Figure 2 shows the variability of the dataset for two geographical attributes (latitude and longitude at the start of sampling (DD)), one climatic attribute (wind speed (m/s) at the start of sampling), and three environmental attributes (depth (m) at the start of sampling, temperature (°C) and water O_2 concentration (mg/L)). The violin diagrams present many of the same summary statistics as the box plots: the white dots designate the medians, the thickened black bars in the center represent the interquartile range, and the thin black lines on either side evoke the rest of the distribution, with the exception of the points that are considered "extreme". On either side of the black lines is an approximation of the kernel density to show the shape of the data distribution. The wider areas of the violin diagrams indicate a higher probability of the variables taking a given value; the thinner areas indicate a lower probability. These diagrams are essential for designing the conditions of the habitats where the samples were taken, as high variability in the data can indicate heterogeneous habitats, while low variability indicates more homogeneous habitats. In this way, these violin diagrams help to highlight unique habitats that may have an impact on Cumacea genetics. The mean, median, standard deviation, 1st and 3rd quartiles help identify general trends in the data, irregularities, or extreme values, and capture the diversity of environmental conditions.

Figure 2a shows a latitude range at the start of sampling from 60.357 to 67.868 DD. The median of this distribution (67.15 DD) is higher than the mean (64.83 DD), indicating an asymmetrical distribution towards lower values. This means that there are some low values pulling the mean down. Unlike the mean, the median is less affected by extreme values and provides an indication of the central disposition of the data. The standard deviation (3.17 DD) shows a moderate dispersion close to the mean (64.83 DD). Quartiles Q1 (61.64 DD) and Q3 (67.64 DD) reveal that the majority of data are clustered around the median (67.15 DD). The curve has an asymmetrical bimodal shape, showing two peaks, suggesting that the samples come from two dominant latitudinal regions at the start of sampling. This could indicate important fluctuations in climatic and environmental conditions in the regions sampled. Similarly, this type of curve is present for longitude (DD) at the start of sampling, as well as for temperature (°C) and O_2 concentration (mg/L) of the water from which the samples were taken (see Figures 2b, 2e and 2f).

Figure 2b shows a longitude distribution at the start of sampling from -31.356 to -12.162 DD. The median (-26.21 DD) is lower than the mean (-23.12 DD), suggesting an asymmetry on the side of the higher values. This means that a few high values pull the mean upwards. The standard deviation (5.52 DD) indicates a relatively wide data range. Quartiles Q1 (-26.77 DD) and Q3 (-18.14 DD) also show great variability in the longitude data at the start of sampling. This suggests a strong environmental gradient, geographical distribution and sample diversity from east to west in the study area.

Figure 2c shows that samples were taken at depths ranging from 316 to 2568 m. The median (1574.70 m) is notably higher than the mean (1412.57 m), showing an asymmetrical distribution in favor of the lowest values. The standard deviation (881.16 m) is quite high, indicating variability in sampling depths at the start of sampling and providing a more global overview of benthic habitats. Quartiles Q1 (579.10 m) and Q3 (2504.70 m) show and confirm a wide distribution of depths. Important fluctuations in standard deviation (881.16 m) may indicate that certain parameters are more variable and could possibly have an impact Cumacea genetics. The curve in this figure has a multimodal shape with three prominent peaks, suggesting that the samples were mainly collected and concentrated at three different depths (around 500, 1500 and 2500 m).

Figure 2d shows a range of wind speeds at the start of sampling, from 2 to 11 m/s. The standard deviation (2.16 m/s) shows a moderate dispersion of the data. Quartiles Q1 (5.25 m/s) and Q3 (7.00 m/s) suggest that most wind speeds fall within a fairly narrow distribution between 5.25 and 7.00 m/s. The mean (6.26 m/s) and median (6.00 m/s) are similar, with a high concentration of data around the median (6.00 m/s), indicating stable wind conditions at the start of sampling.

