



OPEN

# 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)

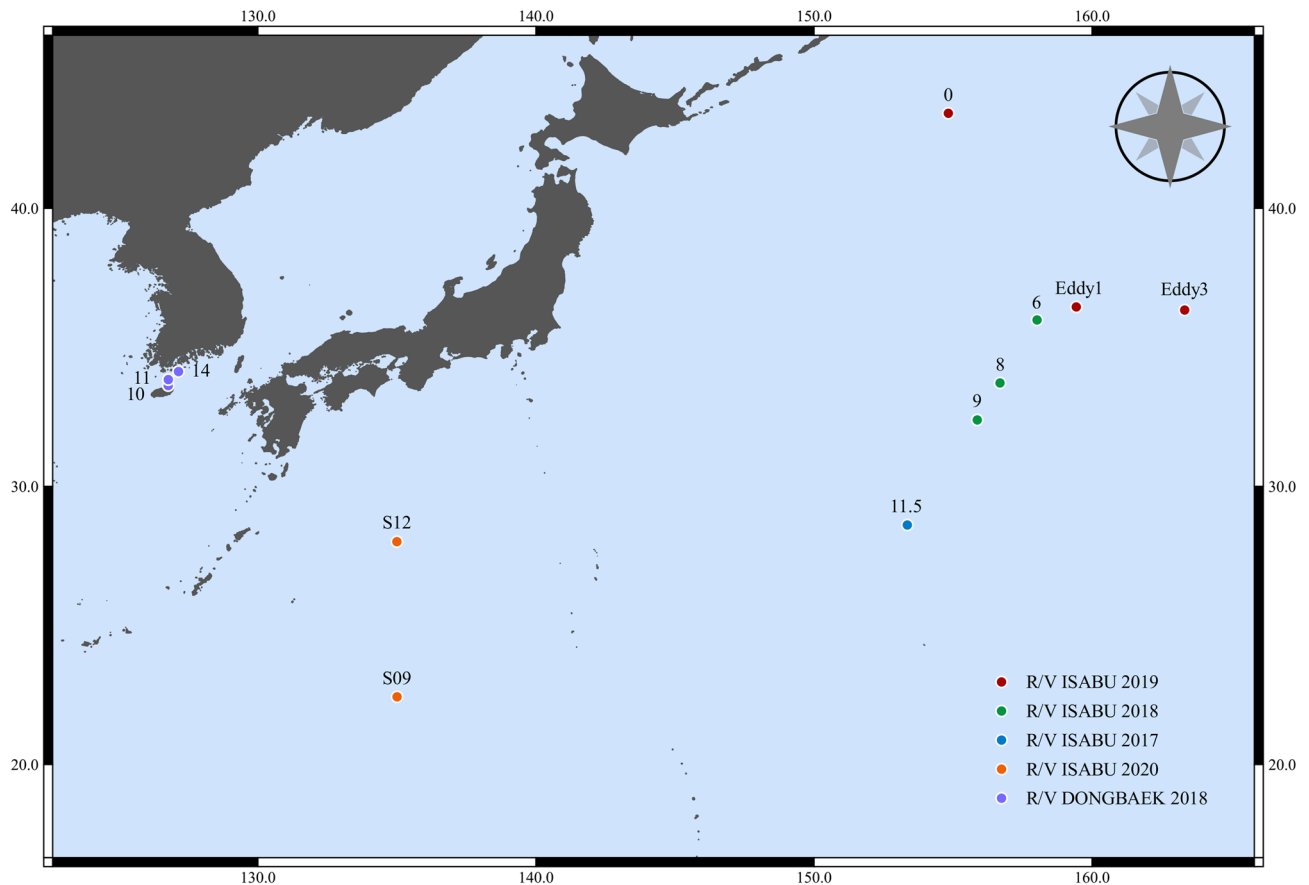
Nayeon Park<sup>1,3</sup>, Jisu Yeom<sup>1,3</sup>, Raehyuk Jeong<sup>2</sup> & Wonchoel Lee<sup>1✉</sup>

One major difficulty in identifying the gelatinous bodied bullet-shaped Siphonophore, Diphyids, is that their shape is deformed following ethanol fixation. Ethanol often is preferred over other fixatives, since samples fixed in ethanol can be used for molecular studies that can supplement morphological findings. To overcome this problem, we obtained protein mass spectra of ten species of Diphyidae found in the waters of the Kuroshio Current (Northwest Pacific and South Coast of South Korea) to test whether MALDI-ToF MS could be used as a methodology for species identification. In addition, a number of morphological characteristics that can be used with ethanol-treated samples was summarized. Concatenated phylogenetic analysis was also performed to determine the phylogenetic relationship by obtaining partial sequences of four genes (mtCOI, 16S rRNA, 18S rRNA, and ITS regions). Based on our integrative analysis, MALDI-ToF MS was evaluated as a potentially fast, inexpensive, and accurate tool for species identification along with conventional morphological and DNA barcoding for Diphyidae.

Over 60% of Siphonophorae Eschscholtz, 1829 which form unique colonies<sup>1</sup> through their life are Diphyidae Quoy & Gaimard, 1827<sup>2</sup>. The family Diphyidae consists of 45 species belonging to eight genera, and it is the most species-diverse family of seven belonging to the suborder Calycophorae Leuckart, 1854. Aside from the currently valid 45 species, there is a handful of species that is considered species inquirenda, which require taxonomic review. Through alternation of generations, Diphyidae goes through polygastric (creating eudoxid phase through asexual reproduction) and eudoxid (creating polygastric phase through sexual reproduction) phases<sup>3</sup>. Polygastric Diphyidae consists of one or two nectophores, of which the small bullet-shaped anterior nectophore contains key information for classification, such as somatocyst, mouthplate, ostial teeth, hydroecium, and longitudinal ridges. In fact, there are many instances where the morphological record of Diphyidae often is dependent on the anterior nectophore. Despite their morphological importance for species identification, anterior nectophores often are easily damaged during fixation, since they consist of gelatin. In particular, using ethanol as a fixative will induce dehydration, shriveling the shape of the species to a point that complicates species identification. For this reason, much effort has been asserted in the last decade to utilize DNA barcoding for species identification. However, DNA barcoding is time consuming and costly with many limitations, such as requiring prior knowledge of a target species to design species-specific primers to maximize sequencing success.

In species identification for pathogens or microbes, much easier and faster results can be achieved through Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-ToF MS). MALDI-ToF MS is a method developed by Karas & Hillenkamp<sup>4</sup>, which uses an aromatic carboxylic acid matrix to crystallize the protein, which is then hit by a laser in a vacuum to cause ionization. The ionized particles are passed through a ToF analyzer, where the mass of the protein is calculated by measuring the time of flight<sup>5</sup>. These protein mass

<sup>1</sup>Laboratory of Biodiversity, Department of Life Science, Hanyang University, Seoul 04763, South Korea. <sup>2</sup>Big Data Biology Lab, Department of Life Science, Chung-Ang University, Seoul 06974, South Korea. <sup>3</sup>These authors contributed equally: Nayeon Park and Jisu Yeom. ✉email: wlee@hanyang.ac.kr



