We thank the reviewer for considering our manuscript important and valid for publication. We value the insight and criticism provided in the review letter as constructing and changed the manuscript substantially. We believe that the result is a much-improved manuscript. Below is a point-by-point reply (in blue).

General comments:

The study presented by Pinko et al. represents a comparative study of two different common LBF species, with different shell and symbiont types, exposed to elevated temperature over few weeks. The two main proxies assessed give insights into photosymbiont performance and holobiont health. Due to subtle differences, the authors conclude that *Sorites orbiculus* will be less affected by climate change than *Amphistegina lobifera*. They also claim insights into distinct effects of preexposure to moderate temperatures regarding the LBFs thermal tolerance. Along the lines of former studies, the experiment shows that there are species-specific thresholds regarding temperature and duration of exposure, and that LBF from the Eastern Mediterranean, which are most likely Lessepsian invaders from the Red Sea, have a relatively high thermal tolerance. The further confirm that the photosymbionts seem to be the 'weaker' member of this symbiotic association, showing the earlier stress response. Hence, the study give further important proof of prior hypotheses on LBF thermal stress responses, and adds to the knowledge of species-dependent thresholds. Hence, I consider it important and valid to publish this data. However, the novel insights are limited, as I do not think that calcification can be considered as a host-specific response (as they suggest), and therefore this study does not assess the relative contribution of host and symbionts (see specific comments).

Specific comments:

1) Calcification cannot be considered as a host proxy, as it is largely influenced by photosynthesis. It is hence, as in many other studies, a holobiont proxy.

It is true that many key elements are involved in the biological machinery of the calcification process. Among these is the mutualistic partnership between the algal symbionts and the foraminiferal host.

Nevertheless, calcification is a physiological trait preformed only by the foraminifera and thus present a direct proxy of its wellbeing (as in other calcifying organisms). This is based on a common observation that stress lowers the physiological activities that involves high consumption of energy. The same is true for photosynthesis, which is a physiological trait of the algal symbionts. Because of the exclusiveness of each parameter we have selected them to disentangle the complex relationship between the two components of the holobiont.

Since this issue was raised by both reviewers, we recognise the need for clarification in the paper and added an explanation of the rational to the text at the end of the introduction.

Although prior studies mostly used possibly less precise methods to assess growth (e.g. increase in surface area or buoyant weight in studies by Schmidt, Prazeres, Stuhr, Hallock and others), it basically gives similar information.

Alkalinity anomaly method provides similar information to that produced by the measure of increase in surface area or buoyant weight. However, its main advantage is that it is much more accurate and thus presents an ideal experimental approach for detecting even subtle differences in performance under different treatments. This technique requires high level of analytical expertise and meticulous work measuring large number of samples (in order to replicate properly) but it can detect differences of even few single micromolar in carbonate production. Thus, this method is highly beneficial over the other common method despite the amount of time and expertise required.

Nearly all studies on LBF stress response assessed at least one holobiont parameter such as growth (often also others to get a better picture, as calcification / growth can be limited due to other factors that are independent of stress, hence, it is not a very good parameter anyways), and one or more photosymbiont parameters. The only study to my knowledge that actually managed to gain host-specific insights was Stuhr et al. 2018 (Scientific Reports) by differentiating between host and symbionts on the protein level. But even here, the influence of photosymbionts stress on host stress cannot be fully excluded.

It is true that stress of the symbiont will affect the host and vice versa. However, it is important to try and find indicators related to each of the component. Stuhr et al. 2018 identified differential expression in protein related specifically to host or symbiont. Under the same logic in this study we examine physiological activity only related to one of the components (calcification of foraminifera and photosynthesis of symbiont algae). The reviewer indicates that even in the data from Stuhr et al. 2018 the host is affected by symbionts stress but still doesn't exclude the fact that the stress is experienced by the foraminifera, we believe that the case is similar with our approach.

Due to the lack of novelty described above, I would suggest to the authors to focus more on the comparison between *A. lobifera* and *S. orbiculus*, and the detected differences in time-related responses (seasons and experimental duration), and emphasize these in more detail.

We changed the discussion to focus more on the comparison of the differences between the holobionts.