Figure 2e shows a temperature distribution where samples were taken between -0.851 and 4.28°C. The mean (1.45°C) is higher than the median (0.71°C), demonstrating an asymmetry towards the higher values. The standard deviation (1.73°C) is relatively high compared with the mean (1.45°C), indicating a wide range of data. Quartiles Q1 (0.07°C) and Q3 (2.66°C) also show a wide dispersion of water temperature values. These results suggest acclimatization of cumaceans to a variety of habitat temperatures.

Figure 2f shows a range of O_2 concentration in the water where samples were taken, from 245.53 to 292.97 mg/L. The median (278.77 mg/L) is higher than the mean (271.88 mg/L), showing an asymmetry towards the lowest values. Quartiles Q1 (258.39 mg/L) and Q3 (290.90 mg/L) show a certain variability in O_2 concentration data, also confirmed by the standard deviation (18.11 mg/L). These latest results reflect a diversity of O_2 requirements, with organisms adapted to low O_2 conditions and potentially influenced by the heterogeneity of biogeochemical cycles, such as photosynthesis, respiration, and organic decomposition, which have an impact on depth-dependent dissolved O_2 levels.

The distribution and diversity of the various cumacean species found in our sample are presented in Figure 3. It shows the most represented species (*Leptostylis ampullacea*, 14.1%; *Leucon pallidus*, 12.5%) and the least represented species (*Bathycuma*

brevirostre, *Leucon spinulosus*, *Styloptocuma gracillimum*, *Campylaspis horrida*, *Campylaspis globosa*, and *Hemilamprops uniplicatus*; all 1.6%). suggesting particular ecological forces that favor or limit certain species, or that certain species have a restricted ecological niche. Unlike less represented species, dominant species may have particular adaptive traits that contribute to their exploitation of food, interspecific competition, or resistance to changing environmental conditions.

Comparing the frequency of each species with the mean and median (see upper right-hand corner of Figure 3) also helps to identify more or less frequent species. As mentioned above, *Leptostylis ampullacea* and *Leucon pallidus*, with above-average frequencies, are dominant species. Thus, the presence of species with frequencies above or below the mean (2.85 cumacean species) and median (2.00 cumacean species) support particular ecological forces that advantage or disadvantage certain species. Furthermore, a below-average median indicates that most species have relatively low frequencies, while a few species have higher frequencies, indicating an asymmetrical distribution.

The distribution of samples of the different cumacean families according to the variety of water masses in which they were collected is illustrated in Figure 4, allowing comparison of the diversity and potential preferences of the different families in each water mass.

Arctic Polar Water (APW) shows a high diversity of families, with a greater presence of the Leuconidae, Diastylidae, and Astacidae. Arctic Polar Water/North Sub-Arctic Intermediate Water (APW/NSAIW) shows a strong presence of Diastylidae and Leuconidae, with a low density of Lampropidae. Iceland Scotland Overflow Water (ISOW) has a high abundance of specimens, including Lampropidae and Diastylidae. Labrador Sea Water (LSW) presents a low diversity of families, with a preponderance of Astacidae. NAW (North Atlantic Water), like Arctic Polar Water (APW), present a high family diversity, with a predominance of Lampropidae, Ceratocumatidae, and Astacidae. North Sub-Arctic Intermediate Water (NSAIW) shows, like Labrador Sea Water (LSW), a low family diversity, represented by the Leuconidae and the Lampropidae. North Sub-Atlantic Deep Water, cold and warm (NSWc abd NSDWw), have high abundances of Diastylidae, with the presence of Leuconidae in warm North Sub-Atlantic Deep Water.

This testifies to the resistance and ecological adaptability of the Diastylidae family to a wide variety of environmental conditions, reminiscent of *Leptostylis ampullacea* in Figure 3, which belongs to the Diastylidae family. Two water masses contain the greatest diversity of cumacean families, with Arctic Polar Water (APW) and North Atlantic Water

(NAW) both having five families; APW: Astacidae, Diastylidae, Lampropidae, Leuconidae and Pseudocumatidae; NAW: Astacidae, Bodotriidae, Ceratocumatidae, Diastylidae, Lampropidae and Pseudocumatidae. This concomitance of different families could be explained by the diversified and resource-rich environments of these two water masses, favoring various complex ecological niches exploited by these families.