**Figure 1.** Map of collecting stations marked with the spots.

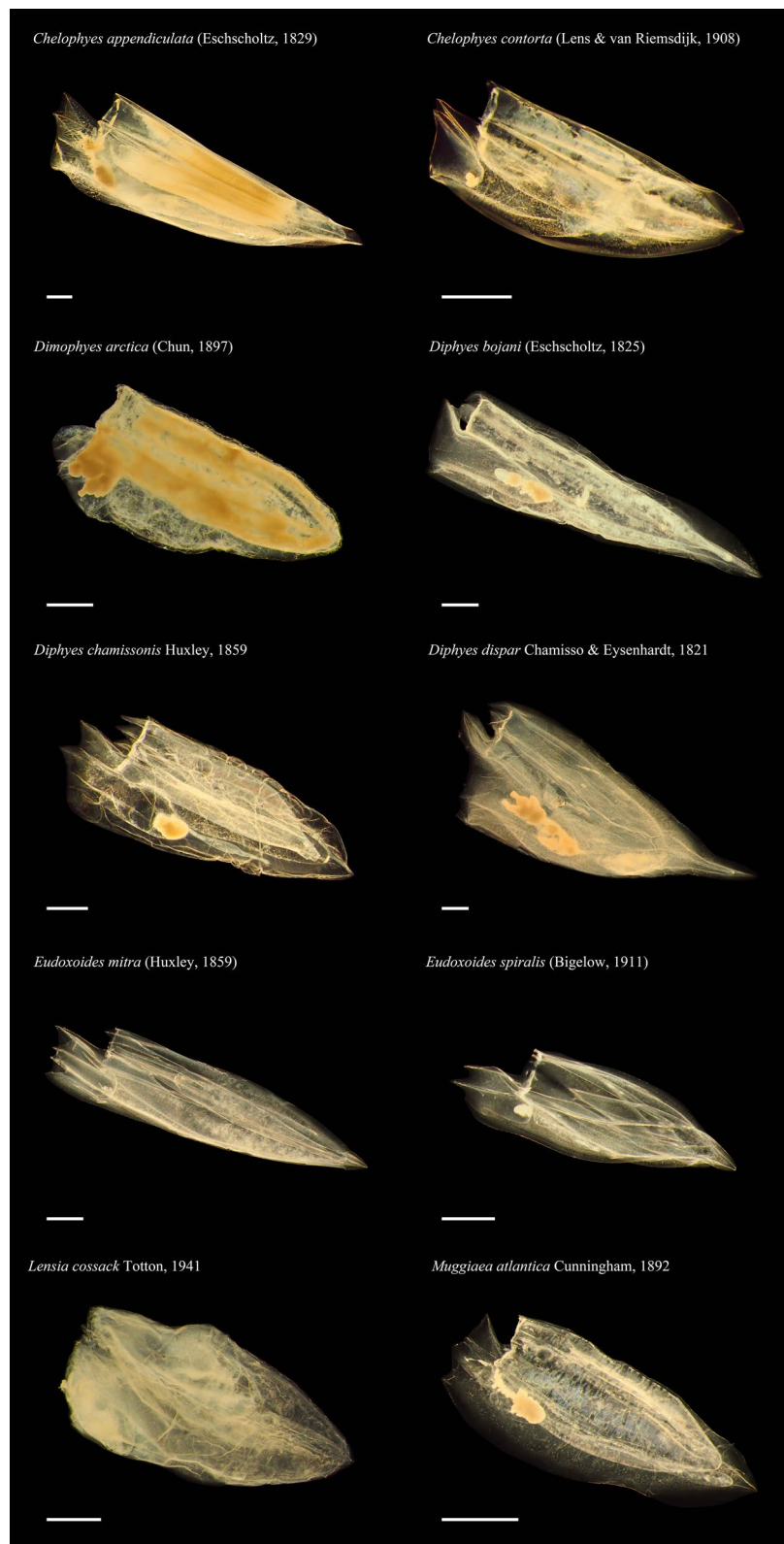
spectra serve as unique fingerprints, enabling species identification. Recent studies have been expanded to apply to eukaryotes, such as mosquitoes, fruit flies, sand flies, fishes, and copepods<sup>6–11</sup>. This method has been attempted on jellyfishes as well<sup>12–14</sup>, but it was used to target specific proteins of nematocysts, not to identify species.

In this study, we evaluated the potential of MALDI-ToF MS as a tool for species identification of jellyfish by testing the method on ten species of Diphyidae found in the Kuroshio Current and off the coast of Korea (Fig. 1). This was the first case in which MALDI-ToF MS was utilized as a tool for jellyfish species identification. In addition, we conducted an integrated taxonomic review of the ten species based on morphological analysis and molecular analysis of four DNA marker sequences (mtCOI, 16S rRNA, 18S rRNA, and ITS regions). The three integrative approaches utilized in this study will heighten our understanding of bullet-shaped Siphonophore, which has been subject to identification difficulties due to morphological ambiguity.

## Results

**Morphological comparison.** A total of ten species of Diphyidae was obtained, and all species were recorded using anterior nectophore (Fig. 2). All specimens were bullet-shaped, except for *Diphyes dispar* Chamisso & Eysenhardt, 1821, and *Eudoxoides spiralis* (Bigelow, 1911), which had modified forms of triangular bullet-shaped and twisted bullet-shaped, respectively. Morphological key features such as shape of the apex, number of longitudinal ridges, hydroecium, ostial teeth, mouthplate, and somatocyst were compared (Table 1).

**Pattern of the protein mass spectra.** Protein mass spectra were obtained from a total of ten species of Diphyidae specimens in a mass peak range between 1.5 and 20 k Dalton (Da) (Table 2). The quality of the mass spectra was controlled separately by eye, and a total of 123 spectra was selected. The same species showed similar peak patterns. On the other hand, noticeable differences in peak patterns were confirmed among the species (Supplementary Fig. S1 online). Figure 3 comprehensively shows the variation in important peaks. Even within the same species, there were slight differences depending on individual. On a larger scale, the peak pattern differed by species. In addition, the peak pattern did not seem to be significantly affected by year of collection, location, and storage period (Supplementary Fig. S2A, B online). Supplementary Fig. S3 online showed the results obtained from Diagonal Discriminant Analysis, which indicated that how strong the most remarkable peaks (protein masses that differed greatly among the species) extracted from the 123 spectra were relatively strong in each species. For instance, *Diphyes chamissonis* Huxley, 1859 had a relatively large amount of 2125–2128 Da compared to the other species. This result confirmed that the level of expression by the peak varied greatly depending on the species. Meanwhile, *Eudoxoides mitra* (Huxley, 1859) and *E. spiralis* showed almost identical features. This was a different result from the overall peak pattern (Fig. 3).



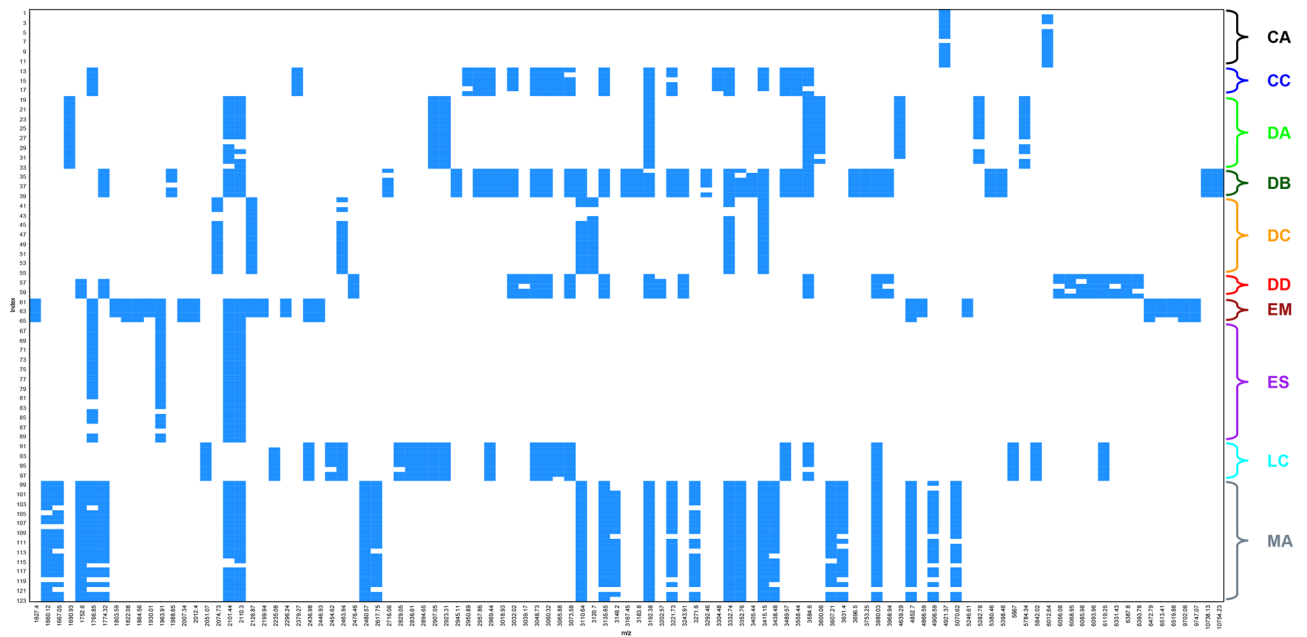
**Figure 2.** Digital images of anterior nectophores from target Diphyids. All specimens are ethanol fixed. Scale bar = 1 mm.

