. lamarckii,* Supplementary Table 3 – 5). Thus, groups according to developmental stages can already be recognized in the Hellinger-transformed processed raw data as presented in PCoAs for*A. aurita* (Fig. 3a), *C. capillata* (Fig. 3B) and *C. lamarckii* (Fig. 3C). However, in some cases stages seem to be highly similar for example for some ephyrae and polyps in *C. lamarckii*(Fig. 3C). This is also shown in a tSNE plot based on class votes from a RF classification model trained using species-stages as classes (Fig. 3C, tSNE plot). Still the RF OOB error for all species on stage level was 0 and most stages in all species are clearly separated in tSNE plots (Fig. 3, tSNE plots). Accordingly, Euclidean distances within the different stages are in the range of intra-specific distances (Fig. 2B). Nevertheless, distances between stages within a species are distinctly higher, but still lower than average inter-specific differences (Fig. 2B) allowing differentiation of stages also in classification approaches.

Upon a more detailed investigation, distinct important peaks, identified by Gini index within a RF model, can be found differentiated between stages (Fig. 3, heat maps depicting peak intensities). Between some stages, several peaks differ not only by the relative intensities but by their presence and absence. In *A. aurita* the peak m/z 3447 is absent in all polyp specimens but present in almost all other specimens except for three medusae (Fig. 3A). Other peaks such as m/z 5352 differ between stages mainly by differences in relative intensities, which are in this case higher in polyps compared to other stages. This is also the case for the peak m/z 4581 in *C. capillata* (Fig. 3B) which is present in all specimens but clearly differs in relative intensities between stages. Other peaks such as m/z 3314 are mainly present in medusae but largely absent in the other stages. Even though in *C. lamarckii* ephyrae and polyps are very similar in general, several peaks are completely absent in the polyps but widely present in the ephyrae and most intense in the medusae (m/z 2251, 2797 and 4733).

Tests on species identification of selected stages previously excluded from the reference library resulted in 100 % successes (Fig. 4) with only one exception: one *C. lamarckii* medusa was misclassified as another species (*H. digita*). All classifications were tested using a *post-hoc* test. Whereas the majority of specimens were correctly classified, some classifications were recognized as false positives. In *A. aurita,* of the analyzed six polyps two were recognized as false positives which may reflect that intraspecific distances within this species were on average the highest between polyps and the other stages. In C. *lamarckii* the lowest classification and *post-hoc* test success was recorded in the medusae. Again, these show the highest intraspecific distances to the remaining stages in this species.

The analysis of mass spectra variability in ephyrae of the two scyphozoan species *A. aurita* and *C. capillata* from NS and BS origins revealed that intraspecific distances within one species from a certain region are not distinctly different from intraspecific differences in specimens from different regions (Fig. 2A). The Euclidean

distances within and between regions were on average lower than the average distances between the different developmental stages (Fig. 2B), demonstrating that the influence of the factor stage was higher than the influence of the factor region. The R2 values of the PERMANOVA also indicate a lower percentage of variance explained by regions in *A. aurita* (4.2%, Supplementary Table 3) than in *C. capillata*(10.7%, Supplementary Table 5).

Discussion

Our study validates proteome fingerprinting as a promising tool for identification of species across different classes of Cnidaria. The success of the current study goes alongside other studies displaying the high validity of this method for classification of metazoan taxa across a variety of animal groups such as a collection of Arthropoda (Laakmann et al., 2013; Rossel and Martínez Arbizu, 2019; Nabet et al., 2021; Kürzel et al., 2022; Paulus et al., 2022), Mollusca (Wilke et al., 2020), and Vertebrates (Mazzeo et al., 2008; Mazzeo and Siciliano, 2016; Rossel et al., 2020) from marine, limnic and terrestrial realms. The general applicability of MALDI-TOF MS for differentiation of cnidarian species was shown before on three staurozoans from the North Sea (Holst et al., 2019) and on nine siphonophores of the family Diphyidae (Park et al., 2021). It was furthermore used in an integrative approach for the differentiation of notoriously difficult to identify Hexa- and Octocorallian species showing tendencies to delimit a species complex that could not be resolved using the investigated molecular genetic markers (Korfhage et al., 2022). In the present study, we confirmed that MALDI-TOF MS is also applicable for the fragile gelatinous tissues of hydrozoan and scyphozoan medusae with water contents of > 95 % (Arai, 1997). In addition, the method was tested using species from a wider range of classes in one analysis for the first time. Also, the effect of ontogenetic stage and environmental conditions on proteomic spectra and in turn their impact on reliable species identification in this taxon by MALDI-TOF MS was largely unknown so far.