The distribution of samples of the different cumacean families according to the type of habitat in which they were collected during sampling is shown in Figure 5, enabling us to compare the diversity of the different families in each type of habitat. In the deep sea, we find a wide variety of families, dominated by Diastylidae and Lampropidae. The shelf presents a wide variety of families, but less so than the deep sea. It is dominated by Leuconidae. The slope has a low diversity of families, with a greater presence of Diastylidae.

The strong presence of families in particular habitats, such as Diastylidae in the deep sea and on the slope, and the Leuconidae on the shelf, suggests that these families have acquired adaptive characteristics (physiological, behavioral, or morphological), which could favor their survival in these specific environments. It also recommends that accessible resources (food and ecological niches) and environmental conditions, such as temperature, O_2 concentration, and sediment type, are key factors in the distribution of cumacean families.

The correlation between genetic sequences and two attributes, one climatic (wind speed (m/s) at the start of sampling) and the other environmental (O_2 concentration (mg/L)), is presented in Figures 6 and 7. This correlation is based on four distance metrics: Least-Squares (LS), Robinson-Foulds (RF), normalised Robinson-Foulds (nRF), and Euclidean. All the attributes presented in the first step of the *aPhyloGeo* software section (see *aPhyloGeo* software) were analyzed and are available on {https://github.com/tahiri-lab/Cumacea_aPhyloGeo} in the img and script python file. However, only these two parameters showed the most interesting mutation rate. Fluctuations in these parameters seem to reflect an adaptive response of these specific genetic sequences to environmental and climatic conditions.

Sequence correlation raises the question of how variations in sequences (i.e., windows) respond to, or vary with, climatic and environmental conditions. Conserved positions (low values) could potentially suggest functionally essential regions that do not readily vary with changing conditions. On the other hand, fluctuating positions (high values) could

present specific adaptations to conditions, in this case wind speed (m/s) and O_2 concentration (mg/L). Analysis of how sequences do or do not vary under these two different conditions can highlight regions in the sequences (i.e., windows) that are sensitive or resistant to fluctuations in these two attributes. In our results, we observe a similar fluctuation in correlation with these two parameters between Figures 6 and 7.

In Figure 6a, the peaks and troughs suggest that some positions across the sample sequences are more conserved and therefore similar (smaller distance), probably indicating potential functional or structural significance, while other positions show greater variability (larger distance). In contrast to Figure 6a, the values in Figure 6b are more concentrated on restricted values, indicating a more uniform fluctuation. In this context, lower variation may indicate that changes in sequences do not completely affect local phylogeny. The same applies to Figure 6c, where normalized distances are rather homogeneous, suggesting that variations in sequence positions have a relatively constant impact on the phylogenic structure of the trees. The Euclidean distance, shown in Figure 6d, appears to be the most sensitive and disparate distance for our data. A higher Euclidean distance means a greater difference between sequences located between positions 520 to 529 amino acids (aa), suggesting that this site is more variable due to numerous mutations or evolutionarily susceptible to change.

Figure 7a is similar to Figure 6a but shows greater variation between different sequences positions. Windows with smaller mean-squared distances are more likely to be conserved during evolution, while windows with larger distances show greater instability. The RF distances in Figure 7b vary with a restricted range of values (50 to 70). These small variations suggest that sequence fluctuations do not alter local phylogeny. Like the previous figure, Figure 7c shows a fairly homogeneous distribution. This means that variations in the positions of individual windows exert a fairly uniform influence on the phylogenetic arrangement of trees after normalization. Like Figure 6d, the Euclidean distance presented in Figure 7d shows the greatest sensitivity and heterogeneity in our data. The position with the highest Euclidean distance (between position 1190 to 1199 aa) shows higher dissimilarity between sequences at this position, which may signify a more fluctuating or evolutionarily unstable site.

All these results will need to be further studied and analyzed in greater depth to better understand them and draw sound conclusions.