

General assessment

The manuscript addresses a potentially interesting topic, namely the effects of ocean acidification on *Mytilus galloprovincialis* and *Phorcus* sp., using the carbon isotopic composition of shells combining two approaches: from natural CO₂ vent systems and from laboratory experiments. However the manuscript presents substantial methodological, experimental, and statistical limitations that compromise the robustness of the discussion and conclusion. In particular, the comparison between organisms belonging to inadequately characterized taxonomic groups, different exposure approaches, different size classes, non-comparable exposure times, and different environmental conditions introduces numerous sources of variability that are not adequately controlled. These issues are further compounded by important gaps in the description of the methods, potential pseudoreplication in the laboratory experiments, and an interpretation of the results that, in several instances, goes substantially beyond what the data can demonstrate. I also consider the lack of species-level identification of the *Phorcus* specimens and the use of individuals from natural populations without excluding differences in population composition among sites to be particularly problematic. Indeed, at least 3-4 species belonging to the *Phorcus* genus are present in the Mediterranean Sea. This issue is especially relevant considering that the authors themselves emphasize in the Introduction the importance of species-specific responses to acidification. Overall, the issues identified do not appear to be resolvable through a simple revision of the manuscript or the Discussion. Some of them would require new experimental information, improved characterization of the organisms and study sites, and a substantial revision of the statistical approach. For these reasons, I recommend rejection of the manuscript.

Main concerns

Materials and Methods – Natural acidified conditions

The comparison between *Mytilus galloprovincialis* and *Phorcus* sp. does not allow any robust biological conclusions. The authors compare a single species, *M. galloprovincialis*, with a group of organisms belonging to the genus *Phorcus*, without providing species-level taxonomic identification of the *Phorcus* specimens collected in the field. This approach is problematic because it prevents the assessment of species-specific responses to acidification. Considering that many *Phorcus* species occur in the Mediterranean, the presence of different species at different sites could introduce a substantial source of biological variability and represent an important bias in the comparison. It is possible that the difficulty of taxonomically identifying *Phorcus* specimens motivated this choice, whereas *M. galloprovincialis* is relatively easy to identify. However, this methodological difficulty cannot be considered negligible, as it fundamentally compromises the comparability of the organisms used in the study.

The Mussel Watch field experiment is also insufficiently described. The manuscript does not clearly state how many cages were used at each site, how many animals were present in each cage, or whether mortality was monitored throughout the experiment. Furthermore, considering that mussels are filter-feeding organisms, the characteristics of the cages and their degree

of permeability to water should be described, particularly to assess whether individuals located in the inner parts of the cages had comparable access to water flow. Moreover, the depth at which sampling was conducted at San Pietro is also not reported, and it is unclear whether it was comparable to the depths at Vullatura and Castello Aragonese. Although the authors state that the San Pietro site has biogeochemical parameters comparable to those of the ambient areas at Castello and is unaffected by volcanic inputs, this statement is not supported by any biogeochemical characterization presented in the manuscript. The authors describe pH, DIC, and TA gradients, but the actual values measured at different sites are not reported.

Another important gap concerns the origin of the *M. galloprovincialis* individuals used in the transplant experiment. The authors emphasize that the two species have different ecological histories; however, the origin of the mussels used in the field experiment is not adequately described. It is essential to specify which population the mussels originated from and whether they were collected from the wild, particularly if they were obtained from a site not shown on the map of the study area. This information is important because the previous environmental history of the transplanted individuals may influence their physiological and isotopic responses during the experiment. There is also insufficient information regarding the initial size of the animals and their growth during the 60-day exposure period. Mussel size is particularly important because it is directly related to growth and therefore to the amount of carbonate deposited during the experiment. It would therefore be important to know whether animal growth was measured and to what extent the environmental conditions allowed normal growth.

The statement that the *Phorcus* individuals represent populations chronically exposed to acidified conditions “from the larval stage through the juvenile, sub-adult, and adult phases” is particularly problematic. This statement cannot be demonstrated based on the data presented. Given the geographical proximity of the different sites, larval dispersal or recruitment from other areas cannot be excluded. No evidence is provided to demonstrate that the individuals were born and spent their entire life cycle at the site where they were collected. Furthermore, only three individuals were collected at some sites. Such a low sample size appears insufficient to represent the natural variability of a population, particularly when the authors collected up to 15 individuals at other sites. I understand the difficulty of obtaining large sample sizes from natural populations; however, with such a low “n”, it is impossible to exclude the possibility that the sampled individuals were transient rather than representative of a resident population. This issue is further aggravated by the absence of species-level taxonomic identification. Finally, the two species were not sampled at the same sites: *M. galloprovincialis* was used across the different sites of the gradient, whereas *Phorcus* was absent from Vullatura. This further complicates the overall comparison between the two organisms.