Average intra- and interspecific distances were clearly different as was also found in studies on different groups of crustaceans (Renz et al., 2021; Paulus et al., 2022). Our results demonstrate that intraspecific variability in scyphozoans was strongly affected by the ontogenetic stage. It is possible to differentiate within species on the level of ontogenetic stage which was previously shown on calanoid copepods (Rossel et al., 2022a). Although the benthic polyp and the pelagic ephyra / medusa stages in metagenetic cnidarians represent two generations with very different morphologies, the different stages formed clear species clusters. However, stage-specific investigations in our study were limited to the class Scyphozoa, and consequently, future studies should also include polyp stages of metagenetic hydrozoan species.

Stage identification by MALDI-TOF MS in cnidarians is mainly of interest if tissue samples of unknown stage (origin) are analyzed since the identification of a certain stage in the life cycle of metagenetic cnidarians is less challenging than the identification of ontogenetic stages in other marine invertebrates as for example in copepods. Still, knowledge on ontogenetic variation of spectra will be of high relevance for defining the minimum requirements on stage-resolution in the applied reference library, i.e. whether the inclusion of adult stages will be sufficient to identify juvenile stages. Our results indicate that mass peak variability between stages does usually not affect species level classification. Even though there are differences between mass spectra from the different stages, differences in the proteome among stages was smaller than interspecific differences. Most specimens from certain stages were still confidently classifiable on species level even if the respective stage had been removed from the reference library. This may allow identification of real samples using a partly incomplete reference library concerning the different stages. For example, scyphozoan ephyra and polyp stages which are difficult to identify to species level by morphological methods (Holst, 2012) could be identified by MALDI-TOF MS even if the reference library is based on adult medusae only. Problems may occur in cases where between-stage distances are high as seen in *C. lamarckii* and *C. hysoscella* . Although in these cases, RF classification success was high, the majority of classifications were rejected by the *post-hoc* test based on assignment probabilities. These classifications would therefore require a re-investigation by morphology or genetic approaches (Rossel and Martínez Arbizu, 2018). With a growing reference library, the amount of correct classifications being recognized as false positives will most likely decrease.

Differences between stages can be seen in a variety of mass peaks. Some are frequently found to be present

or absent in some stages. However, the majority of peaks, also those driving the stage differences in the RF model, differ only in relative mass-peak intensity. Thus, mass spectra rather seem to change continuously with development than abruptly with the onset of the next stage. Certain proteins or peptides may already be expressed before transition to the next developmental stage or still be expressed after transition and therefore be recorded in similar stages. This would be comparable to some kind of intermoult status assumed to cause misclassification in different stages of *Calanus* species classified using MALDI-TOF MS (Rossel et al., 2022a).

Other factors influencing mass spectra variability may be environmental differences impacting the physiology (Karger et al., 2019) and/or underlying differences between populations (Müller et al., 2013; Benkacimi et al., 2020). Mass spectra variability of the two scyphozoans from the North Sea and the Baltic Sea were not as strongly influenced by their sampling localities as was previously shown for copepods (Peters et al., 2022). Specimens tend to cluster according to regions (supplementary figure 1) but clusters also align with sampling and/or measurement occasions. There was no effect of specimen origin on species classification by the RF model. In other taxa, groupings based on MALDI-TOF MS data according to sample location were found, for example in calanoid copepods clustering according to origin lake (Riccardi et al., 2012). Thus, it can be assumed that ecological factors influence proteomic fingerprints, however the strength of the effect may depend on taxon physiology. Copepods osmoregulate by changes of the osmolarity of their hemolymph (Roddie et al., 1984; Lee et al., 2012) and changes in protein expression were found under osmotic stress (DeBiasse et al., 2018). Differences in salinity may have lower effects on the measured proteome in osmoconformers (Rivera-Ingraham and Lignot, 2017). Although, changes in salinity can impact cnidarian larval settlement and reproduction (Glon et al., 2019; Dańko et al., 2020; Schäfer et al., 2021), protein expression associated with osmoregulation may be less affected since cnidarians are osmoconformers with only slight differences in the osmolarities of sea water and the gelatinous tissues (Wright and Purcell, 1997; Graham, 2001).

The fact that preserved specimens of the same species which were stored for different time periods or were measured on different dates clustered together, demonstrates the reliability of the approach. Previous studies have shown that siphonophore tissues preserved more than one year were still useful for MALDI-TOF MS (Park et al., 2021). Our results now confirm that even cnidarian samples preserved much longer with storage times up to 111 months can still be successfully used.

Moreover, the fact that scyphozoan specimens reared in laboratory cultures clustered together with their conspecifics collected in the field corroborates the assumption that the effects of environmental factors on proteomic fingerprints are low in this taxon. This would facilitate the applicability of the method since the creation of a reference library for different environments would not be necessary. However, verifying this demands investigation of further species from other regions. To include considerations about variability based on sample origin for future reference libraries, thoroughly planned experiments should be carried out to investigate which ecological factors have a major impact on mass-spectra variability.

