

In the manuscript entitled “A comparison of approaches to estimate muscle forces, validated by *in vivo* bite forces.” Ginot and Blanke use different methods to estimate the (physiological) cross-sectional area of the mandible closer muscles in a grasshopper, and then compare mechanical predictions based on the obtained values with *in vivo* bite force data. The authors conclude that a dissection-based method gives the best result, whereas cross-sectional area measurements via 3D image analyses leads to a significant over- or underestimation of bite forces.

The authors address a highly relevant and important question: how variable/accurate are different methods to measure muscle cross-sectional area? The answer to this question has significant implications for modelling efforts, which often may be the only way to estimate performance (for example in rare or small specimen). As such, I find the aim of the manuscript highly valuable. However, and unfortunately, I have strong concerns about the analysis and interpretation of the data, and am thus currently unable to recommend publication of this work. I appreciate that it is frustrating for the authors that they now have to deal with a third referee – I tried my best to keep my comments brief and focussed. I also carefully read through the first reviews, to avoid any repeats.

I first briefly summarise my main concerns; further detailed comments then follow. I tried my best to be helpful, clear, constructive and specific. I am happy to be contacted by the authors to discuss any points in detail, where this can aid a more speedy and collegiate review process.

Sincerely

David Labonte, London 23/04.

Main concerns

1. From my reading of the methods, and an equivalent comment by the first referee, it would appear that the authors used the PCSA of both mandible closer muscles (ie left and right hemisphere combined) to estimate bite forces. Unless my understanding of the sensor used is incorrect, this is a physical error: Only one of the two bite plates is connected to the sensor, and they thus measure the force produced by only one of the two closer muscles. The estimated forces are thus all too large by a factor of two. Correction of this error would seem to significantly alter the conclusions: The dissection-based estimate would be too low, and the Convex-hull-methods would be in closer agreement with the measured data. I notice that the authors replied to the same concern raised by referee 1, and here briefly respond to their arguments.
 1. The authors argue that the agreement between the estimate obtained via the area of both closer muscles and *in vivo* bite force measurements suggests that their estimate is correct. This argument is circular (there is no reason to assume *a priori* that any of the methods should yield a result that is in close agreement with the experimental data – this is what the authors seek to test), and unphysical: unless I misunderstand the sensor design, the sensor measures the force applied by one mandible only, so the relevant muscle volume is that of only one muscle, regardless of how well an estimate using the entire volume agrees with the *in vivo* data).
 2. The authors ask which of the two closer muscle volumes should be used, in light of the asymmetry: The volume of whichever muscle was connected to the mandible which applied a force to the bite plate connected to the sensor. Did the authors measure all individuals with the same side?
 3. The authors point out that the asymmetry itself poses interesting biological questions. I agree, and I do not have a sound functional interpretation for why such asymmetry may arise. But this has no bearing on the problem at hand. Provided that my understanding of

the sensor design is correct, the two mandibles can produce disparate forces without causing any problem to the measurement – the different forces are equilibrated by reactions from mechanically decoupled parts of the sensor. In “real” biting, equilibrium may then involve different opening angles of the left and right mandible.