Materials and Methods – Laboratory experiments

The comparison between the natural and laboratory experiments also presents important limitations. The animals used in the two approaches differ in size, exposure duration, environmental conditions, and probably origin (lacking information). In the case of *M. galloprovincialis*, the laboratory animals were juvenile specimens (mean length of approximately

1.5 cm) collected from the northern Adriatic coast near Chioggia, whereas the organisms used in the field experiment were probably adults (again no info available). Size therefore represents another potentially important variable that could explain differences between the two experiments. Moreover, the two environmental seawaters used can be highly different, one from North Italy from a lagoon and the other from South Italy near an island, performed also in different years. The manuscript does not report that the water parameters used in the laboratory were comparable to those present in the natural environment at Ischia. Differences in water chemistry, salinity, temperature, nutrient availability, and other environmental parameters could contribute to the differences observed. The exposure durations also differ: approximately 60 days in the field experiment and 90 days in the laboratory experiment for *M. galloprovincialis*, whereas the laboratory experiment with *Phorcus* lasted eight weeks. These differences make impossible a rigorous comparison between the different experiments. The laboratory experiments also lack sufficient information on the actual amount of food available, and, in the case of mussels, the concentration of microalgae supplied (cells/mL). Similarly, the actual pH values measured in the different treatments are not adequately reported. It is therefore unclear whether the nominal values set by the system actually corresponded to the conditions to which the animals were exposed throughout the experiment. For *Phorcus*, the presence of several species within the genus makes taxonomic identification of individuals from the different populations particularly important. Comparing populations from southern and northern Italy without verifying their species identity may introduce additional variability into the dataset and makes it difficult to attribute the observed differences specifically to acidification.

Isotopic analyses

The description of the shell sampling procedures is insufficient to ensure study reproducibility. It is unclear how many individuals were selected from each cage, how they were selected, and which specific portion of the shell was analyzed. Although the authors generally identify the growing shell margin for *M. galloprovincialis* and the aperture for *Phorcus*, they do not provide sufficiently detailed information on the size and precise location of the analyzed shell portion. For example, it is unclear how the newly deposited shell portion was identified, whether it represented a defined percentage of the total shell length, and whether an initial shell size measurement was available before the experiment. Did the authors use image analysis to quantify the newly deposited shell portion? Similar information is missing for *Phorcus*. These details are particularly important for the laboratory experiments, where the objective is to interpret the isotopic signal as a consequence of the experimental exposure. For the field experiment, it is also difficult to understand how the shell portion corresponding to growth during the 60-day exposure period was identified in a standardized manner. The statement "All samples consisted of multiple shell specimens" is particularly problematic. It is unclear whether "sample" refers to a single isotopic sample obtained by pooling several shells or to multiple shells analyzed separately. This distinction is fundamental because the two approaches result in completely different experimental units and sample sizes. If several shells were pooled into a single sample, inter-individual variability cannot be assessed. If the shells were analyzed individually, the number of individuals analyzed and how these measurements were subsequently used in the statistical analysis must be clearly reported. As currently written, this

statement makes it impossible to determine the actual n used in the analyses and the true experimental units. This is an important issue both for study reproducibility and for the correct interpretation of the statistical results.

Statistical analysis

The main statistical concern is pseudoreplication. In the laboratory experiments, multiple individuals were maintained within the same tank, yet the statistical analyses appear to treat individual animals as independent observations. However, the experimental unit is the tank rather than the individual animal. Individuals maintained within the same tank share the same experimental conditions and therefore cannot be considered fully independent replicates of the treatment. In the mussel experiment, 20 individuals were maintained in each tank, with three replicate tanks per treatment. Therefore, the effective experimental sample size is $n = 3$ tanks per treatment, rather than the total number of individuals analyzed. The reported ANOVA degrees of freedom ($df = 1,73$), however, suggest that individual mussels were treated as independent observations. This may substantially inflate the degrees of freedom and result in artificially high statistical significance. The same concern applies to the *Phorcus* laboratory experiment. The analysis should therefore account for the hierarchical structure of the experimental design, either by considering the tank as the experimental unit or by using an appropriate mixed-effects model with tank included as a random effect. Treating individuals within the same tank as independent replicates is not appropriate for this experimental design.

Discussion – Natural acidified conditions

Section 4.1 presents a potentially interesting result, namely the enrichment in $\delta^{13}\text{C}$ observed in *Mytilus galloprovincialis* shells at low-pH sites. However, I believe that the interpretation proposed by the authors goes substantially beyond what the data can demonstrate. In particular, the conclusion that biological modulation is capable of “overriding” the ambient geochemical signal is not adequately supported by the dataset. The main issue is the absence of $\delta^{13}\text{C}$ -DIC measurements along the environmental gradient and of measurements of the main organic carbon sources. Without these data, it is not possible to distinguish the contribution of ambient DIC from that of metabolic carbon or to quantify the relative importance of the different carbon sources. Consequently, the statement that respiratory CO_2 represents the dominant carbon source for calcification remains purely speculative. Similarly, the hypothesis that acidification induces prolonged valve closure, isolation of the extrapallial fluid, and consequently an increased metabolic contribution has not been experimentally tested. No behavioral, physiological, respiratory, or calcification measurements are available to verify this mechanism. The section concerning primary productivity is also particularly problematic. The statement that photosynthesis produces biomass enriched in ^{13}C that is subsequently assimilated by the organisms is not adequately supported. Again, measurements of the $\delta^{13}\text{C}$ of organic matter and its sources are lacking, making it impossible to substantiate the proposed mechanism. Similar caution is required for the ontogenetic interpretation of *Phorcus* sp., where the $\delta^{13}\text{C}$ pattern is attributed to a progressive reduction in vital effects without independent measurements of growth or calcification rates. Several alternative explanations remain possible, including changes in diet,