From our results we conclude that proteomic fingerprinting is a reliable method to differentiate and identify cnidarian species including different scyphozoan life-history stages. Especially in the context of identifying specimens deformed beyond recognition from samples fixed for monitoring purposes, this time- and cost-effective method represents a valid alternative method to molecular genetic identification tools.

Author contributions

All authors devised the study. SH, SL and SR carried out MALDI-TOF MS measurements. SR and JP analyzed the data. SR, JP and SH drafted the manuscript. All authors critically reviewed the manuscript.

Acknowledgement

We thank Josephine Goldstein, Hermann Neumann and Rebekka Schüller for providing samples and morphological species identification as well as Peggy Weist and Jil Kühne for their support in proteomic sample processing. This work was also supported by the DFG initiative 1991 "Taxon-omics"(grant no.RE2808/3-1 and RE2808/3-2). HIFMB is a collaboration between the Alfred-Wegener-Institute, Helmholtz-Center for

Polar and Marine Research, and the Carl-von-Ossietzky University Oldenburg, initially funded by the Ministry for Science and Culture of Lower Saxony and the Volkswagen Foundation through the 'Niedersächsisches Vorab' grant program (grant no. ZN3285). This work was funded by the Federal Ministry of Education and Research (Grant No. 03F0499A) and the Land Niedersachsen. This is publication number 20 of Senckenberg am Meer Proteome Laboratory.

Data availability statement

All raw data is stored at Data Dryad (DOI:XXX). No further data was collected during this study.

References

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x.

Arai, M. N. (1997). *A Functional Biology of Scyphozoa* . Springer Science & Business Media.

Benkacimi, L., Gazelle, G., El Hamzaoui, B., Bérenger, J.-M., Parola, P., and Laroche, M. (2020). MALDI-TOF MS identification of Cimex lectularius and Cimex hemipterus bedbugs. *Infection, Genetics and Evolution* 85, 104536. doi: 10.1016/j.meegid.2020.104536.

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

Bouillon, Gravili, C., F, P., Gili, J.-M., and Boero, F. (2006). An Introduction to Hydrozoa. *Mémoires du Muséum national d'Histoire naturelle* 194, 1–591.

Breiman, L. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science* 16, 199–231.

Brugler, M. R., González-Muñoz, R. E., Tessler, M., and Rodríguez, E. (2018). An EPIC journey to locate single-copy nuclear markers in sea anemones. *Zoologica Scripta* 47, 756–776. doi: 10.1111/zsc.12309.

Bucklin, A., Peijnenburg, K. T. C. A., Kosobokova, K. N., O'Brien, T. D., Blanco-Bercial, L., Cornils, A., et al. (2021). Toward a global reference database of COI barcodes for marine zooplankton. *Mar Biol* 168, 78. doi: 10.1007/s00227-021-03887-y.

Chen, X.-F., Hou, X., Xiao, M., Zhang, L., Cheng, J.-W., Zhou, M.-L., et al. (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, 1536.

Dańko, A., Schaible, R., and Dańko, M. J. (2020). Salinity Effects on Survival and Reproduction of Hydrozoan Eleutheria dichotoma.*Estuaries and Coasts* 43, 360–374. doi: 10.1007/s12237-019-00675-2.

DeBiasse, M. B., Kawji, Y., and Kelly, M. W. (2018). Phenotypic and transcriptomic responses to salinity stress across genetically and geographically divergent Tigriopus californicus populations.*Molecular Ecology* 27, 1621–1632. doi: 10.1111/mec.14547.

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

Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*31, 3718–3720. doi: 10.1093/bioinformatics/btv428.

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

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

Glon, H., Haruka, Y., Daly, M., and Nakaoka, M. (2019). Temperature and salinity survival limits of the fluffy sea anemone, Metridium senile (L.), in Japan. *Hydrobiologia* 830, 303–315. doi: 10.1007/s10750-018-3879-2.

Graham, W. M. (2001). Numerical increases and distributional shifts of Chrysaora quinquecirrha (Desor) and Aurelia aurita (Linné) (Cnidaria: Scyphozoa) in the northern Gulf of Mexico. in *Jellyfish Blooms: Ecological and Societal Importance* Developments in Hydrobiology., eds. J. E. Purcell, W. M. Graham, and H. J. Dumont (Dordrecht: Springer Netherlands), 97–111. doi: 10.1007/978-94-010-0722-1_9.

