

As you will see, the reviewers raised a number of substantial criticisms that require me to reject your paper. While the reviewers (and I) were very impressed with the scope of the dataset, there were a considerable number of issues with the manuscript that make it unclear as to whether it could be revised sufficiently to be publishable. These issues range from statistics to calculations to the framing of the study and its claims. Perhaps a substantially rewritten manuscript with new analyses could be acceptable, but it is not clear to me whether or not that is feasible. My one broad suggestion is to align the goals of the study with a feasible set of questions that can be robustly addressed with the dataset in hand and appropriate statistics. This may narrow the scope, but it may also allow for a more rigorous study in the end.

A thoroughly revised manuscript might warrant further consideration, but it will require that you make extensive amendments to your paper, including the re-analysis or collection of new data. Even in this case, however, it may not be possible to address the most substantial concerns raised by the reviewers. We are currently under great pressure for space and it takes two very enthusiastic recommendations by the reviewers for a manuscript to be accepted.

>We thank the editor for the helpful feedback.

Reviewer 1 Comments for the Author:

Summary:

This paper tests whether linearly configured (streamlined) colonial invertebrates are faster and more efficient than other architectures. The motivation for the work stated in the paper is bio-inspired design of vehicles. The idea itself is an interesting one, and there are some interesting results in the paper itself.

>We thank Reviewer 1 for their kind feedback.

Overview:

In general, I found two main issues with the paper that need to be addressed in order to strengthen the results and the stated conclusions.

The first issue is that the conclusions in the abstract and the discussion are inconsistent with one another.

>We reworked the abstract to match the discussion (see specifics in replies to other comments below)

Part of the issue is that the hypotheses seem interdependent in some places, so it isn't easy to test them independently.

>We removed the interdependence structure of the primary hypotheses by switching the angle of inquiry on the second hypothesis (whether COT should decrease with speed) away from the

results from the first hypothesis test, and toward previous research (such as Bi & Zhang 2019) predicting that COT should increase with speed in jet-propelled systems.

There is also never a quantitative argument relating either of them explicitly to frontal drag, which is the crux of the paper.

>We removed our claims on drag from the introduction and added some discussion of the relevant forces in the discussion section.

The second is the need for post-hoc analysis to explore the driving source of significance in the results where it was found. Addressing the second point could help with the first one, as the driving sources might explain some of the unexpected results.

>We omitted the t-tests and replaced them with ANOVAs with Tukey post hoc contrasts (to explore the source of significance) with adjusted p-values. We added two supplemental tables S4 and S5 for architecture ANOVAs on the swimming speed and COT data respectively.

Below are some of the main questions and comments that I had while reading the paper.

Questions/Comments:

Lines 23-26 (Summary Statement): Reading the rest of the paper, the results were mixed in some cases (see below). This reads more like a hypothesis than a summary statement of what was found, and the results were more complex than expected. Is this what is intended?

>We edited the Significance Statement to: "Linear arrangements in multi-jet propelled marine colonial invertebrates are faster than less streamlined architectures without incurring higher costs of transport, offering insights for bioinspired underwater vehicle design."

Lines 32-36 (Abstract): What is meant by swimming performance? Performance can mean many things; since this is part of the hypothesis to be tested, the specific metric should be clearly stated.

>We rephrased swimming performance to swimming speed.

As written, the second hypothesis reads that faster-swimming taxa are more energetically efficient because the drag is lower. What is being compared here? Similarly streamlined taxa? Similarly swimming taxa? Again, this may be clarified in the main paper. Still, the ambiguity in the abstract generates immediate skepticism, as there are many reasons why something might swim faster but be less efficient. In both of these cases, I suggest being more precise in stating the hypotheses so that the reader

has a clear sense of the paper's aims and what is specifically being tested in each case.

>We rewrote the hypothesis structure in the Abstract: “We hypothesize that colonial architecture drives differences in swimming speed between salps due to differences in how frontal area scales with the number of propeller zooids in the colony. Based on findings from other jet-propelled systems, we hypothesize that swimming speed leads to higher cost of transport in salps.”