2. The authors use a muscle stress estimate which itself is affected by some of the (fair and valid) concerns they raise with their work: Taylor obtained stress estimates from *in vivo* force measurements: bite forces were converted into a stress *using an estimate of the PCSA* – but how accurate was that measurement, and how does it compare to the methods used by the authors in this work? As the authors convincingly show, different methods to measure the PCSA can differ substantially. In order to use Taylor’s stress value, the authors would thus need to establish how their PCSA measurements compare to the method used by Taylor.
3. I politely ask the authors to provide a more explicit definition of the PCSA – the metric which is the focus of their work. To explain my concern briefly: the cross-sectional area of muscle fibres varies with contractile state (as muscle is approximately incompressible). At which contractile state do the authors define the PCSA, and how does this state compare to the state of the fibres they measured experimentally? One choice may be to use the cross-sectional area of a relaxed fibre, as is commonly done in the vertebrate literature. However, the authors report mandibles were fully closed, so presumably the muscle fibres were measured in contracted state? Either way, there is no *a priori* reason why the cross-sectional area of a relaxed fibre must correspond to the area at which the muscle fibre produces maximum stress (which occurs for optimal overlap between thin and thick filaments). This also adds further difficulty to the use of the stress-data from the work by Taylor, who did not investigate/discuss whether the force measurements were obtained for an “optimal” angle/fibre length (nor is there a reason to assume that the apodeme area used in the paper to estimate the PCSA reflects the PCSA at maximum force). We have previously argued that PCSA should be defined strictly as the area of the fibre at the stretch where it produces maximum stress (as this would allow meaningful comparison of stresses/areas across muscles), but this would require knowledge of the force-length relationship. One way to resolve this problem may be to focus on a relative comparison of the area measurements, ie the variation between the methods, rather than an absolute comparison to *in vivo* bite forces.
4. For the above two reasons, I am not convinced the authors will be able to draw strong conclusions on which PCSA estimate is most “accurate” on the basis of the data they provide. To avoid the need for additional experiments, I suggest to resolve this by splitting the analysis in two parts:
 1. Address the question which of the area estimates are consistent with simple theoretical expectation. The authors currently assess the correlation between PCSA estimates and bite forces, but they can make this section stronger – there should be a linear relation between force and area, with an intercept of zero. This expectation may be assessed quantitatively as follows: Plot the measured bite force against the area estimate on a log-log scale. The slope extracted should not be significantly different from unity. Where it is, the area method is erroneous. From the presented results, it would seem this may exclude one measurement technique (which is a rather interesting result). The authors currently provide two plots similar in spirit (Fig 5), but these plots suffer from the problems highlighted above, and treat the stress-estimate as an error-free value.
 2. Address the question what stresses would be required for the different PCSA estimates to explain the observed bite forces. These stresses can be conveniently extracted from the same analysis as the slopes – they are related to the intercept of the respective log-log regressions. The advantage of this analysis is that it will also return confidence intervals, so permitting a more robust statistical comparison with whatever range of stresses may be reasonable expected (see next point).
 3. Compare the stress estimates with independent estimates/reasonable expectations. The authors would need to carefully consider error propagation/uncertainty in this process.

For example, the regression estimate via the Taylor-data has uncertainty. If the authors wish to draw strong conclusion on which method is “best” in absolute terms, they will need to consider this and other experimental uncertainties via error propagation (or similar), and demonstrate statistically that some methods produce results that are within the error margin, and some produce results which are not. I think that it is quite a tough task, and it may be preferable to instead focus on the relative differences between the methods (which are rather striking!), and to provide only a softly-worded plausibility assessment of different stress estimates/sources of error in PCSA estimates.

Detailed comments

Introduction

1. L39: Another appropriate citation is: Heethoff, M. & Norton, R. A. A new use for synchrotron X-ray microtomography: three-dimensional biomechanical modeling of chelicerate mouthparts and calculation of theoretical bite forces Invertebrate Biology, Wiley Online Library, 2009, 128, 332-339
2. L42: I believe “fewer degrees of freedom” is correct (countable)
3. L50. Please write the full term at first use (yes, it is in the abstract, but it will make the reader’s life easier)
4. L55: I politely disagree – the methods still “only” estimate morphological data – without a stress estimate, no force can be computed. How, then, are these “true mechanical models” of absolute forces?
5. L62: The authors themselves have used μ CT data for mechanical modelling to great success in several studies. Why not include these references here?