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

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

Häussermann, V. (2004). Re-description of Phymactis papillosa (Lesson, 1830) and Phymanthea pluvia (Drayton in Dana, 1846) (Cnidaria: Anthozoa), two common actiniid sea anemones from the south east Pacific with a discussion of related genera. *Zoologische Mededelingen* 78, 18–28, 345–381.

Holst, S. (2012). Morphology and development of benthic and pelagic life stages of North Sea jellyfish (Scyphozoa, Cnidaria) with special emphasis on the identification of ephyra stages. *Mar Biol* 159, 2707–2722. doi: 10.1007/s00227-012-2028-0.

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

Holst, S., and Laakmann, S. (2014). Morphological and molecular discrimination of two closely related jellyfish species, Cyanea capillata and C. lamarckii (Cnidaria, Scyphozoa), from the northeast Atlantic. *Journal of Plankton Research* 36, 48–63.

Jarms, G., and Morandini, A. (2019). *World Atlas of Jellyfish* .

Kaiser, P., Bode, M., Cornils, A., Hagen, W., Martínez Arbizu, P., Auel, H., et al. (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., and 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, e0210590. doi: 10.1371/journal.pone.0210590.

Korfhage, S. A., Rossel, S., Brix, S., McFadden, C. S., Ólafsdóttir, S. H., and 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. Available at: https://www.frontiersin.org/article/10.3389/fmars.2022.838201.

Krijthe, J. (2015). R wrapper for Van der Maaten's Barnes-Hut implementation of t-Distributed Stochastic Neighbor Embedding. Available at: https://github.com/jkrijthe/Rtsne [Accessed March 24, 2022].

Kürzel, K., Kaiser, S., Lörz, A.-N., Rossel, S., Paulus, E., Peters, J., et al. (2022). Correct Species Identification and Its Implications for Conservation Using Haploniscidae (Crustacea, Isopoda) in Icelandic Waters as a Proxy. *Frontiers in Marine Science* 8. doi: doi: 10.3389/fmars.2021.795196.

Laakmann, S., Gerdts, G., Erler, R., Knebelsberger, T., Martínez Arbizu, P., and 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, 862–76. doi: 10.1111/1755-0998.12139.

Lafri, I., Almeras, L., Bitam, I., Caputo, A., Yssouf, A., Forestier, C.-L., et al. (2016). Identification of Algerian Field-Caught Phlebotomine Sand Fly Vectors by MALDI-TOF MS. *PLOS Neglected Tropical Diseases* 10, e0004351. doi: 10.1371/journal.pntd.0004351.

Lawley, J. W., Gamero-Mora, E., Maronna, M. M., Chiaverano, L. M., Stampar, S. N., Hopcroft, R. R., et al. (2021). The importance of molecular characters when morphological variability hinders diagnosability: systematics of the moon jellyfish genus Aurelia (Cnidaria: Scyphozoa). *PeerJ* 9, e11954. doi: 10.7717/peerj.11954.

Lee, C. E., Posavi, M., and Charmantier, G. (2012). Rapid evolution of body fluid regulation following independent invasions into freshwater habitats. *Journal of Evolutionary Biology* 25, 625–633. doi: 10.1111/j.1420-9101.2012.02459.x.

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

Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R news* 2, 18–22.

Martell, L., Selsø, K., Titelman, J., and Hosia, A. (2022). Setting the baseline for the dynamics of siphonophores and hydromedusae in Oslofjorden. *Marine Ecology Progress Series* 686, 71–89. doi: 10.3354/meps13991.

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

Mazzeo, M. F., and Siciliano, R. A. (2016). Proteomics for the authentication of fish species. *Journal of proteomics* 147, 119–124.

McFadden, C. S., Benayahu, Y., Pante, E., Thoma, J. N., Nevarez, P. A., and France, S. C. (2011). Limitations of mitochondrial gene barcoding in Octocorallia. *Molecular Ecology Resources* 11, 19–31. doi: 10.1111/j.1755-0998.2010.02875.x.

McFadden, C. S., Brown, A. S., Brayton, C., Hunt, C. B., and van Ofwegen, L. P. (2014). Application of DNA barcoding in biodiversity studies of shallow-water octocorals: molecular proxies agree with morphological estimates of species richness in Palau. *Coral Reefs*33, 275–286. doi: 10.1007/s00338-013-1123-0.

Mills, Calder, D., Marques, A., Migotto, A., Haddock, S., and Dunn, C. (2007). "Cnidaria: Hydrozoa: polyps, Hydromedusae, and Siphonophora, Light and Smith's Manual: Intertidal Invertebrates of the Central California Coast," in (Berkeley: University of California Press), 151–168.

Miranda, L. S., Mills, C. E., Hirano, Y. M., Collins, A. G., and Marques, A. C. (2018). A review of the global diversity and natural history of stalked jellyfishes (Cnidaria, Staurozoa). *Mar Biodiv*48, 1695–1714. doi: 10.1007/s12526-017-0721-4.