Lines 43-47: The abstract reads with much stronger statements than the results and sometimes omits that the authors' hypotheses weren't supported. In multiple places, for example: lines 357-358, 409-410, 421-423, 559-561, they explicitly state that the results are incongruent with one or more hypotheses. None of this is reflected in the abstract. The abstract should accurately reflect both the congruent and incongruent results, especially when the incongruences are in multiple places and in some cases completely contrary to the hypothesis.

>We edited the abstract results section to: “We found that linear colonies generally swim faster due to their differential advantage in frontal area scaling with an increasing number of zooids. While we did not find any significant differences in cost of transport between architectures, we found that higher swimming speeds predict lower costs of transport in salps”. Some of the line examples provided refer to results addressing secondary hypotheses (such as the relationships with pulsation rate) which are not the primary goal of the manuscript and thus don't belong in the abstract.

Lines 43-46 and 80-113: The authors claim that frontal drag accounts for the differences in swimming speeds, but it doesn't appear that the drag was measured/calculated.

>Measuring drag falls outside the scope of the study. We narrowed down our claims to the effects of relative frontal area (streamlinedness) on swimming speeds, which incorporates any underlying forces involved.

In addition, the hypothesis relies entirely on the assumption that frontal drag is the dominant energetic contributor to the system.

>The independent effect of varying relative frontal area can be accounted for even in the presence of additional sources of variation. We only rely on the assumption that frontal-area-driven resistance forces are a non-zero contributor with direct predictable relationships with colonial architecture. Other factors are likely to explain the residual unexplained variation. We added lines to the Discussion delineating alternative explanatory factors, inviting future studies to unravel their influence on salp swimming.

When considering length of chains, for example, other considerations, such as vortex shedding and added mass along the lateral sides, can contribute to energetic cost and can also scale with length. There is also an assumption that the system's stability and efficiency are independent of the chain's length, which is not necessarily true and impacts efficiency. Some of these effects could account for some of the unexpected results mentioned in the paper.

>We added further discussion on factors that may impair performance/efficiency due to an increased number of linear propellers (skin/friction drag, vortex drag).

Even if linear architectures are shown to be faster and more efficient, I'm not sure you can claim it is due to frontal drag unless you somehow quantify and compare the drag in the analysis. It would be a more substantial claim if the authors addressed why factors other than frontal drag can be ignored or controlled for and tie frontal drag to the quantitative results.

>We added further discussion/intro of the unaccounted mechanical factors that may confound the results. If friction drag, chain stability, and vortex shredding were more important than frontal form drag, we predict that linear chains would appear slower than other more stable and compact architectures.

For the statistical analysis throughout: the  $R^2$  values and the linear equations should be given, I didn't see that they were anywhere?

>We added adjusted  $R^2$  values for each regression. We also made the equations used for each regression more explicit.

In the cases where significance was found, post hoc analysis should be done to determine what drives the significance. In particular, there is quite a bit of variation in the linear species, and I wonder if one of the taxa is driving a good deal of the significance.

>We changed the t-tests to ANOVAs with post-hoc pairwise analyses to identify the drivers of significance.

Also, a table of the number of individuals collected for each species for each type of measurement for clarity. If I read the SI correctly, the samples for non-linear architectures were, in many cases, much lower than linear architectures, some only having 1-2 specimens. Even when there were several specimens available, it would be helpful to know the relative sample sizes and have it summarized in a table.

>We added Supplementary Table S3 summarizing  $N$ s and other key variables across both the video/speed data and the respirometry experiment data.

Widely varying numbers of taxa and specimens will affect statistical considerations such as p-value thresholds and, in turn, the strength of the statements. In cases where there were only

one or two specimens, it might be more appropriate to report the analyses without those taxa and only add them where they can provide significant information.

>Even when sample sizes were low for certain species, we chose not to omit data and instead, we've made our sample sizes and comparisons transparent (see Table S3). Owing to the challenges of working with fragile and patchily occurring open ocean jelly plankton we think it is important to include all of the data.

Generally, the idea of the research is very interesting and could be quite useful for bio-inspired design if the sources of significance and overall results are clearly and consistently reported to the reader in each section of the paper.

>We thank Reviewer 1 for their generous feedback.