Furthermore, the methods section is very short and lacking a lot of details, descriptions etc., and many crucial information has unfortunately been moved to the supplementary materials. The same applies to some of the results, e.g. the statistics, which should be at least indicated in the text or the figures where significant. Even with the results provided in the supplementary materials, it is not possible to fully judge where statistically significant variation were detected due to the poor representation and lacking explanations.

We have added the information to the methods and results sections and the statistical indication of significance (p value) throughout the manuscript.

I am also wondering why no further parameters were tested, and calcification and photosynthesis normalized by individual, which is very unusual and prohibits comparison with other studies.

Calcification rate and net oxygen production are quantitative very accurate parameters. As such, they were chosen for this study that aimed to recognize

even subtle differences between treatments and between species. Normalising these parameters by individual (as done by Evans et al., 2015, Titelboim et al., 2019) is meant to be more informative and more comparable than other ways of presenting this data that in many cases doesn't include the amount of foraminifera in the experiment or when normalized to mg is less specific because it can represent either more small specimens or less adult specimens.

It is also less precise if not all individual were of the same size (which they most likely were not).

Specimens were confined in size as mentioned in the manuscript (line 87-88). We have now added the specific size fraction used for the experiments.

And I wonder, if you measured photosynthetic activity via oxygen production, why didn't you simply also measure respiration via oxygen consumption in darkness? This would have provided another valuable indicator for holobiont condition, and would have allowed to calculate gross oxygen production. So far, you only provide net values for photosynthesis, not considering that respiration is likely to be higher under high temperatures, which results in lower net photosynthesis, even though the actual production of oxygen may be constant.

Indeed. We have measured net photosynthesis. The use of this parameter clearly takes into account the fact that respiration lowers net photosynthesis. We understand from this comment that this was not clear and we adjusted the text to always say "net oxygen production/photosynthesis).

Lastly, I think this study could further be improved by discussing some interesting observational details such as bleaching (Did you observe it? In which species? Mottled or more gradual? Or mortality? Reproduction?) as well as by including time into the statistical evaluation of the results (e.g. two-factorial analysis of variance with time and temperature as factors).

There were no observations of bleaching, mortality, reproduction etc (otherwise they would have been mentioned). This is in fact another indication that the parameters examined in this study are the most appropriate ones as the accuracy provided by them can identify stress before a fatal response was reached.

We will include in the revised version additional statistical analyses, including two-factorial analysis of variance with time and temperature as factors.

Technical comments:

L2: "Roommate problems or successful collaboration?" this title sounds catchy, but the question is still as valid as before. . . I don't think the paper is resolving this question,

Following our response to comment 1 we believe that our title does describe the content of the paper and is not only valid but the formulation of it as a question actually emphasis the complexity of the relationship between the holobiont components.

and also the first part of the title is very broad (e.g. climate change includes more than just temperature stress) and should be a bit more specific

We except this part of the comment and changed the first part of the title to: "Foraminiferal holobiont thermal tolerance under future warming…."

L11: ". . . hyaline diatom-bearing *Amphistegina lob*ifera and the proceallaneaous dinoflagellate-bearing *Sorites* . . ."

Changed

L12-13: see discussion above

Changed according to the comments of reviewer 1.

L16-L17: "future warming will significantly shift the relative contribution. . ." this is taking the implications way too far. You only see small differences in their response to 35◦C in respect of timing. Please be more specific (in general) and stick to what you actually show.

This was changed to "may" according to the comment by reviewer 1. Further, the word "significantly" was deleted to make the sentence less affirmative.

L18-21: You mention pre-exposure for the first time here, and it is rarely described in the manuscript in general. What do you mean by this? The season? Or different acclimation temperatures? I also don't understand why you suggest that it reduces thermal tolerance. Please reconsider these statements and adjust to your results and the discussion.

Following the review process, we have reconsidered this part of the manuscript and recognized that there are not enough evidences to make this case. Thus, we have decided to eliminate the results of the February experiment and the discussion regarding pre-exposure.

L27-28: ". . . one of the regions most affected. . ."

Changed

L30: "Symbiont-bearing large benthic foraminifera . . ."

Changed

L34-35: To my knowledge, none of these studies really provided evidence for temperature control on symbionts composition. Some suggested that there may be a connection, but statements saying that they are "strongly controlled" would definitely require further proof, especially since other studies did show extremely flexible relationships (e.g. several Lee et al. studies, Pochon et al. 2007, Schmidt et al. 2016)

We have changed it to: "The symbiont composition of LBF was suggested to be controlled by temperatures"

L45: delete "calcifiers"

Deleted

L55-56: there are much earlier studies that describe these species in much more detail (such as Hansen & Burchard 1977 and Hottinger 1977) that deserve to be cited here.