Methods

6. L108: From my understanding of this sensor design, only one plate transmits the force to the piezo force transducer. Please clarify.
7. L108: It appears that the authors do not correct for the (likely) misalignment between applied force and the force that can be measured (the force perpendicular to the bite plate). I am ok with this simplification (the error is probably small compared to inter-individual variation), but this limitation should be mentioned and acknowledged, in particular where the authors seek to verify model predictions against experimental data.
8. L121: Please be more specific. What does “lack of proper measurement” mean (explained in discussion – move here)?
9. L123: Please add a definition of opening angle (move here from below, where it first appears?).
10. L124: Why are cockroaches the right comparison here? From my reading of Fig. 4 in Tom Weihmann’s work, the range 60-80 degrees is more or less the entire measured opening angle range in cockroaches, and the average force varies systematically by a factor of more than 3 in this region (which appears to be a variation of similar order if not larger than the variation across different PCSA metrics reported by the authors). A variation of similar magnitude in this opening angle range was found in our work on ants (Pueffel et al 2023). Is a likely variation in bite force of that magnitude really small in light of the model-test the authors aim to conduct?
11. L162: And the apodemes?
12. L164: Please clarify: were the mandibles still attached when the muscles were weighed (and apologies if I have missed this)?
13. L196: Please add a brief explanation of the difference between convex hull volume and 3D convex hull
14. L210: Please provide an explicit definition of in- and out-lever at first appearance.

15. L213: Meaning and significance of “initially” unclear here – the “force amplification” remains correct instantaneously throughout the contraction?
16. L216-221: This is a nice method! Please state explicitly the implicit assumption that the line connecting the two condyles is a good proxy for the axis of rotation (and thus that the triangle height lies in a plane of rotation by definition). Please also clarify the implicit assumption about the orientation of the muscle force (to my understanding, the calculation conducted by the authors assumes that the muscle force is perpendicular to the height of the triangle formed with the insertion point – that is clearly wrong across opening angles, but may be roughly right for the angles measured by the authors? I now saw that they account for this later in the manuscript. My concern here would probably be addressed if the authors provided a clear, explicit definition of what they mean by in- and outlever straight away).
17. L223: Please provide a clear and explicit definition of PCSA. To clarify the difficulty I have here (because I appreciate the authors give an equation which may be seen as a definition): the cross-sectional area of the fibre changes as the fibre contracts. So at which contractile state do the authors define PCSA (and how does this compare to the state in which they measured the fibres)? The stress also varies with fibre length, and is not necessarily maximum for an unstretched fibre.
18. L224: A point of contention (sorry): I know that the PCSA is often defined like this in the vertebrate literature, but from a physical perspective, I politely object to this definition: The PCSA of a fibre does not depend on its pennation angle. If the authors want to stick with this definition nonetheless, I ask them kindly to label it as “effective PCSA”, and specify that they define it as the PCSA of a non-pennate muscle which exerts the same force.
19. L226: I believe it is “classic” (classical is the music). I am not sure this term fits here, though – “typical”? “established”?
20. L235: How was this measured in this study?
21. Can the authors please be explicit somewhere: the volumes and PCSAs they calculate are for both head hemispheres?
22. L238: Just “..know the maximum muscle stress” (though this depends on the definition of PCSA used by the authors, see above). I note that the stress can be expressed in any unit, as long as it has dimension force per area.
23. L243: Please specify: (I) did you use the general regression line from Taylor, ie using data from all species, or just data from arthropods? (II) Provide an accuracy of this estimate (there is variation about the regression line).
24. L237-261: I have the following objection to this approach: The calculation of stress values in Taylor 2000 in itself relied on a measurement of a muscle area (to convert the measured forces into a stress). This area measurement was approximate, and differs from all methods used in this work: from my reading of the paper, Taylor estimated the PCSA via an approximate insertion area on the apodeme. The area is thus likely overestimated, and the stresses thus underestimated. Can the authors clarify why this stress estimate is a reasonable proxy that is not influenced by the very methodological problems they point out in their paper?
25. L277: The effect of muscle stress variation due to variations in thick-thin filament overlap is likely much larger than the effect of the MA variation. I understand that the authors cannot account for this in this work, but I think they should alert the reader that it is a significant factor.