\*\*\*\*\*

Reviewer 2 Comments for the Author:

Damian-Serrano et al. have performed kinematic and energetic measurements upon a variety of salp species to evaluate how differences in the morphology of these colonies relates to their swimming. The central idea is that colonies organized in a linear arrangement offer less projected area to the oncoming flow, which results in faster and more economical swimming.

The authors' present a tour-de-force of experimental measurements from 17 species, collected in the field, under highly challenging conditions. The data collected here are truly impressive and offer a real gem to the literature, especially given the difficulty to find and work with these animals.

>We thank Reviewer 2 for their generous feedback.

As itemized below, the trouble with this study comes from the analysis of the measurements. There are numerous statements of findings not supported by any statistics, and the statistical comparisons between morphological categories are evaluated by t-tests without mention of standard adjustments like a Bonferroni correction.

>We removed anecdotal comparison statements from the text. We changed the T-tests to ANOVAs with post-hoc pairwise analyses and reported the adjusted p-values for those instead.

There are lesser issues, such as the sample sizes (though itemized in the supplemental materials) are largely not mentioned in the manuscript, and values are not plotted in the figures.

>We added a table to supplement (S3) with a sample size summary per species and some mean values of descriptive variables such as swimming speeds and zooid sizes. We also

reported sample sizes in the Methods: “We analyzed the swimming behavior of salp colonies arranged in linear (six species, 64 specimens), bipinnate (three species, 17 specimens), whorl (three species, 10 specimens), cluster (two species, eight specimens), and transversal (one species, two specimens) architectures, with oblique and helical architectures represented by a single specimen.”

There is mention of evolutionary trends in the Discussion, but it is unclear to what extent shared ancestry factors into the measured patterns. It is not necessarily a requirement (in my opinion) to perform independent contrast analysis, or some other way of correcting for phylogeny, but a consideration of phylogeny would help inform the extent to which the measurements (especially those related to COT) are due to ancestral relationships.

>We agree that the PICs are not a requirement in this case, since the relationships between “traits” are mechanical and thus can be considered present-time-driven. At no point do we assume or expect the relationships between speed, shape, or efficiency to be a result of them co-evolving with one another along the branches, but rather a product of present mechanics derived from colonial form. We added details in the discussion around potential phylogenetic signal in the residuals or in the error terms for other unaccounted variables.

In short, the statistics employed in this study are not up to contemporary standards to the extent that I am not convinced of any of the findings. Fortunately, I know these investigators are capable of a rigorous data analysis and I hope they are willing to revise this work.

#### SPECIFIC COMMENTS

L38 - Unclear meaning of “scaling frontal cross-sectional area,”

>We changed it to “architectures with a constant motion-orthogonal frontal area as well as in those where this area increases with the number of zooids,”

L57 - Perhaps “neurophysiologically” or “nervous” instead of “neurologically”. “Neurology” has biomedical connotations.

>We changed it to neurophysiologically

It is difficult to make out from figure 1 where the oral and atrial siphons are located. What are the open circles and closed circles on the zooid shapes?

>We added in the caption an indication that the full black circles are viscera and the open circles are siphons.

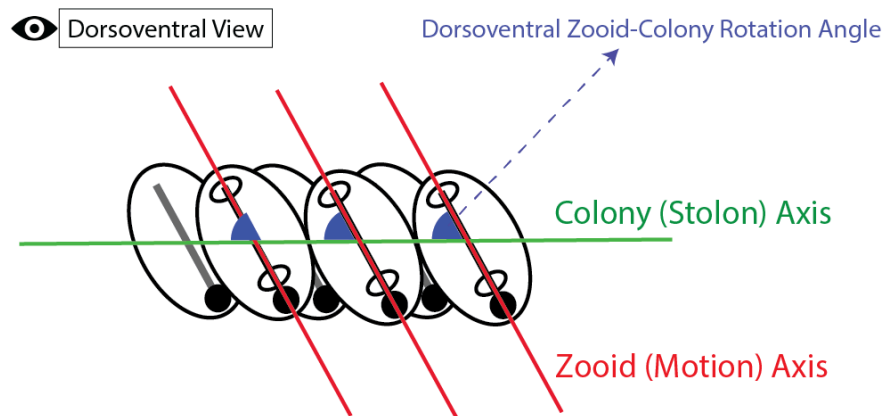
L80-105 - There is an issue with the consideration of frontal area with respect to drag here and in the rest of the paper. The thesis offered here is that more zooids serve to enhance thrust, but their arrangement with respect to the