metabolism, microhabitat, and seasonality. Finally, I do not consider the statement that seasonal variability in $\delta^{13}\text{C}$ -DIC cannot explain the observed pattern to be sufficiently justified, since this variable was not directly measured at the study sites. Comparison with values reported in the literature is not sufficient to exclude this possibility.

Overall, an isotopic observation is used to construct a highly specific physiological and biogeochemical explanation without measuring the variables required to test the proposed mechanisms.

Discussion – Laboratory experiments

The interpretation of the laboratory experiment is also more speculative than the data allow. The authors observe a decrease in $\delta^{13}\text{C}$ in *Mytilus galloprovincialis* under reduced pH and interpret this result as a direct geochemical response to the reduction in CO_3^{2-} availability. However, the experiment does not allow the authors to identify CO_3^{2-} reduction as the specific cause of the observed isotopic change. Lowering pH by adding CO_2 affects several components of the carbonate system simultaneously, and the study does not provide evidence that the observed $\delta^{13}\text{C}$ shift can be specifically attributed to changes in CO_3^{2-} . The authors further conclude that, under controlled laboratory conditions, the shell acts as a relatively passive recorder of environmental chemistry. However, controlled conditions do not eliminate physiological effects. The mussels were still actively feeding, growing, respiring, and calcifying, and no measurements of growth, calcification, respiration, or metabolic activity were provided. Therefore, the study cannot determine whether the observed isotopic change was driven primarily by environmental carbonate chemistry, physiological responses, or a combination of both. The data support an association between reduced pH and changes in shell $\delta^{13}\text{C}$, but they do not demonstrate the specific mechanism proposed by the authors. The interpretation of *Phorcus* sp. is even weaker. A non-significant trend is used to support the existence of high physiological plasticity, different calcification strategies, changes in energy allocation, and stochastic calcification events. None of these hypotheses is directly supported by measurements of growth, metabolism, respiration, or calcification. Inter-individual variability alone does not demonstrate physiological plasticity or different calcification strategies. The laboratory result for *Phorcus* is in fact not statistically significant ($p = 0.144$), and therefore cannot be used to support an isotopic response to acidification. Likewise, the significant treatment effect observed in natural *Phorcus* populations cannot simply be interpreted as an effect of acidification because the individuals originate from different natural populations, the species were not taxonomically resolved, sample sizes can be variable, and several environmental factors may be confounded with pH. The strong ontogenetic effect observed in the *Phorcus* data ($F_{2,39} = 29.67$; $p < 0.001$) further demonstrates that $\delta^{13}\text{C}$ is strongly influenced by growth stage. However, the criteria used to assign individuals to the different ontogenetic stages (juvenile, sub-adult, and adult) are not described. It is unclear whether these categories were based on shell size, morphological characteristics, or another criterion. This information is particularly important because shell size and growth rate may directly influence the isotopic signal. Moreover, without independent measurements of growth or calcification rates, the subsequent interpretation in terms of changes in vital effects remains speculative.

Overall, also in this section the authors transform descriptive isotopic patterns into untested mechanistic explanations. The experiments demonstrate, at most, an association between acidification and changes in shell carbonate $\delta^{13}\text{C}$; they do not demonstrate either predominantly geochemical control of the signal in *Mytilus* or metabolic control in *Phorcus*.

Conclusions

The methodological and statistical issues described above, considered together, substantially compromise the robustness of the conclusions of the manuscript. The problems are not limited to the overinterpretation of some results but also concern the difficulty of reliably identifying the experimental units, assessing their independence, determining the taxonomic composition of the *Phorcus* populations, verifying the actual exposure conditions, and identifying the sources of carbon incorporated into the shells. These issues cannot be resolved through a simple revision of the text or a reformulation of the Discussion. Additional experimental information and, for some of the main issues, new data would be required to adequately address these concerns. Furthermore, although the authors compare natural and laboratory approaches to study ocean acidification, the two experiments differ in several additional factors, including species, animal size, origin, and exposure duration. Therefore, differences in isotopic responses cannot be attributed specifically to the experimental approach (natural versus laboratory), as they may instead result from these uncontrolled differences. Overall, I consider these limitations to be fundamental to the experimental design and interpretation of the results. Therefore, I recommend rejection of the manuscript in its present form.