| Family                        | Diphyidae            |                            |                      |                      |                    |                             |                      |                      |                |                  |
|-------------------------------|----------------------|----------------------------|----------------------|----------------------|--------------------|-----------------------------|----------------------|----------------------|----------------|------------------|
| Genus                         | <i>Chelophyes</i>    |                            | <i>Dimophyes</i>     | <i>Diphyes</i>       |                    |                             | <i>Eudoxoides</i>    |                      | <i>Lensia</i>  | <i>Muggiaea</i>  |
| Species                       | <i>appendiculata</i> | <i>contorta</i>            | <i>arctica</i>       | <i>bojani</i>        | <i>chamissonis</i> | <i>dispar</i>               | <i>mitra</i>         | <i>spiralis</i>      | <i>cossack</i> | <i>atlantica</i> |
| Author                        | Eschscholtz, 1829    | Lens & van Riemsdijk, 1908 | Chun, 1897           | Eschscholtz, 1825    | Huxley, 859        | Chamisso & Eysenhardt, 1821 | Huxley, 1859         | Bigelow, 1911        | Totton, 1941   | Cunningham, 1892 |
| Number of Nectophores         | 2                    | 2                          | 2                    | 2                    | 1                  | 2                           | 2                    | 1                    | 1              | 1                |
| Number of Nectophore Types    | 2                    | 2                          | 2                    | 2                    | 1                  | 2                           | 2                    | 1                    | 1              | 1                |
| Comparison of Nectophore Size | Anterior > Posterior | Anterior > Posterior       | Anterior > Posterior | Anterior > Posterior | X                  | Anterior > Posterior        | Anterior > Posterior | X                    | X              | X                |
| Shape of Anterior Nectophore  | Bullet-shape         | Bullet-shape               | Bullet-shape         | Bullet-shape         | Bullet-shape       | Triangular Bullet-shape     | Bullet-shape         | Twisted Bullet-shape | Bullet-shape   | Bullet-shape     |
| Shape of Apex                 | Pointed              | Pointed                    | Blunt                | Pointed              | Pointed            | Pointed                     | Pointed              | Pointed              | Blunt          | Pointed          |
| Longitudinal Ridges           | 5                    | 4                          | X                    | 5                    | 5                  | 5                           | 5                    | 5                    | 5              | 5                |
| Hydroecium Depth              | Medium               | Medium                     | Medium               | Deep                 | Deep               | Deep                        | Medium               | Medium               | Shallow        | Deep             |
| Hydroecium Shape              | Claw-shaped          | Claw-shaped                | Rounded              | Rounded              | Rounded            | Rounded                     | Rounded              | Rounded              | Flatted        | Rounded          |
| Ostial Teeth                  | X                    | X                          | X                    | O                    | O                  | O                           | X                    | X                    | X              | X                |
| Mouthplate                    | Divided              | Divided                    | Undivided            | Undivided            | Undivided          | Undivided                   | Divided              | Divided              | Divided        | Divided          |
| Shape of Somatocyst           | Fusiform             | Oblique Club-shaped        | Fusiform             | Fusiform             | Fusiform           | Fusiform                    | Fusiform             | Fusiform             | Ovoid          | Filiform         |
| Length of Somatocyst Stalk    | Short                | Long                       | Short                | Short                | Short              | Short                       | Short                | Short                | Short          | Short            |
| Somatocyst/Anterior Nectosac  | 1/2–3/4              | 1/2                        | 1/2–3/4              | 1/2                  | 1/2–1/3            | 1/3–1/4                     | 1/2–1/3              | 1/2–1/3              | 1/3            | 4/5              |

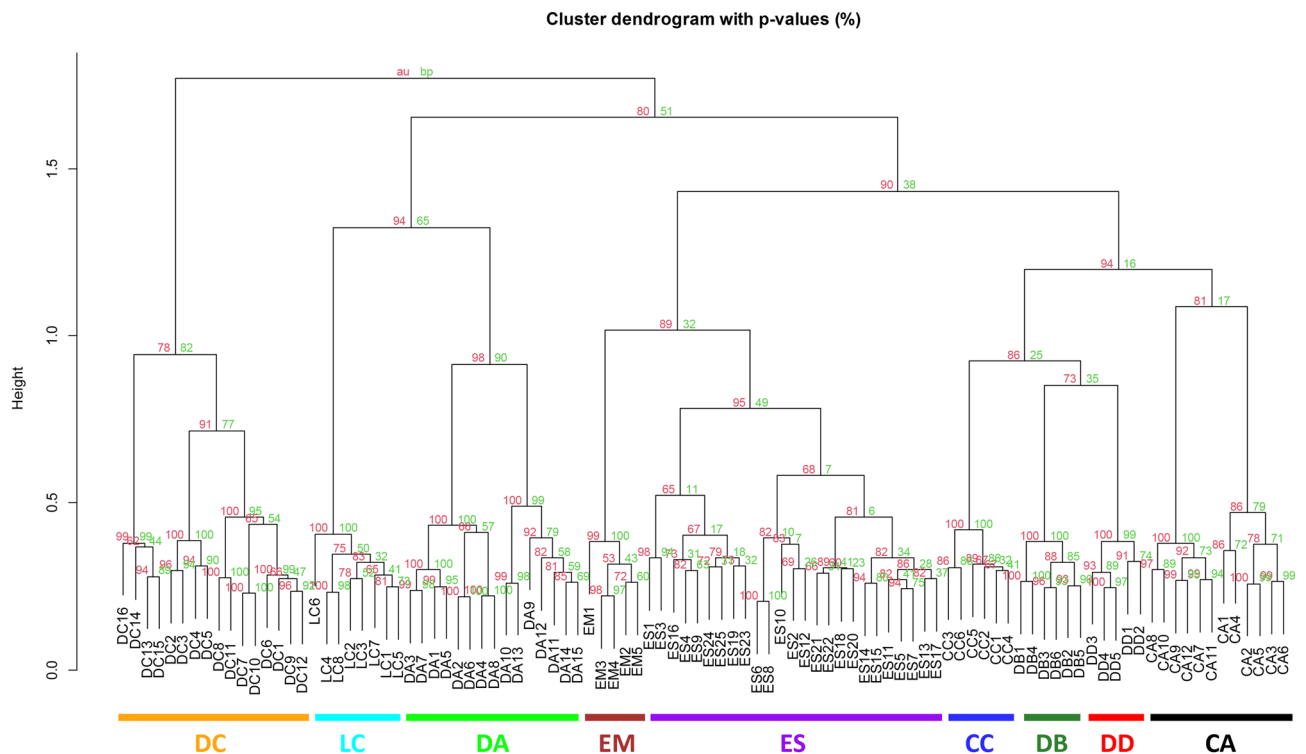
**Table 1.** Morphological comparisons between the ten Diphyidae species from ethanol-fixed samples. O: presence; X: absence.

| Species                         | Abbreviation | Year. month | Cruise       | Station | Mass spectra | DNA sequences |
|---------------------------------|--------------|-------------|--------------|---------|--------------|---------------|
| <i>Chelophyes appendiculata</i> | CA           | 2018. 08    | R/V ISABU    | 9       | 6            | –             |
|                                 |              | 2019. 10    | R/V ISABU    | Eddy3   | 6            | 1             |
| <i>Chelophyes contorta</i>      | CC           | 2020. 05    | R/V ISABU    | S09     | 6            | 3             |
| <i>Dimophyes arctica</i>        | DA           | 2019. 10    | R/V ISABU    | 0       | 15           | 1             |
| <i>Diphyes bojani</i>           | DB           | 2018. 08    | R/V ISABU    | 9       | 6            | 2             |
| <i>Diphyes chamissonis</i>      | DC           | 2018. 09    | R/V DONGBAEK | 10      | 12           | 1             |
|                                 |              | 2018. 09    | R/V DONGBAEK | 11      | 4            | 4             |
| <i>Diphyes dispar</i>           | DD           | 2017. 10    | R/V ISABU    | 11.5    | 5            | 3             |
| <i>Eudoxoides mitra</i>         | EM           | 2020. 05    | R/V ISABU    | S12     | 5            | 3             |
| <i>Eudoxoides spiralis</i>      | ES           | 2020. 05    | R/V ISABU    | S12     | 8            | –             |
|                                 |              | 2018. 08    | R/V ISABU    | 6       | 8            | –             |
|                                 |              | 2018. 08    | R/V ISABU    | 8       | 2            | 1             |
|                                 |              | 2019. 10    | R/V ISABU    | Eddy1   | 1            | –             |
|                                 |              | 2019. 10    | R/V ISABU    | Eddy3   | 6            | 1             |
| <i>Lensia cossack</i>           | LC           | 2019. 10    | R/V ISABU    | Eddy1   | 8            | 1             |
| <i>Muggiaea atlantica</i>       | MA           | 2018. 04    | R/V DONGBAEK | 14      | 25           | 3             |
| <i>Hippopodius hippopus*</i>    | HH           | 2019. 10    | R/V ISABU    | Eddy3   | –            | 1             |

**Table 2.** Collecting information and the number of specimens used in the experiment to obtain DNA sequences and protein mass spectra. \* Outgroup.