Moura, C. J., Lessios, H., Cortés, J., Nizinski, M. S., Reed, J., Santos, R. S., et al. (2018). Hundreds of genetic barcodes of the species-rich hydroid superfamily Plumularioidea (Cnidaria, Medusozoa) provide a guide toward more reliable taxonomy. *Sci Rep* 8, 17986. doi: 10.1038/s41598-018-35528-8.

Müller, P., Pflüger, V., Wittwer, M., Ziegler, D., Chandre, F., Simard, F., et al. (2013). Identification of cryptic Anopheles mosquito species by molecular protein profiling. *PLOS ONE* 8, e57486. doi: 10.1371/journal.pone.0057486.

Nabet, C., Kone, A. K., Dia, A. K., Sylla, M., Gautier, M., Yattara, M., et al. (2021). New assessment of Anopheles vector species identification using MALDI-TOF MS. *Malaria Journal* 20, 1–16.

Niare, S., Berenger, J.-M., Dieme, C., Doumbo, O., Raoult, D., Parola, P., et al. (2016). Identification of blood meal sources in the main African malaria mosquito vector by MALDI-TOF MS. *Malar J* 15, 87. doi: 10.1186/s12936-016-1152-6.

Niare, S., Tandina, F., Davoust, B., Doumbo, O., Raoult, D., Parola, P., et al. (2018). Accurate identification of Anopheles gambiae Giles trophic preferences by MALDI-TOF MS. *Infection, Genetics and Evolution* 63, 410–419. doi: 10.1016/j.meegid.2017.09.009.

Okamura, B., Gruhl, A., and Bartholomew, J. L. (2015). "An Introduction to Myxozoan Evolution, Ecology and Development," in *Myxozoan Evolution, Ecology and Development* , eds. B. Okamura, A. Gruhl, and J. L. Bartholomew (Cham: Springer International Publishing), 1–20. doi: 10.1007/978-3-319-14753-6_1.

Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633.

Park, N., Yeom, J., Jeong, R., and Lee, W. (2021). Novel attempt at discrimination of a bullet-shaped siphonophore (Family Diphyidae) using matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-ToF MS). *Sci Rep* 11, 19077. doi: 10.1038/s41598-021-98724-z.

Paulus, E., Brix, S., Siebert, A., Martínez Arbizu, P., Rossel, S., Peters, J., et al. (2022). Recent speciation and hybridization in Icelandic deep-sea isopods: An integrative approach using genomics and proteomics. *Molecular Ecology* 31, 313–330. doi: 10.1111/mec.16234.

Peters, J., Laakmann, S., Rossel, S., Arbizu, P. M., and Renz, J. (2022). Perspectives of species identification by MALDI-TOF MS in monitoring - stability of proteomic fingerprints in marine epipelagic copepods. Preprints doi: 10.22541/au.166671183.32080869/v1.

Pruski, S., and Miglietta, M. P. (2019). Fluctuation and diversity of Hydromedusae (Hydrozoa, Cnidaria) in a highly productive region of the Gulf of Mexico inferred from high frequency plankton sampling.*PeerJ* 7, e7848. doi: 10.7717/peerj.7848.

R-Core-Team (2022). R: A language and environment for statistical computing. Available at: https://www.R-project.org/.

Renz, J., Markhaseva, E. L., Laakmann, S., Rossel, S., Martínez Arbizu, P., and Peters, J. (2021). Proteomic fingerprinting facilitates biodiversity assessments in understudied ecosystems: A case study on integrated taxonomy of deep sea copepods. *Molecular Ecology Resources* .

Riccardi, N., Lucini, L., Benagli, C., Welker, M., Wicht, B., and 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, 484–492.

Rivera-Ingraham, G. A., and Lignot, J.-H. (2017). Osmoregulation, bioenergetics and oxidative stress in coastal marine invertebrates: raising the questions for future research. *Journal of Experimental Biology* 220, 1749–1760. doi: 10.1242/jeb.135624.

Roddie, B. D., Leakey, R. J. G., and Berry, A. J. (1984). Salinity-temperature tolerance and osmoregulation in Eurytemora affinis (Poppe) (Copepoda : Calanoida) in relation to its distribution in the zooplankton of the upper reaches of the Forth estuary. *Journal of Experimental Marine Biology and Ecology* 79, 191–211. doi: 10.1016/0022-0981(84)90219-3.