Added

L59: ". . . Israel, during. . ."

Changed

L60: the picked size fraction is crucial when it comes to assessing growth/calcification as it is strongly linked to ontogenetic phase. Hence, please provide this information in the manuscript and not in the supplemental materials.

Added

L62: the same accounts for the sample sizes. It is important to know in order to judge the power of the study.

Added

And “. . . 60-ml airtight. . .”

Changed

L 68: there is still a lot of detail missing regarding the culture conditions: was there water flow? What kind of water were they in? Were they fed? pH? . . .

There was no water flow, this is implicit from the description of airtight Erlenmeyer that are important for measurement of oxygen and alkalinity. We use natural sea water as indicated in line 88. We do not feed them because we use natural seawater filtered above 45 µm, that contain the foraminifera and symbiont algae nutrition. pH of water through the experiment was added to the results and supplement.

L65: Which temperatures? In the baths or the flasks? What means regularly? Once a week or once per hour would both be regularly but are very different.

The flasks are incubated in the water baths and thus the temperature of the water within the flasks is controlled by the water in the bath. These were monitored using HOBO data loggers that recorded temperature every 1 hour.

L71: Why did you use calcification rate as the only parameter? Please explain.

February experiment is now eliminated from this manuscript because of the differences in methodologies.

L74: Same for: why did you only include *A. lobifera* for the spring experiment?

February experiment is now eliminated from this manuscript

L77: Which “constant conditions”?

These details were at the supplementary file but are now written also in the manuscript itself

L79-80: How many samples didn’t show oxygen production? And any suggestion why?

A total of three samples didn't show normal oxygen production (this is detailed in Supplementary Table S1) and in any case the number of replicates per treatment didn’t decrease below three. The word " apparent " was replaced with " similar values of net oxygen production as other samples" which describe the scenario better, we cannot speculate why this happened.

L81: How was temperature adjusted? What instruments did you use to control this?

The temperature is controlled by heating circulators with a thermostat and were adjusted manually.

L82: where are these temperatures expected? In the Med Sea? The Red Sea?

Added "in the Eastern Mediterranean"

L85: Was the water filtered? Pre-conditioned temperature-wise?

Water were filtered through 45µm to ensure nutrition for the holobionts but also reduce noise in the oxygen measurement. Other than that, the water were not treated.

L85-90: Please give more details on the method. Did you do this in the 60-ml flasks? What was the time frame? Were foraminifera pooled?

Water from all the Erlenmeyer flasks were replaced at the same time, immediately transferred to air-tight syringes, and then measured for their alkalinity and dissolved oxygen. Measurements of all samples were conducted within the next two days. To ensure no changes in water properties accurse in this time frame standard material is measured before and after the first and last sample of the set, respectively). Foraminifera were kept inside the Erlenmeyer flasks throughout the experiment.

L91-93: Same details are missing here, as well as references. What instruments did you use? What light? Which temperatures? ". . . µg L-1 ..." and please define RDO.

This follows the same protocol as described in the last comment. We added the technical details of the sensor, fixed "µg L-1" and defined RDO in the text.

What was the accuracy value?

The accuracy of the optical dissolved oxygen sensor was better than ± 0.01 mg/L.

And in general, you normalize both parameters by specimen. For better comparability, they are usually normalized by species size (given by surface area or weight etc.). Please consider doing so.

See response to Specific comment 4

L95-97: Please give n for each parameter, treatment, time point, described which data got transformed, and which tests exactly got chosen "accordingly".

We added all of this to the manuscript in 2.3 Statistical analysis

L99: " . . .cases where normality . . . non-parametric Kruskal-Wallis test . . ."

Changed

L100: Please name the "proper" post-hoc tests.

This is case specific and the post-hoc test used in each case is detailed in the supplementary tables. As mentioned in the previous comment, this is now also added to 2.3 statistical analysis in the manuscript.

L106: you cannot say whether there would be differences in *A. lobifera* between winter and spring, but sounds as if you do. Please rephrase.