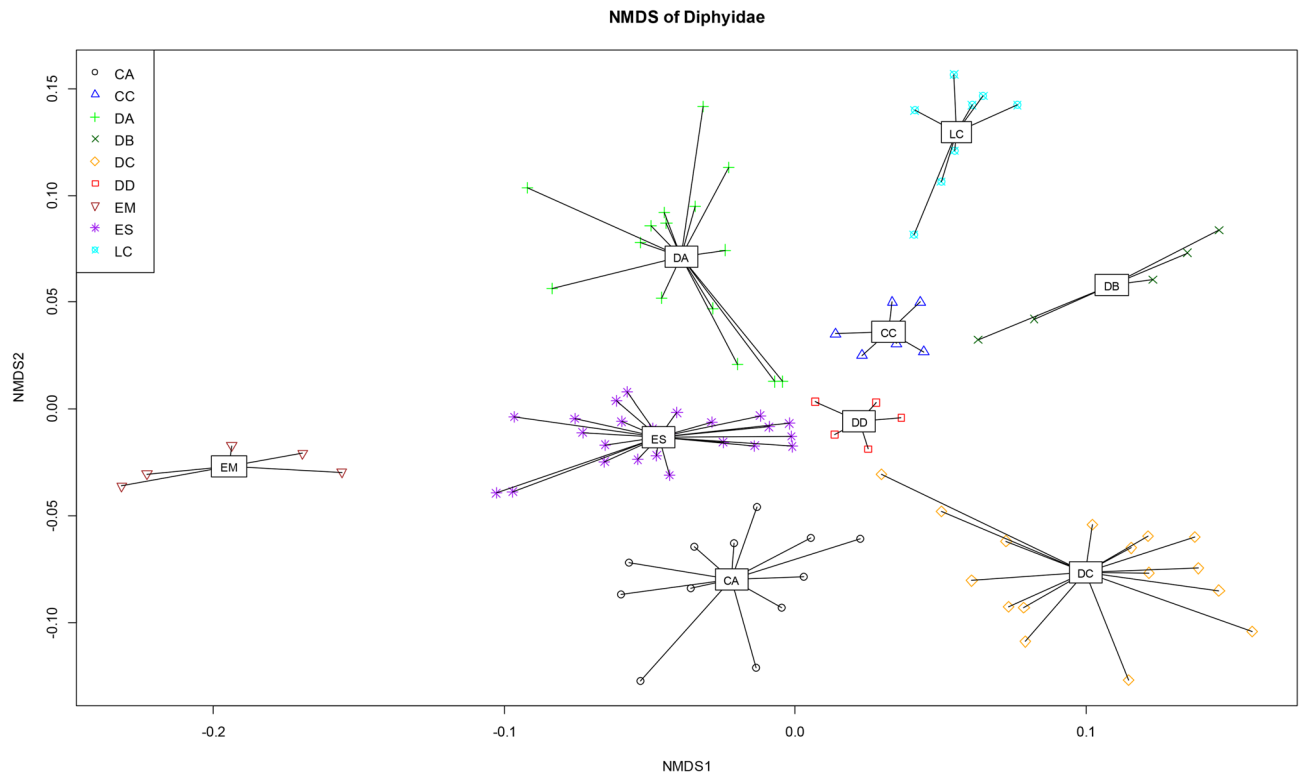


**Figure 3.** Peak pattern plot of the protein mass spectra of ten Diphyids. X-axis: representative peaks with the highest variation among 123 spectra; Y-axis: number of specimens are displayed in 4 intervals.



**Figure 4.** Cluster dendrogram with the  $p$  values (%) of nine Diphyids (98 spectra). The number of each node dictates the AU (Red) and BP (Green).

**Clustering.** Cluster dendrograms with  $p$  values (%) were obtained with two datasets using a Hellinger-transformed matrix. Dataset 1 included nine species (98 spectra) (Fig. 4) excluding *Muggiaea atlantica* Cunningham, 1892, and Dataset 2 contained ten species (123 spectra) (Supplementary Fig. S4 online) including *M. atlantica*. In Dataset 1, all nine species were divided into individual clusters. Dataset 2 was divided into clusters except for *Chelophyes appendiculata* (Eschscholtz, 1829). The spectra of *M. atlantica* were classified as the outermost cluster. Neither dataset was separated at the genus level, but *E. mitra* and *E. spiralis* clustered closely. These results were consistent with those confirmed in Supplementary Fig. S3 online and were supported by high AU  $p$  values

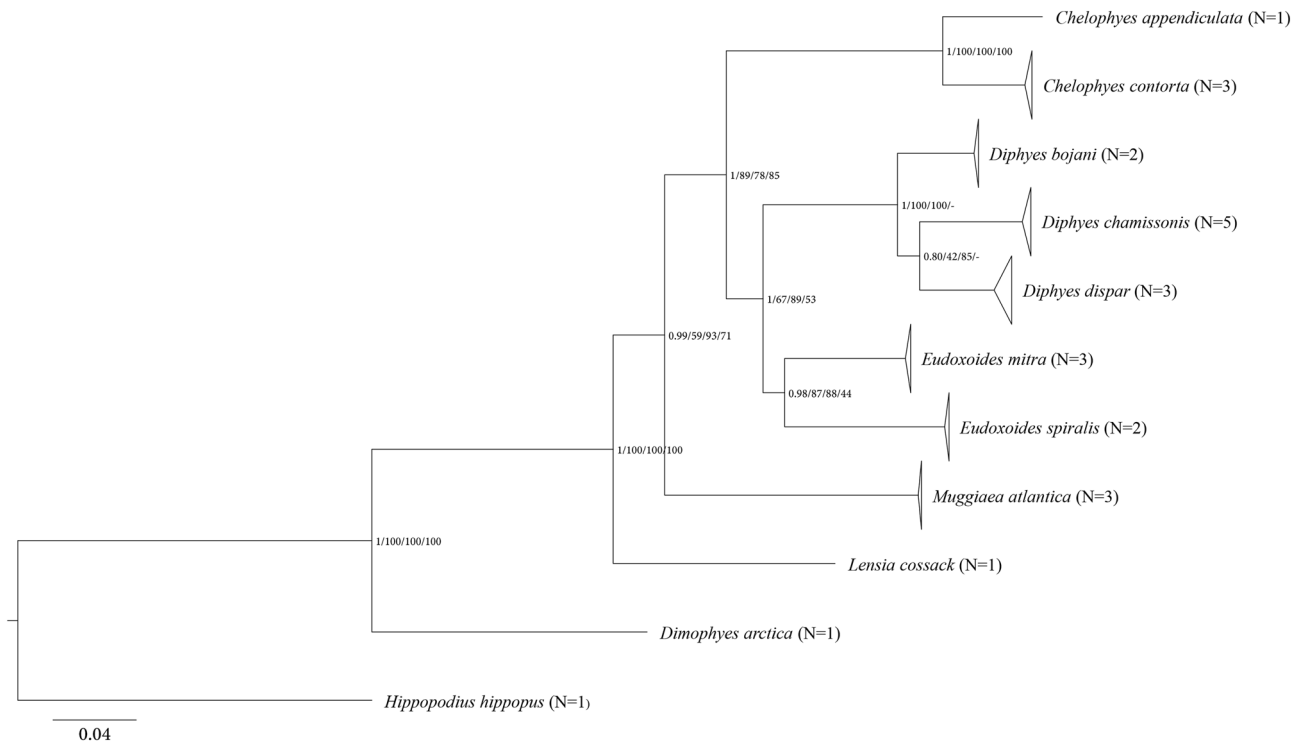


**Figure 5.** NMDS plot of the protein mass spectra of nine Diphyids from the Hellinger-transformed matrix (98 spectra).

and rather low BP  $p$  values. The BP  $p$  values are at risk of being significantly lower if data are skewed to one side, while the AU  $p$  values are more reliable because it is more improved and unbiased<sup>15</sup>.

**Non-metric multi-dimensional scaling (NMDS).** The NMDS analysis was performed based on the Hellinger-transformed matrix of two datasets (Dataset 1-Fig. 5; Dataset 2-Supplementary Fig. S5 online) covered in the Clustering section above. In Dataset 1, the spectra formed groups of the same species regardless of collection year, location, and storage period (homogeneity test:  $p < 0.05$ , PERMANOVA test:  $p = 0.001$ ). The number of spectra grew respective to the range of the group, but there was no overlap with other species. In Dataset 2, *M. atlantica* was separated clearly, consistent with the clustering results. However, except for *E. mitra*, *Diphyes bojani* (Eschscholtz, 1825), and *Lensia cossack* Totton, 1941, the remaining six species overlapped (homogeneity test:  $p < 0.05$ , PERMANOVA test:  $p = 0.001$ ). The stress values shown in Fig. 5 and Supplementary Fig. S5 online were 0.2004 and 0.1706, respectively, which were in a usable range according to Clarke, who suggested that stress values  $< 0.20$  were usable<sup>16</sup>. Considering that our dataset contained 98 and 123 samples of different variables, and since the stress values increased with numbers of samples and variables, the stress values were acceptable.