Rodriguez, E., Barbeitos, M. S., Brugler, M. R., Crowley, L. M., Grajales, A., Gusmao, L., et al. (2014). Hidden among Sea Anemones: The First Comprehensive Phylogenetic Reconstruction of the Order Actiniaria (Cnidaria, Anthozoa, Hexacorallia) Reveals a Novel Group of Hexacorals.*PLOS ONE* 9, e96998. doi: 10.1371/journal.pone.0096998.

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

Rossel, S., Kaiser, P., Bode-Dalby, M., Renz, J., Laakmann, S., Auel, H., et al. (2022a). Proteomic fingerprinting enables quantitative biodiversity assessments of species and ontogenetic stages in Calanus congeners (Copepoda, Crustacea) from the Arctic Ocean. *Molecular Ecology Resources* n/a. doi: 10.1111/1755-0998.13714.

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

Rossel, S., and Martinez Arbizu, P. (2018). 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, 1421–1434. doi: 10.1111/2041-210X.13000.

Rossel, S., and Martinez 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, 9182. doi: 10.1038/s41598-019-45718-7.

Rossel, S., and Martinez Arbizu, P. (2020). Unsupervised biodiversity estimation using proteomic fingerprints from MALDI-TOF MS data.*Limnology and Oceanography: Methods* . doi: https://doi.org/10.1002/lom3.10358.

Rossel, S., Uhlenkott, K., Peters, J., Vink, A., and Martinez Arbizu, P. (2022b). Evaluating species richness using proteomic fingerprinting and DNA-barcoding – a case study on meiobenthic copepods from the Clarion Clipperton Fracture Zone. *Marine Biodiversity* .

Ryan, C., Clayton, E., Griffin, W., Sie, S., and 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, 396–402.

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

Schafer, S., Gueroun, S. K. M., Andrade, C., and Canning-Clode, J. (2021). Combined Effects of Temperature and Salinity on Polyps and Ephyrae of Aurelia solida (Cnidaria: Scyphozoa). *Diversity* 13, 573. doi: 10.3390/d13110573.

Schuchert, P. (2012). *North-West European Athecate Hydroids and their Medusae* .

Schuchert, P. (2020). DNA barcoding of some Pandeidae species (Cnidaria, Hydrozoa, Anthoathecata). *rsdz* 125, 101–127. doi: 10.5281/zenodo.1196029.

Tran, A., Alby, K., Kerr, A., Jones, M., and Gilligan, P. H. (2015). Cost Savings Realized by Implementation of Routine Microbiological Identification by Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry. *J Clin Microbiol* 53, 2473–2479. doi: 10.1128/JCM.00833-15.

Wilke, T., Renz, J., Hauffe, T., Delicado, D., and Peters, J. (2020). Proteomic Fingerprinting Discriminates Cryptic Gastropod Species.*Malacologia* 63, 131–137.

Wright, D. A., and Purcell, J. E. (1997). Effect of Salinity on Ionic Shifts in Mesohaline Scyphomedusae, Chrysaora quinquecirrha. *The Biological Bulletin* 192, 332–339. doi: 10.2307/1542726.

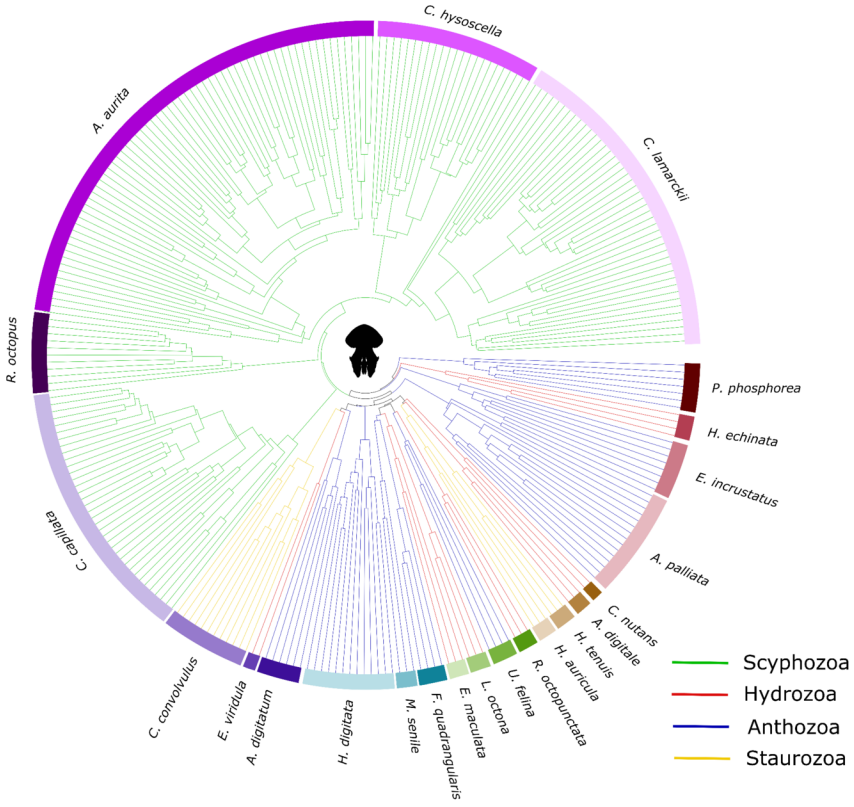

Figure 1: Circular representation of hierarchical clustering of 278 specimens from 23 Cnidaria species based on MALDI-TOF mass spectra. Colored bars highlight species clusters and colored branches indicate class affiliation. Species showing ambiguous clustering are highlighted.