February (winter) experiment is now eliminated from this manuscript

L108: What is the "x" in your box plots? Please indicate significant differences by a letter report. And give n in captions. Also provide full species names and specify what the whiskers represent (SE?).

Figure changed as suggested

L112&L130: keep descriptions consistent

Noted and changed

L113f: Please avoid expressions like "much higher", "a decrease" or "substantially different". What does that mean? Please provide statistical test results and/or how big is the difference (twice as high, ~20 %...)

We deleted "much" and "substantially" however the word "decrease" is not interpretive and is very clear especially at week 3 and 4

L120: “The symbionts’ photosynthetic . . .”

Changed

and how is the “sensitivity pattern” different? Apart from one week in the 35◦C treatment, they look very similar to me.

This is what we mean. We rephrased it to: "indicates faster response than that presented by calcification rates"

L122: As mentioned before, please provide overview of statistical results here, ideally in figure.

Added

L123: If you mention this “abnormal value” please state in which way it was abnormal and why you suspect this to be related to handling. You say it’s not used for average and SE calculations. Does that also mean the further statistical analyses?

We added "abnormally high value" to indicate the way it is abnormal and considering the suggestion of reviewer 1 removed the statement about sample handling

Fig. 3: please jitter weeks more, so they are easier to recognize. In which week did you have the extreme value?

The graph is modified and also includes now the week of the extreme value (week 2)

L128: “significant negative response” I am not sure the word significant is used correctly here.

Removed the word "significant"

L134: “week” and what means “massive bleaching”? Please describe, give proportions etc.

This will be added to the manuscript

L135-136: “... between 25◦C and 32◦C, and was thus clearly ...” As described in the general comments, please describe it more. What may be the reason for this bleaching? Were both species affected? *A. lobifera* could have a lower light tolerance than *Sorites* and could therefore bleach. Or since it only affected the lower temperature range, it could actually be related to reproduction. Why didn’t you exclude the bleached specimens from the analyses? I suggest to include the data anyways, e.g. only for comparison in the supplementary, or to conduct analyses on those samples that had no bleaching. Moreover, as the 35◦C had no bleaching, please provide at least these values to compare with former weeks.

This will be added to the manuscript

L142-143: Well, I think “clear differences in thermal tolerance” is a bit exaggerated. I would call them rather subtle.

We deleted this as part of the first reviewer comments

L144: This is not true, the calcification seems lower at 35◦C (significances missing), at least in the winter populations. Plus, the experimental exposure in spring was shorter, so the response may have just been delayed, as suggested by the reduced oxygen production.

As mentioned before, we excluded the winter experiment since it was done in slightly different conditions than the spring experiment. Furthermore, the difference between the species is not just in the magnitude but also in the timing. *Sorites orbuculus* reduced calcification only after the second week, which indicate it is more resilient than *A. lobifera*. It is true that this is a short-term experiment and therefore (and following the comment of reviewer 1) had changed the sentence to be less affirmative.

L147: please rephrase, something here doesn't make sense. And I again don't agree that you can state that there is a "strong dependence", as many other factors have been shown to be at least as important. Please also include some newer references here.

Rephrased and references added

L149&L161: "dinoflagellates" and change to "Symbiodinium" to Symbiodiniaceae (as you seem to be aware this taxonomic system has been revised)

Changed

L153: ". . . control a holobionts ..."

Changed to "… control the holobionts…"

L154-155: What "mechanism to cope with thermal stress was observed"? Please

describe, I do not think this paper actually showed 'shuffling'.

Explained

L156: ". . . explain the observation . . ."

Changed

L159: ". . . describes ..."

Changed

L164: "... 35◦C, whereas in ..."

Changed

L165: ". . . inhibited, indicating that it is. . ."

The sentence was completely changed following comment by reviewer 1.

L166: That means you refer to *Sorites* only? When is "earlier"?

Does this refer to old line 162? If so, yes this refer to the symbionts of *Sorites*, this is why we write " Symbiodinium". Earlier, is from the first measurement after one week contrary to calcification that only decrease after the second week. This is now clarified in the text.

L168: An apparent higher sensitivity (earlier/stronger response of the symbionts than the holobiont) was also observed by other studies such as Prazeres et al. 2017, Stuhr et al. 2017, Schmidt et al. 2016. . .