**Phylogenetic analysis.** A total of 100 new partial sequences for 24 specimens of ten Diphyidae species and a specimen of outgroup species were obtained successfully using four markers (mtCOI, 16S rRNA, 18S rRNA, ITS regions) (Table 2). For five datasets (Concatenated: ~4117 bp, mtCOI: ~916 bp, 16S rRNA: ~691 bp, 18S rRNA: ~1803 bp, ITS regions: ~797 bp), K2P genetic distances (Supplementary Table S1–5 online) were calculated, and phylogenetic trees were constructed using four algorithms (BI, NJ, ML, and MP). There were differences between the topologies according to phylogenetic tree construction algorithm. In the concatenated tree, the genera of Diphyidae formed each monophylum (Fig. 6) (BI  $> 0.99$ ; NJ  $> 59$ ; ML  $> 67$ ; MP  $> 53$ ). The most distantly related species in Diphyidae was *Dimophyes arctica* (Chun, 1897) (BI = 1; NJ = 100; ML = 100; MP = 100), followed by *L. cossack* (BI = 1; NJ = 100; ML = 100; MP = 100) and *M. atlantica* (BI = 0.99; NJ = 59; ML = 93; MP = 71). The concatenated dataset provided more well-resolved and supported relationships than did the single marker dataset. The topology of the mtCOI-based phylogeny (Supplementary Fig. S6 online) formed was monophyletic by genus, as in the concatenated tree. However, this phylogenetic tree differed from the concatenated tree (Fig. 6) in that *M. atlantica* and *L. cossack* formed the monophyletic branch (BI = 0.98; NJ = 62; ML = 60; MP = 37) and the location of *Chelophyes* Totton, 1932 (BI = 1; NJ = 91; ML = 78; MP = 85). The topology of the phylogenetic tree based on 16S rRNA (Supplementary Fig. S7 online) appeared paraphyly in the genus *Eudoxoides* Huxley, 1859. In addition, the topology in *Diphyes* Cuvier, 1817 (BI = 1; NJ = 100; ML = 96; MP = 98/BI = 1; NJ = 95; ML = 90; MP = 87) differed from that of the concatenated tree (Fig. 6). The phylogenetic tree based on 18S rRNA (Supplementary Fig. S8 online) failed to recover species within the genus *Chelophyes*, *Diphyes*, and *Eudoxoides*. In fact, the distance results for 18S rRNA (Supplementary Table S4 online) showed



**Figure 6.** Molecular phylogenetic tree of Diphyidae species based on 25 concatenated sequences. All positions containing gaps and missing data were eliminated. The number of each node dictates the BI/NJ/ML/MP bootstrap values. 'N' indicates the number of each sequence obtained in this study.

little difference within 0.005. Nevertheless, as shown in the previous trees, the genera *Dimophyes* Moser, 1925, *Muggiaea* Busch, 1851, and *Lensia* Totton, 1932 were separated (BI = 1; NJ = 100; ML = 100; MP = 98/BI = 0.64; NJ = 52; ML = 66; MP = 57). The phylogenetic tree based on ITS regions (Supplementary Fig. S9 online) showed paraphyly in *Eudoxoides*, like the tree of 16S rRNA. On the other hand, there were differences from the other phylogenetic trees in terms of the locations of *Muggiaea* and *Lensia* (NJ = 79; MP = 92/NJ = 77; MP = 41).

## Discussion

Diphyidae is the only family of the Calycophore Siphonophorans with distinctive bullet-shaped anterior nectophore, making it morphologically unique. However, the species of Diphyidae share many similar morphological features, making it difficult to distinguish species and genera. In fact, the initially described species of Diphyidae were all classified as the genus *Diphyes* and later were split into 45 species of eight genera based on minor differences in features such as ridge number, depth of hydroecium, length of somatocyst, and ostial teeth. Currently, the largest genus within the family is *Lensia*, with 26 species<sup>2</sup>, while the genus *Dimophyes* only had one species recorded. In this study, ten species belonging to six genera of Diphyidae were targeted. One major difficulty with taxonomic analysis of gelatinous Diphyids is that their body consists of >90% water, and they are immediately dehydrated when fixed with ethanol. This shrivels their body shape, making it difficult to identify key morphological features, increasing the likelihood of misclassification. In a previous Park & Lee study<sup>17</sup>, they used formalin as a fixative for studying *Chelophyes* and *Eudoxoides*, and the formalin-fixed samples were superior in terms of morphology compared to the ethanol-fixed samples in this study (Fig. 1). Table 1 shows the key morphological features that can be observed even after ethanol fixation. Among them, features including hydroecium, ostial teeth, mouth plate, and somatocyst were relatively easy to observe. However, the boundaries of the longitudinal ridges for many species became unclear from folds induced by dehydration, making it difficult to discern species. In fact, *D. arctica* was the only species that was ridge-less with only folds present.

Based on our results, MALDI-ToF MS is a viable alternative to species identification for non-intact gelatinous ethanol-fixed specimens. The ten species targeted in this study showed different spectral patterns (Fig. 3, Supplementary Fig. S1 online), and similar peaks were observed among the same species, regardless of time, sampling location, and duration of storage (Supplementary Fig. S2A, B online, Table 2). These results were consistent with the results of Karger et al.<sup>18</sup>, in which protein mass spectra varied slightly depending on time of collection and sampling location, but not enough to skew species identification. Peak variation between the storage periods of three and 75 days for 70% ethanol-fixed samples was covered by Dvorak et al.<sup>6</sup>, but our study was the first to compare 99% ethanol-fixed samples that had been stored for more than 1 year. In addition, peaks were detected from 4-year-old samples. When MALDI-ToF was attempted with formalin-fixed samples, peaks below 3000 Da resembled those of the matrix only. In comparison to the ethanol-fixed *C. appendiculata*, the strong peak observed at 4920 Da also was observed in the formalin-fixed spectra but with considerably less intensity,

rendering it unusable (Supplementary Fig. S2C, D online). This is believed to be due to the cross-linking nature of the formalin-induced proteins<sup>19</sup>.

The Hellinger-transformed matrix was used for visualization, as per recommendation by Rossel & Martinez Arbizu<sup>20</sup> as the best transformation method for species identification. With the exception of *M. atlantica*, the dataset of the other nine species showed valid discrimination in both cluster analysis and NMDS (Fig. 4, 5). However, without prior morphological species identification, *C. appendiculata* from Fig. 4 could have been divided into two species. The two groups of *C. appendiculata* differed in sampling time and location. The peak pattern (Fig. 3) and raw spectra (Supplementary Fig. S2A, B online) were similar (especially 4925 Da), nonetheless, and even formed a single group in the NMDS plot (Fig. 5). The difference in detailed peaks appeared to have been greatly overestimated, causing the division into two groups. However, in agreement with aforementioned by Karger et al.<sup>18</sup>, the difference seemed minimal to skew or influence species identification. Granted, it is necessary to obtain and compare more samples with varying conditions (peaks, timing, etc.) to fully understand these variations.

No notable associations between the morphological features mentioned in Table 1 and the clusters (Figs. 4), but Fig. 4 showed that the blunt apex species (*D. arctica* and *L. cossack*) were bound to the same cluster. It was evident from cluster analysis and NMDS of the dataset containing ten species (Supplementary Fig. S4, 5 online) that *M. atlantica* was skewed to one side. The NMDS plot of the other species overlapped with one another, rendering them indistinguishable. This seemed likely due to the more accentuated *M. atlantica* showed peaks than those of the rest of the species, resulting in a Hellinger-transformed matrix more centralized toward *M. atlantica*. The peak pattern (Fig. 3) and top-ranking feature (Supplementary Fig. S3 online) showed significantly different peak expressions per species. Peaks shared by many species tended to have decreased t-scores of approximately -10, and peaks rarely found in other species, such as 3414–3416, 1752–1753, and 3331–3332 Da, showed high levels of expression. Supplementary Fig. S4, 5 online grouped the dataset of ten species with a large peak difference, resulting in the graph skewing to one side and resembling that of a phylogenetic tree with an incorrect outgroup. Therefore, it would be more advantageous to compare species by species for higher accuracy. Similar results were seen with DNA analysis. Our results showed that *M. atlantica* differed from the other species by 0.270–0.465 (mtCOI) and 0.257–0.432 (16S rRNA) (Supplementary Table S2, 3 online). According to Zheng et al.<sup>21</sup>, the intra-family mtCOI and 16S rRNA K2P distance of hydrozoan ranged from 0.104 to 0.248 (average = 0.192) and 0.073–0.287 (average = 0.198), respectively. In this context, the difference between *M. atlantica* and other species was beyond the family level. Furthermore, phylogenetic analysis showed that *M. atlantica* was the third-most distant in terms of relationship within Diphyidae (Fig. 6; Supplementary Fig. S6–8 online). It is plausible that the difference in genetic distance may have affected protein expression and pattern differences. However, *D. arctica* was far more genetically distant than *M. atlantica*, but the protein patterns did not reflect this. Based on this result, it cannot be said that protein peak patterns directly reflect phylogenetic relationship. In fact, clustering results were not separated at the genus level with the exception of *Eudoxoides* nor were the topologies of the cluster analysis and phylogeny tree consistent with one another because the MALDI-ToF MS method does not account for composition or molecular function of proteins, but instead simply compares the pattern of the protein peaks<sup>22</sup>. Granted, as mentioned earlier, a difference in genetic distance can be reflected in protein expression, just not in proportion to genetic distance.