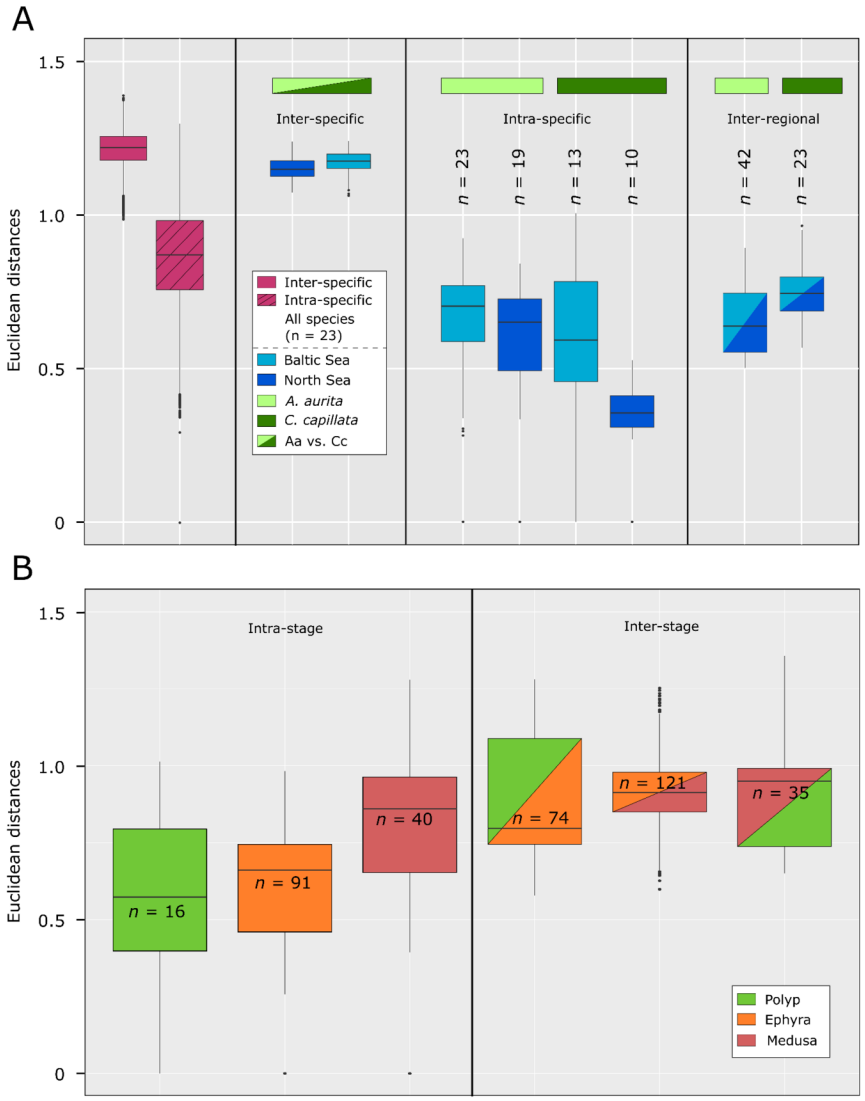

Figure 2: Euclidean distances based on MALDI-TOF mass spectra.**A** Inter- and intraspecific distances of all analyzed species (red) and selected species (blue) from different localities. **B**Intra-stage- and inter-stage distances between polyp, medusa and ephyra stages in scyphozoan species (*Aurelia aurita* , *Cyanea capillata* , *Cyanea lamarckii*). Distances were compared between species from the same geographic region. Two-colored boxes indicate distances between different regions or stages.