prevailing direction of motion determines the frontal area, and hence, drag on the colony. However, I think readers can reasonably wonder what the implications of the colony arrangement are for the hydrodynamic interactions between zooids, which could affect thrust and drag production. For example, isn't it not possible that a zooid in a chain encounters enhanced drag due (as opposed to a cluster, for example) the jet of the zooid in front of it? It seems to me that the arrangement of zooids could have adverse or beneficial effects on thrust production as well. Your working hypothesis is fine, but you should acknowledge the complexity that it neglects. Alternatively, you may have evidence from prior work for rejecting this complexity, which should be explained.

>We cited Sutherland et al 2024 & Sutherland & Weihs 2017 paper in the Discussion while explaining how jet plumes are non-interacting in linear chains. Non-linear colonies have far broader separation (visibly so) between jet plumes than linear chains (personal observation from several coauthors while diving and imaging jet plumes using fluorescein).

L134- It is not clear on what plane and with respect to what axis these angles are defined. A schematic drawing would help.

>We added a supplementary figure with the schematic of the DV zooid-colony rotation angle (Fig. S10).



Faster motion is generally associated with a higher cost of transport, which may hold true for some jetters (e.g., 10.1088/1748-3190/ab57e4). I think I get why that might not be expected here, but it would be helpful for readers if you explained why that expectation may not hold true in this case. It would additionally be helpful to describe relationships between COT and speed in other jet-propelled animals.

>We added a paragraph to the Introduction stating that "In some single-jetters, swimming speed can be inversely proportional to the cost of transport (Bi & Zhu 2019) due to a highly inefficient refill phase in the jetting cycle which requires costly acceleration forces to reach high swimming speeds. In an asynchronous multi-jet system, inertia and continuous pulsation help maintain

speed without pronounced acceleration phases. We expect higher swimming speeds to be explained by form factors related to drag reduction, thus we hypothesize that energetic efficiency (cost of transport per unit of distance) to be higher in taxa with higher swimming speeds.”

L179 - These systems need to be described in greater detail. Are they all two-camera systems? How were the two cameras synced? In what volume can they interrogate? What are their frame rates and spatial resolution?

>We added details on the two-camera stereovideography system: “The primary stereovideography system was comprised of two synchronized high-resolution cameras (Z Cam E2 and Sync Cable; 4K at 60 or 120 fps) with 17mm f/1.8 lenses (Olympus M.Zuiko Digital) housed in custom aluminum housings (Sexton Company). Each field of view was 23 x 42 mm and in-focus depth was 20-25 mm. The image from the right-hand camera was viewed using an external monitor (Aquatica Digital), and illumination was provided with two 10,000-lumen lights (Keldan). An L-shaped plastic framer helped the videographer position colonies in the field of view of both cameras. Before diving, the stereo system was calibrated in a swimming pool using a cube with reflective landmarks. Calibration images were processed using the CAL software package (SeaGIS measurement science).” + “These two-dimensional videos were collected using a Sony FDR-AX700 4K Camcorder (3840x2160 pixels, 60-120 fps) with a Gates Underwater Housing using brightfield illumination (Colin et al 2022) or darkfield illumination.”

L185 - How were the positions in EventMeasure recorded? How was the 3D spatial calibration performed?

>We added details on how the calibration was performed (see above) and on how the points were selected in EventMeasure in the Methods.

L219 - Explain that coordinates were selected manually (if true).

>We added an explanation of the manual selection to the text.

L298 - It is unclear to me how t-tests could resolve differences between so many species. If they were performed in all possible pairwise comparisons, then that approach would lack statistical power and something like an ANOVA (with post-hoc comparisons) or a mixed-effects model would be better. If I am misunderstanding the design, then please explain in the text.

>We switched from t-tests to ANOVA with post hoc comparisons for both speed and COT comparisons across colonial architectures and reported the adjusted p-values in Tables S4 and S5.

When there were repeated measures, were the mean values used for a species? It should have been or there would be pseudo replication in the analysis. A statistical analysis that takes



advantage of those repeated measures would offer a more powerful approach for resolving a significant trend.