Peter et al. showed mtCOI NJ trees including six (*D. bojani*, *D. chamissonis*, *D. dispar*, *E. mitra*, *E. spiralis*, *M. atlantica*) of our target species<sup>23</sup>. The same topologies were supported in the phylogenetic trees of this study, further extending phylogenetic locations of four species (*C. appendiculata*, *Chelophyes contorta* (Lens & van Riemsdijk, 1908), *D. arctica*, *L. cossack*) based on a total of four gene concatenated sequences using four algorithms (BI, NJ, ML, MP) (Fig. 6; Supplementary Fig. S6–99 online). Each dataset consisted of the same specimens, all of which formed a monophyletic group with clear discrimination. However, depending on the marker and algorithm used, the topology at the genus level differed. In addition, the phylogenetic tree produced with 18S rRNA (Supplementary Fig. S8 online) could not distinguish the phylogenetic position of the genera *Chelophyes*, *Diphyes*, and *Eudoxoides*. This result was consistent with the results of a previous study<sup>24</sup>, which suggested that the 18S rRNA marker was not useful in species differentiation at the genus or species level.

The association between the phylogenetic tree and morphological features was noted in genus-specific characteristics (Claw-shaped hydroecium in *Chelophyes*; Ostial teeth in *Diphyes*). Species with claw-shaped hydroecium and ostial teeth characteristics were monophyletic, respectively. It also confirmed that *D. arctica*, without longitudinal ridges, and another nine species were divided distinctly and phylogenetically. In concatenated and 16S rRNA (Fig. 6, Supplementary Fig. S7 online), eight species that had a pointed apex were established as the monophyletic branch, and the blunt apex species, *D. arctica* and *L. cossack*, were separated. This was similar to what appeared in Fig. 4 (the aforementioned binding of *D. arctica* and *L. cossack*). Furthermore, the results of the phylogenetic analysis suggested that *D. arctica* and *L. cossack* were located at the outermost location within Diphyidae, close to the nearest ancestor. Taken comprehensively, these results indicated that the ancestral Diphyidae initially had neither ridges nor an apex and subsequently evolved to become more angular, with ridges forming, and apex becoming pointed. Indeed, Abylidae L. Agassiz, 1862, Siphonophorans, known to have many angles, were of evolutionary descendant and were located in the innermost phylogenetic position of Diphyidae<sup>25</sup>. However, at present this conviction is incomplete, and we need to obtain more information about a wider variety of species.

We have for the first time confirmed that MALDI-ToF can be a valid method of species identification for the morphologically “difficult” group Diphyidae. This method used in conjunction with morphological analysis and DNA information helped to ensure efficient and accurate species identification. There are still some limitations of using this method alone. First, MALDI-ToF analysis cannot yield any insights regarding phylogenetic relationships. It is still noteworthy that *E. mitra* and *E. spiralis* were clustered closely and showed similar peak expressions in the top ranks despite significant differences in peak pattern (Supplementary Fig. S1, 3 online). We

suspected that there might be differences between the representative peaks selected from matrices and the real peaks. Based on this finding, it was expected that the absence of phylogenetic analysis could be supplemented by genus-specific peaks. Second, the lack of existing data on Diphyidae spectra limited the methodology to application as a supplement to morphological and DNA analyses. In comparison to DNA barcoding, MALDI-ToF MS is cheaper and less time consuming, with a higher sample throughput. Therefore, it is expected that fast accumulation of data will be possible. For this to be possible, it will be necessary to establish a shared database system among researchers, such as the NCBI. Finally, organisms, such as Diphyidae, which are polymorphic by generational rotation or colony formation, might differ in protein expression depending on life stage. To address this, researchers in future studies will have to focus on profiling data for all stages of a species based on spectra of both posterior nectophore and eudoxids.

## Conclusions

For the first time, we applied a fast and accurate using MALDI-ToF MS method for identification of ten species of Diphyidae that are morphologically difficult to discriminate, and assessed the potential for use. Protein mass spectra showed clear differences between species and confirmed the validity of the identification tool. Ethanol-but not formalin-fixed samples should be used, and old ethanol-fixed samples can also be used because they are not affected by storage period. This novel method has breakthroughs for cost and time. In addition, the morphological characteristics and partial sequences of four genes (mtCOI, 16S rRNA, 18S rRNA, ITS regions) were obtained, and the associations were compared to each other. The tools and reference data established in this study can be used as a fundamental source for multilateral understanding of Siphonophore.

## Methods

**Sample collection.** The samples used in this study were collected in 2017–2020, during a, R/V ISABU cruise in the Northwest Pacific Ocean area affected by the Kuroshio Current, and in 2018, during an R/V DONGBAEK cruise in the Jeju and along the southern coast of Korea (Fig. 1). On the R/V ISABU, a Multiple Opening/Closing Net and Environmental Sensing System (MOCNESS,  $1 \times 1 \text{ m}^2$ , mesh size: 200  $\mu\text{m}$ ) was towed between a depth of 200 m and the surface. On the R/V DONGBAEK, we towed a plankton net (mesh size: 200  $\mu\text{m}$ ,  $\Phi$ : 60 cm) vertically from the bottom to the surface. The mouth of the net was equipped with a flowmeter (Hydro-Bios, Germany) to determine the volume of filtered water during each tow. Immediately after towing, we split the samples into 1/2 aliquots using a Folsom plankton splitter. Each aliquot was fixed with neutralized 5% formalin solution or 99% ethanol. In this study, we only used ethanol-fixed samples.

**Morphological analysis.** Siphonophore specimens were sorted from 99% ethanol samples using Live Insect Forceps (26029-10, Fine Science Tools Inc., Foster City, CA, USA) under a stereomicroscope (Olympus SZX7, Tokyo, Japan). We identified specimens using descriptions, illustrations, and terminology from the literature<sup>2,26–29</sup>. For morphological records, digital photographs of specimens were produced at various focal points using a digital camera (Olympus PEN Lite E-PL3, Tokyo, Japan) connected to the stereomicroscope with side lights on the dark field. The obtained multi-focus photos were stacked using Helicon Focus 7 software (Version: 7.5.1). Afterward, only the object was cropped and moved to a black background using Adobe Photoshop CS6 software (Version: 13.0).

**MALDI-ToF MS.** Part of the gelatinous substance of each individual separated with Live Insect Forceps or the entire individual were put into a 1.5 ml tube to prepare for protein mass measurement. After evaporating the ethanol, the samples were dried in a vacuum dryer for 30 min. 10–20  $\mu\text{l}$  of  $\alpha$ -Cyano-4-hydroxy-cinnamic acid (HCCA) matrix (Acetonitrile 50%; Ultra-pure water 47.5%; Trifluoroacetic acid 2.5%; supersaturated HCCA (30 mg for a total of 1 ml of matrix)) was added to each tube and incubated at room temperature for at least 30 min. Matrix-specimen solutions were centrifuged at 12,000 rpm for 1 min. 2  $\mu\text{l}$  of the supernatant was loaded onto the target plate of the MALDI-ToF MS equipment (AXIMA Confidence MALDI ToF-Mass Spectrometer; Shimadzu), and the solutions were air-dried completely to crystallize them. Protein mass spectra were measured in the range of 1–20 k Dalton on MALDI-MS Application Launchpad 2.9.2 (Shimadzu Biotech) software using positive-ion linear mode with a laser power of 80. For each loading spot, 100 profiles were repeated ten times and summed into one spectrum. Each protein mass spectrum was exported in ASCII format and imported to Data Explorer version 4.5 software for range trimming to 1.5–20 kDa.

**Protein mass spectra analysis.** Data processing and analysis of protein mass spectra were conducted using MALDIquantForeign (ver. 0.12)<sup>30</sup>, MALDIquant (ver. 1.19.3)<sup>31</sup>, and MALDIrppa (ver. 1.0.5)<sup>32</sup> packages in the RStudio program (version: 1.4.1103, RStudio Inc., Boston, MA, USA; R version: 4.0.5). The workflow in R referred to manual<sup>33</sup> and literature<sup>22</sup>. Spectra were proceeded square root transforming and smoothing (Savitzky-Golay method), baseline removing (SNIP baseline estimation method), normalizing (Total-Ion-Current-calibration (TIC) method) implemented in MALDIquant. Significant peaks were detected with a signal-to-noise ratio (SNR) of 6 and a half window size of 7. Peaks were binned repeatedly with the “binpeaks” command from MALDIquant, with a tolerance of 0.001 and a minimum frequency of 0.05. The feature matrix obtained through this process was Hellinger-transformed and visualized through cluster analysis, diagonal discriminant analysis (DDA), and non-metric multi-dimensional scaling (NMDS) analysis.