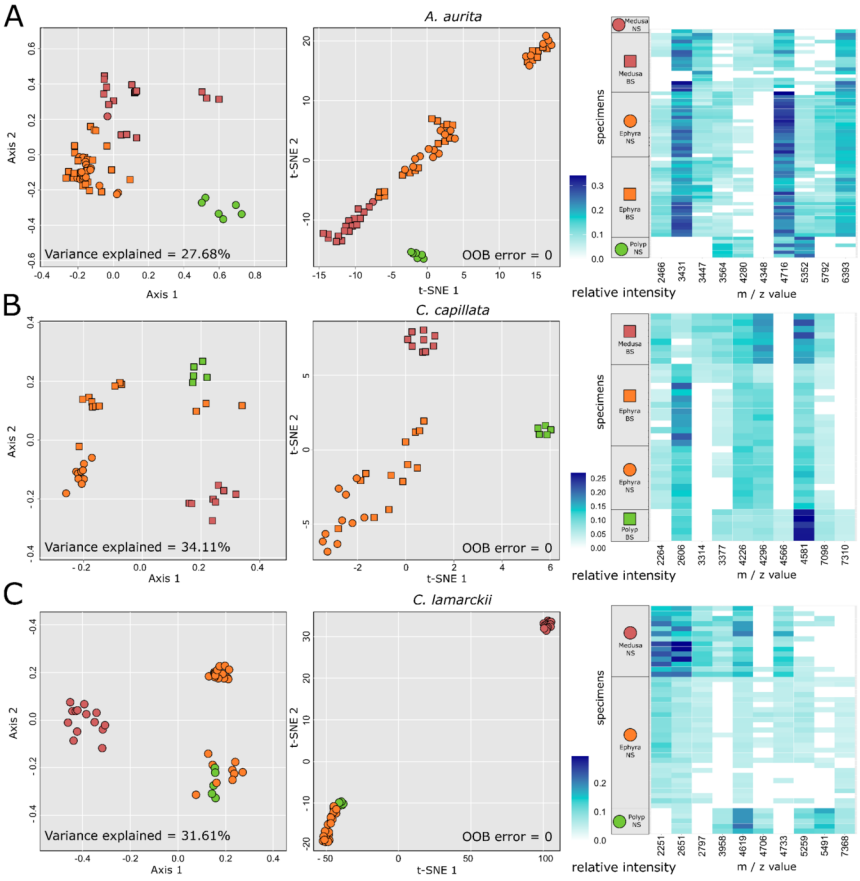

Figure 3: Within-species representation of the different stages based on MALDI-TOF mass spectra. Left: PCoA of Hellinger-transformed and processed raw data. Middle: t-SNE plot based on class votes from a RF model on developmental-stage level. Right: Heat maps depicting relative intensities of the ten most important mass peaks for developmental-stage differentiation identified using RF classifier. **BS** Baltic Sea.**NS** North Sea.

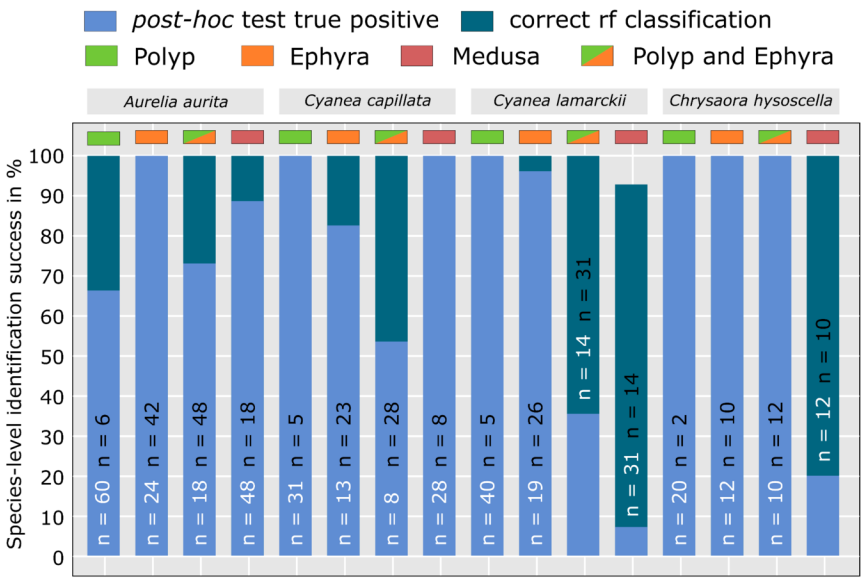

Figure 4: Classification success on species level of stages excluded from the classification model. Dark blue bars represent RF classification success. Light blue bars indicate classification success by the *post-hoc* test for the respective group. Colored boxes indicate the excluded/classified class. Boxes in shaded colors indicate the exclusion of two stages at once. White lettering indicates library size. Black lettering provides the number of classified specimens.

Supplementary Figure 1: Euclidean distance matrix of *A. aurita*and *C. capillata* specimens investigated in this study. Green and Orange shaded bars indicate sample occasions. Blue shaded bars illustrate sample origin. Distances are shown by colors ranging from 0 to around 1.2.