>Each specimen was measured only under a single “treatment”, no colonies were measured at different swimming speeds or colony sizes, thus there are no repeated measures across the different factors (colonial architecture, zooid number etc.).

Since the sample sizes are so variable among the species, it would be helpful to report them along with each mean value and p-value provided in the text, or perhaps a new table.

>We reported sample sizes for speed and respirometry across species in the new Table S3 in the supplement.

L317-327 - Since you have not performed stats to demonstrate differences between species, then you cannot offer statements about the relative speed of species. Those differences could be merely due to chance and not a species-specific signal.

>We removed non-supported claims on species differences.

An issue with either comparisons between species of colony formation is that you have not taken into account shared ancestry. In the absence of phylogenetic information, it could very well be that the differences in formation observed may be attributed to unmeasured traits that have to do with similarity in relatedness. Given contemporary expectations for comparative methods, the authors should address these concerns either by incorporating considerations of phylogeny in the statistics or by discussing the possible role played by the relatedness of the species later in the paper.

>A complete phylogenetic analysis of ancestral states of colony formation was reported in Damian-Serrano et al (2023) IOB. While colonial architecture is phylogenetically structured, we assume its effects on swimming are independent of phylogeny, thus the contrasts within the scope of this study are independent of evolutionary effects. We added a portion to the discussion on the potential role of relatedness in the unexplained variation and on the evolutionary history of colonial architecture.

L336 - It increasing looks problematic that t-tests were used for the statics in these p-values (see concerns above).

>Those are not t-tests, those are slopes in a linear regression and their p-values. We clarified this in the text and added adjusted R<sup>2</sup> values. We removed the t-test asterisks and brackets from the box and bar plot the figures, and added supplementary tables S4 and S5 with the ANOVA Tukey posthoc contrasts and their adjusted p values.

L397 - This should be explained in Methods, not here. How does the model handle

repeated measures? Species?

>We removed these redundant Methods from the Results. We explain in the Methods how we use multiple experiments (and specimens) from each species to obtain multiple values of the differences between variable swimming rates and a single (mean point) anesthetized rate value. We reworked the text in the Methods to make it clearer: “To capture variability within species, we calculated the mean respiration rate of anesthetized specimens for each species and subtracted it from each intact specimen’s total respiration rate to get multiple swimming-specific rate values within each species.”. Each swimming specimen is independent of each other (they are different specimens collected separately, not the same one under multiple measures) within each species. We did perform paired analyses (intact and anesthetized within the same individual at different times) for sensitivity purposes, finding that using different specimens is just as reliable, but the results from those analyses are not included in manuscript.

L389-419 - As for speed, you cannot offer statements about differences in energetics between species without stats because these values could merely reflect chance events.

>We removed these anecdotal claims, only including statistically supported claims.

L422 - Under what test?

>We used the linear regression non-zero vs. zero slope p-value included in the summary of the `lm()` function output in R. This comes from an internal t-test on the least-square values of the data points relative to the line under a flat slope vs an inclined one.

In the interest of data transparency, the figure legends should make the sample sizes clear and ideally should also plot the individual measurements.

>We overlaid semitransparent points with the individual measurements onto the specimen mean-collapsed regression plots in Figures 3,4, and 6. We also added the individual specimen points (with jitter and semitransparent) to the species-mean bar plots in Figure 5. The box plots in Figure 2 already show dispersal well and were constructed using the individual points, thus we did not add individual measurement points to those. We included information on sample sizes in a new table (S3).

L569-582 - Incorporating these phylogenetic patterns into your analysis would be very helpful. e.g. how many independent evolutions are included in the species that you sampled? For the species that you considered, it would be informative to know if their ancestors were linear or not.

>We added details from our prior study, Damian-Serrano et al. (2023), to answer that question in the discussion. “Linear chain architectures evolved independently in *M. hexagona*, *S. zonaria*, *I. punctata*, and before the common ancestor of *Iasis* and *Salpa*.” and “slower swimmers (such

as transversal chains, helical chains, whorls, and clusters in the genus *Pegea* and the Cyclosalpidae family) had also evolved from oblique and linear ancestors.”