The cluster dendrogram was established with the pvclust (ver. 2.2) R package<sup>15</sup> using Ward's 2D clustering algorithm with Euclidean distances and 10,000 bootstrapping replications. The approximately unbiased (AU) *p* values and bootstrap probability (BP) values were calculated in the dendrogram. DDA was performed with the sda<sup>34</sup> function to find the peaks with the highest variation among the Diphyidae species and to calculate

the t-score for feature ranking. An NMDS plot was created with the vegan (ver. 2.5.6) R package<sup>35</sup> based on the Bray–Curtis Dissimilarity distance with  $k=2$ . ANOVA was conducted utilizing the betadisper function provided with the vegan package to test the multivariate homogeneity of group dispersion. PERMANOVA was performed using the Adonis tool to test the fit of the data with 999 permutations. In addition, a plot using the ‘peakPatterns’ function of MALDIrppa (ver. 1.0.5) was obtained and simplified to a minimum frequency of 0.8.

**DNA extraction, amplification, and sequencing.** For DNA extraction, the remaining parts of the individuals used to measure the protein mass were transferred to ultra-pure water for three hours to wash and remove the ethanol. A LaboPass™ Genomic Isolation-Tissue miniprep kit (Cosmogenetech Co., Seoul, Korea) was used to extract genomic DNA following the manufacturer’s protocols. We amplified four genetic markers: mitochondrial cytochrome c oxidase subunit I (mtCOI); 16S ribosomal RNA (16S rRNA); 18S ribosomal RNA (18S rRNA); and Internal transcribed spacer regions (ITS regions) via polymerase chain reaction (PCR) using the AccuPower® PCR PreMix (Bioneer Co., Daejeon, Korea), and performed thermal cycling using a TaKaRa Thermal Cycler Dice Touch TP350 (Takara Bio Inc., Kusatsu, Japan). PCR products consisted of 5 µl premix, 15 µl ultra-pure water, 3 µl DNA template, and 1 µl each of the forward and reverse primers to achieve a 25 µl total reaction volume per tube.

The MtCOI genes (about 700–800 bp) were amplified using modified jgLCO1490 & jgHCO2198 primers with thermo-cycling conditions: initial 5 min at 95 °C, followed by 42 cycles of 20 s at 95 °C, 1 min at 48 °C, and 1 min at 72 °C, and ending with a final 5 min at 72 °C referenced and modified from Geller et al.<sup>36</sup>. COF & COR primers were used with the following thermo-cycling conditions: 5 cycles of 50 s at 94 °C, 50 s at 45 °C, and 2 min at 70 °C, followed by 30 cycles of 50 s at 94 °C, 50 s at 50 °C, and 2 min at 68 °C<sup>37</sup>. 16S rRNA (about 600 bp) genes were amplified using SHA & SHB primers with the following thermo-cycling conditions: 30 cycles of 20 s at 94 °C, 45 s at 50 °C, and 2 min at 68 °C<sup>37</sup>. 18S rRNA (about 1800 bp) genes were amplified using EukA & EukB primers with the following thermo-cycling conditions: 30 cycles of 10 secs at 94 °C, 1 min at 38 °C, and 3 min at 72 °C, and a final 2 min at 94 °C<sup>38,39</sup>. ITS regions (about 750 bp) genes were amplified using IFS & IRS primers with the following thermo-cycling conditions: 24 cycles of 20 secs at 94 °C, 45 s at 51 °C, and 1 min 30 s at 72 °C<sup>37</sup>.

Amplifications were confirmed by electrophoresis on a 1% agarose gel (AGAROSE I™, Amresco Inc., Solon, OH, USA; LaboPass™ Buffer 50X TAE, Cosmogenetech Co., Seoul, Korea) with Staining STAR (Dynebio Inc., Seongnam, Korea) for 20 min at 100 V with a 100 bp DNA ladder (Bioneer Co., Daejeon, Korea). PCR products were purified using a LaboPass™ PCR Purification Kit (Cosmogenetech Co., Seoul, Korea) following the manufacturer’s protocols. Purified PCR products were sent to Bionics Inc. (Seoul, Korea) for DNA sequencing. For sequencing, an ABI automatic capillary sequencer was used with the same set of primers as used for amplification. Due to its long length, the internal primer also was used for the 18S rRNA sequencing<sup>40</sup>. All obtained sequences were visualized using Finch TV software (ver. 1.4.0) (<https://digitalworldbiology.com/FinchTV>; Geospiza Inc., USA). The quality of each sequence was evaluated, and low-resolution peaks were checked by comparing forward and reverse strands. BLAST<sup>41</sup> search confirmed the obtained sequences as Siphonophores without contaminants. Sequence information from this study was deposited in the GenBank database (MZ230437–MZ230486, MZ230526–MZ230550, MZ292030, MZ292031, MZ292870–MZ292892).

**Phylogenetic analysis.** Each dataset consisted of sequences of 24 specimens of ten Diphyids species. We used a specimen of *Hippopodius hippopus* (Forsskål, 1776) belonging to the family Hippopodiidae Kölliker, 1853, as an outgroup. Sequences were aligned using the ClustalW algorithm<sup>42</sup> with default parameters embedded in Molecular Evolutionary Genetics Analysis 7 software (MEGA7, Version: 7.0.26)<sup>43</sup>. The genetic distances between alignment sequences were calculated using the Kimura 2-parameter (K2P) model<sup>44</sup> with complete deletion by MEGA7 software.

Phylogenetic analyses were performed using the Neighbor-Joining (NJ), Maximum Parsimony (MP), Maximum Likelihood (ML), and Bayesian Inference (BI) approaches. The best-fit evolutionary model for phylogenetic analysis was calculated using jModelTest software (Version: 2.1.7)<sup>45,46</sup> with the Akaike information criterion (AIC)<sup>47–49</sup> (Concatenated: GTR + I + G; mtCOI: GTR + I + G; 16S rRNA: TVM + I + G; 18S rRNA: TIM2 + I; ITS regions: TIM2 + I + G). Gap calibration was conducted using FastGap (Version: 1.2)<sup>50,51</sup>. NJ, MP, and ML analyses were performed using MEGA7 software, PAUP4 software<sup>52</sup>, and IQTree web server<sup>53</sup>, respectively, with 1000 bootstrapping replicates for phylogenetic tree reconstruction<sup>54</sup>. The BI tree was constructed using MrBayes software (Version: 3.2.6)<sup>55–57</sup> based on the following model parameters. Markov Chain Monte Carlo (MCMC) was run with the following parameters: ngen = 1,000,000, nchains = 4, samplefreq = 100, savebrlens = yes, and printfreq = 1000. The BI trees were constructed using the “sump” command with burnin = 250 to summarize the parameters and the “sumt” command with burnin = 250 to summarize the tree. All trees were visualized using FigTree (version: 1.4.2). The concatenated dataset of ML was partitioned and applied to the model according to each marker. For NJ, MP, and ML, no separate partitioning was performed on the concatenated dataset.

### Data availability

The datasets generated and/or analyzed during the current study are available in Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>) and this published article (and its supplementary materials).