Results

26. L326: I was unable to find estimates/CIs for allometric slopes in SI 6.
27. L330ff: Please specify the average length and deviation with a number of significant figures that reflects measurement accuracy (you were not able to resolve 10nm let alone 1nm, I assume? Check throughout manuscript for other instances of this problem). A logged length

- still has a unit associated with it (and so does the stress). The estimated stress should have an error associated with it, reflecting the error of the regression from which it was estimated.
28. L346: A key result here would be to inspect if the relative differences between left and right hemisphere for all metrics is independent of the measurement method. Why not conduct an ANOVA with method and hemisphere as main effects, and inspect their interaction?
 29. L365: Unless I misunderstood something, fibre length was measured with the same (one) method for CT data (in contrast to PCSA, for which different 3D methods were used)? “3D” methods is a bit ambiguous/misleading – perhaps “image-based” and “weight-based” (or similar) is clearer here.
 30. L366-268: Why not use an ANOVA with method and hemisphere as main effect, and inspect the interaction?
 31. L377L Please provide all test statistics in a table in the SI
 32. L379ff: The analysis in this section could be improved by distinguishing more explicitly between (a) the correlation between area and bite forces; and (b) the required stress to render this correlation into an equality. Strictly, the expectation should be that all bite forces are correlated with PCSA such that the slope of log-log transformed data should not be different from unity (or, equally, that a regression in linear space has an intercept of zero, and a slope of unity). That is to say that, through inspection of the CIs, the authors can make a strong inference on the quantitative agreement between area and force, regardless of stress (they currently only report only a weaker inference based on correlation). Subsequently, the authors may inspect the intercept of the log-log regression or the slope of the linear regression to obtain a direct estimate of the stress – this again would have the advantage of presenting a quantitative test: both metrics would have a confidence interval, and the authors can then test whether this confidence interval includes reasonable values.

Discussion

33. L415: A reduction in area by a factor of three from changes in diameter alone would imply a reduction in diameter by a factor $\sqrt{3} = 1.75$. Can the authors confirm this approximate difference via small number of manual measurements on both the μ CT data, and the images from which fibre length was measured? This would significantly increase the confidence in the interpretation that the difference in PCSA estimates indeed reflects shrinkage.
34. L420: Is there not at least one other option: the weight estimate could be inflated, because the authors also weighed the apodeme, (perhaps also the mandibles, see above), and all connective tissue around and between muscle fibres? As referee 1 commented, it would also appear likely that additional liquid mass is measured. It may well be so that this method is often used in larger muscles, but it could be a problem here, due to the larger surface-to-volume ratio of small muscle. A manual measurement of fibre diameters as alluded to above would help to discount this alternative explanation.
35. L428: Can it really be safely ruled out? The Huang-, Max_Entropy, and Yen-method, for example, appear to yield cross-sectional area estimates about twice as large as the “default” method used in the paper. Sure, the correlation still appears poor (though I believe this appearance is at least partially an artefact of this particular representation of the data), but that question seems different from the question of an average agreement in magnitude. So I would ask the authors kindly to convince me why the opposite is not true: The *a posteriori* comparison seems to strongly suggest that the difference in magnitude may well be an artefact of the used threshold method. As suggested above, a manual measurement of fibre diameters from unsegmented CT data would help to resolve this.
36. L431-433: On the basis of the above observations, and in particular Fig S1, I am not convinced by this conclusion.
37. L415-433: I would argue that the finding that shrinkage may occur in itself is not surprising (though the seeming magnitude of the effect arguably would be). However, what I am really surprised by is not addressed by the authors: The seeming absence of a correlation between the area estimate and the bite forces. This result would imply differential shrinkage between

larger and smaller muscle volumes. Can the authors comment on this – has this been observed elsewhere? What are possible explanations for this differential shrinkage? I would expect that the fibre diameter is approximately constant across individuals, and differences in volume are largely due to difference in fibre number and length. Naively, I would also expect shrinkage to be a constant fraction of the fibre diameter. Why is it not?

38. L465: I recommend rephrasing this carefully (the same issue appears elsewhere, including the title): The authors used one single biomechanical model, and this model has no direct relation to the choice of “classic” dissection methods vs 3D imaging methods – only how the authors measure one of the involved parameters differs.

Figures & Tables

1. Please provide the units for all numbers in tables.
2. Please specify the meaning of abbreviations in both table captions.
3. Please provide the paired bite force-PCSA data in the SI.
4. Fig 2: Please provide a scale bar for A4
5. Fig 3: Please provide a scale bar for A