References added

L170: Here you mention bleaching again: so were these specimens that you measured calcification on already bleached? If there was, this may also indicate that there was another stress factor such as too high light intensity or the wrong light spectrum (e.g. Hallock once showed that blue light facilitated high growth rates but at the same time led to bleaching). Please discuss your observations.

No, these were not bleached. However, the decrease in photosynthesis is considered as a negative response of the symbiont algae and thus compared to the observation made by Hallock regarding bleaching. We added clarification of this in the text.

Light intensity in the experiments is not a factor causing bleaching since for *Amphistegina* this light was used before in several published (Titelboim et al., 2019) and unpublished experiments. For *S. orbiculus*, this is not the case since through the 3 and 4 week experiments no bleaching was observed.

L172: The resolution is a very important point! One week is a long time for a foraminifera! Plus, so far I don't even know how long your calcification measurements or photosynthesis measurements took. They don't calcify continuously all day long, so the time frame may strongly influence the results. The same accounts for the photosynthesis, which varies over the time of the day.

Indeed, the weekly resolution limits our ability to identify changes in a shorter-term time interval. However, the aim of this study is not to quantify changes in the daily cycle of calcification and photosynthesis but to examine the response over longer time intervals (measurement per week).

The description of handling time and measurement is not relevant to this point and hopefully this was clarified in the response to comment on lines 85-90, and further in the text in the method section.

L174: Please specify the time of the onset.

The word "onset" was replaced with "magnitude" due to comment of reviewer 1

L175: What do you mean by "very cold"? I think that is very relative... give a tem- perature range of what is usually encountered in the Med Sea in winter and spring, and ideally state what were the temperature measured during sampling in the methods section.

"Very" was deleted, and we also specified what are the low temperatures

L178: ". . . symbionts, or both. However, while the *A. lobifera* spring . . . "

Changed

L180: In which way did they respond "negatively"? Please be a bit more specific so the reader does not have to go back to each of the studies you cite to find you what you mean, and "... that, while ..."

Added: "…negatively responded by lowering calcification rates already to 30ºC…"

L181: ". . . temperatures benefits the . . ."

Changed

L186-187: There have been studies modeling the future changes in distribution and hence contribution, which should be cited here (e.g. Weinmann et al. 2013, 2017)

References added

L187: If you mention pre-exposure here (as in the abstract) please elaborate a bit more in which way they had different pre-exposures in the methods as well as the discussion parts.

Pre exposure is no longer discussed in this manuscript

L188-189: I don't understand this statement at all. Why? Where is you evidence for that?

This part was deleted

L189: Again, I find "clearly shows" a bit exaggerated.

"Clearly" is deleted

Supplement 1: Is the number of replicates given the value before or after exclusion of same samples? Why are the numbers different for the different time points? How did you deal with this unbalanced design in your statistical analyses?

The number of replicates is the number at the beginning of the experiment before the exclusion of samples after the acclamation part. This is now clarified in the Supplementary Table S1.

You used once filtered and once unfiltered water. Why? And why did you pre-condition the spring *Sorites* to another temperature than the rest? This is very crucial information and must not be excluded from the actual manuscript! I am not sure if you can compare your data the way you do with all these differences.

Considering these differences, February experiment and the comparison between spring and winter populations was excluded from the manuscript. Each species was pre-conditioned to its specific optimal temperature.

Sometimes you give two numbers after the comma, sometimes six or other. . . please be consistent (and usually its three).

The supplement with statistical analysis will be replaced and this comment will be noted for the new file

Why do you give four stars (they are actually called asterisk)? Usually, these are used to indicate the level of significance, from one (lower end of significance) to three (highly significant).

Since this presentation of results is not clear we will replace the statistical analysis to include actual p-values between each treatment

Why is some text red? Please explain in captions.

Significant results are marked in red. This will be added to the captions of the new supplementary

Table S2.2 and others: What are "1, 2 and 3" in your column headers? I cannot under- stand your statistical results if I don't know what is which group.

Since this presentation of results is not clear we will replace the statistical analysis to include actual p-values between each treatment

Put spacing equally before and after "=", but not after "(".

Noted and will be changed in the new file

Table S5.2: "Stars indicate homogenous groups and thus significant differences be- tween them"? This makes no sense to me, because if they are homogenous, they are similar, so no difference. . .

As mentioned in previous comments we will replace the statistical analysis to include actual p-values between each treatment