Received: 3 June 2021; Accepted: 14 September 2021

Published online: 24 September 2021

## References

- Pugh, P. The vertical distribution of the siphonophores collected during the SOND cruise, 1965. *J. Mar. Biol. Assoc. U.K.* **54**, 25–90 (1974).
- Grossmann, M. M., Collins, A. G. & Lindsay, D. J. Description of the eudoxid stages of *Lensia havock* and *Lensia leloupi* (Cnidaria: Siphonophora: Calycophorae), with a review of all known *Lensia* eudoxid bracts. *Syst. Biodivers.* **12**, 163–180 (2014).
- Dunn, C. W. & Wagner, G. P. The evolution of colony-level development in the Siphonophora (Cnidaria: Hydrozoa). *Dev. Genes. Evol.* **216**, 743–754 (2006).
- Karas, M. & Hillenkamp, F. Laser desorption/ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal. Chem.* **60**, 2299–2301 (1988).
- Clark, A. E., Kaleta, E. J., Arora, A. & Wolk, D. M. Matrix-assisted laser desorption/ionization–time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology. *Clin. Microbiol. Rev.* **26**, 547–603 (2013).
- Dvorak, V. *et al.* Identification of phlebotomine sand flies (Diptera: Phlebotomidae) by matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Parasit. Vectors* **7**, 1–7 (2014).
- Rossel, S. & Arbizu, P. M. Revealing higher than expected diversity of Harpacticoida (Crustacea: Copepoda) in the North Sea using MALDI-TOF MS and molecular barcoding. *Sci. Rep.* **9**, 1–14 (2019).
- Feltens, R., Görner, R., Kalkhof, S., Gröger-Arndt, H. & von Bergen, M. Discrimination of different species from the genus *Drosophila* by intact protein profiling using matrix-assisted laser desorption/ionization mass spectrometry. *BMC Evol. Biol.* **10**, 1–15 (2010).
- Mazzeo, M. F. *et al.* Fish authentication by MALDI-TOF mass spectrometry. *J. Agric. Food Chem.* **56**, 11071–11076 (2008).
- Kaufmann, C., Schaffner, F., Ziegler, D., Pflueger, V. & Mathis, A. Identification of field-caught *Culicoides* biting midges using matrix-assisted laser desorption/ionization time of flight mass spectrometry. *Parasitology* **139**, 248–258 (2012).
- Laakmann, S. *et al.* Comparison of molecular species identification for North Sea calanoid copepods (Crustacea) using proteome fingerprints and DNA sequences. *Mol. Ecol. Resour.* **13**, 862–876 (2013).
- Lassen, S., Wiebring, A., Helmholz, H., Ruhnau, C. & Prange, A. Isolation of a Nav channel blocking polypeptide from *Cyanea capillata* medusae—a neurotoxin contained in fishing tentacle isorhizas. *Toxicon* **59**, 610–616 (2012).
- Lazcano-Pérez, F., Arellano, R. O., Garay, E., Arreguín-Espinosa, R. & Sánchez-Rodríguez, J. Electrophysiological activity of a neurotoxic fraction from the venom of box jellyfish *Carybdea marsupialis*. *Comp. Biochem. Physiol. C: Toxicol. Pharmacol.* **191**, 177–182 (2017).
- Helmholz, H., Naatz, S., Lassen, S. & Prange, A. Isolation of a cytotoxic glycoprotein from the Scyphozoa *Cyanea lamarckii* by lectin-affinity chromatography and characterization of molecule interactions by surface plasmon resonance. *J. Chromatogr. B* **871**, 60–66 (2008).
- Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
- Clarke, K. R. Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* **18**, 117–143 (1993).
- Park, N. & Lee, W. Four new records of family Diphyidae (Hydrozoa: Siphonophorae) in Korean waters. *J. Spec. Res.* **9**, 131–146 (2020).
- Karger, A., Bettin, B., Gethmann, J. M. & Klaus, C. 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 (2019).
- Tayri-Wilk, T. *et al.* Mass spectrometry reveals the chemistry of formaldehyde cross-linking in structured proteins. *Nat. Commun.* **11**, 1–9 (2020).
- Rossel, S. & Martínez Arbizu, P. Effects of sample fixation on specimen identification in biodiversity assemblies based on proteomic data (MALDI-TOF). *Fron. Mar. Sci.* **5**, 1–13 (2018).
- Zheng, L., He, J., Lin, Y., Cao, W. & Zhang, W. 16S rRNA is a better choice than COI for DNA barcoding hydrozoans in the coastal waters of China. *Acta Oceanol. Sin.* **33**, 55–76 (2014).
- Yeom, J., Park, N., Jeong, R. & Lee, W. Integrative description of cryptic *Tigriopus* species from Korea using MALDI-TOF MS and DNA barcoding. *Front. Mar. Sci.* **8**, 495 (2021).
- Peter, S. Molecular characterization and phylogenetic analysis of *Diphyes dispar* (Siphonophora: Diphyidae) from the Laccadive Sea, off the south-west coast of Arabian Sea, Indian Ocean. *Int. J. Fish Aquat. Stud.* **4**, 30–35 (2016).
- Dunn, C. W., Pugh, P. R. & Haddock, S. H. Molecular phylogenetics of the Siphonophora (Cnidaria), with implications for the evolution of functional specialization. *Syst. Biol.* **54**, 916–935 (2005).
- Munro, C. *et al.* Improved phylogenetic resolution within Siphonophora (Cnidaria) with implications for trait evolution. *Mol. Phylogenet. Evol.* **127**, 823–833 (2018).
- Totton, A. K. Siphonophora of the Indian Ocean together with systematic and biological notes on related specimens from other oceans. *Disc Rep* **27**, 1–162 (1954).
- Totton, A. K., Bargmann, H. E. & British Museum (Natural History). *A synopsis of the Siphonophora*. (British Museum (Natural History), 1965).
- Mapstone, G. M. *Siphonophora (Cnidaria, Hydrozoa) of Canadian Pacific waters*. (NRC Research Press, 2009).
- Nishiyama, E. Y., Araujo, E. M. & Oliveira, O. M. Species of *Lensia* (Cnidaria: Hydrozoa: Siphonophorae) from southeastern Brazilian waters. *Zoologia (Curitiba)* **33**, 2337 (2016).
- MALDIquantForeign: Import/Export routines for MALDIquant (2015).
- Gibb, S. & Strimmer, K. MALDIquant: a versatile R package for the analysis of mass spectrometry data. *Bioinformatics* **28**, 2270–2271 (2012).
- Palarea-Albaladejo, J., Mclean, K., Wright, F. & Smith, D. G. MALDIrppa: quality control and robust analysis for mass spectrometry data. *Bioinformatics* **34**, 522–523 (2018).
- Gibb, S. & Strimmer, K. *Species Identification using MALDIquant manual*. <http://www.strimmerlab.org/software/malldiquant/>. (2015).
- Ahdesmäki, M. & Strimmer, K. Feature selection in omics prediction problems using cat scores and false nondiscovery rate control. *Ann. Appl. Stat.* **4**, 503–519 (2010).
- vegan: Community Ecology Package. R package version 2.5-2. 2018 (2018).
- Geller, J., Meyer, C., Parker, M. & Hawk, H. Redesign of PCR primers for mitochondrial cytochrome c oxidase subunit I for marine invertebrates and application in all-taxa biotic surveys. *Mol. Ecol. Resour.* **13**, 851–861 (2013).
- Schuchert, P. DNA barcoding of some Pandidae species (Cnidaria, Hydrozoa, Anthoathecata). *Rev. Suisse Zool.* **125**, 101–127 (2018).
- Medlin, L., Elwood, H. J., Stickel, S. & Sogin, M. L. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. *Gene* **71**, 491–499 (1988).
- Collins, A. G. Towards understanding the phylogenetic history of Hydrozoa: hypothesis testing with 18S gene sequence data. *Scientia Marina* (2000).
- Strychar, K. B., Hamilton, L. C., Kenchington, E. L. & Scott, D. B. *Cold-water Corals and Ecosystems* 679–690 (Springer, Berlin, 2005).

41. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
42. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
43. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
44. Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
45. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* **9**, 772–772 (2012).
46. Guindon, S. & Gascuel, O. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Syst. Biol.* **52**, 696–704 (2003).
47. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974).
48. Yamaoka, K., Nakagawa, T. & Uno, T. Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations. *J. Pharmacokinet. Biopharm.* **6**, 165–175 (1978).
49. Hurvich, C. M. & Tsai, C.-L. Regression and time series model selection in small samples. *Biometrika* **76**, 297–307 (1989).
50. Simmons, M. P. & Ochoterena, H. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**, 369–381 (2000).
51. Young, N. D. & Healy, J. GapCoder automates the use of indel characters in phylogenetic analysis. *BMC Bioinform.* **4**, 6 (2003).
52. Swofford, D. L. & Sullivan, J. Phylogeny inference based on parsimony and other methods using PAUP\*. *Phylogenet. Handb. Pract. Approach DNA and Protein Phylogeny* **7**, 160–206 (2003).
53. Trifunopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).
54. Felsenstein, J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
55. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
56. Ronquist, F. & Huelsenbeck, J. P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574 (2003).
57. Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

## Acknowledgements

This study was supported by the Korea Institute of Ocean Science and Technology (KIOST, PE99888), Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1I1A2043807), and the National Institute of Biological Resources (NIBR), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIBR202130203). We would like to express sincere gratitude to the captains and crew members of the R/V ISABU and DONGBAEK for their help during the sampling process. In addition, we would like to express our deep appreciation to Prof. Jinwon Lee, Ms. Yourae Kim, and members of the Biodiversity Laboratory, Hanyang University, for advice on how to obtain the protein mass spectra, DNA analysis, etc.

## Author contributions

All authors contributed to the writing of the manuscript. W. L. designed the study; N. P. and J. Y. obtained specimens and data; and R. J. conducted R processing.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-98724-z>.

**Correspondence** and requests for materials should be addressed to W.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021