nature portfolio

Peer Review File

Diverse mechanical properties, composition, and performance of shark dermal denticles

Corresponding Author: Dr Sheng-Feng Shen

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Attachments originally included by the reviewers as part of their assessment can be found at the end of this file.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The manuscript “Diverse Forms of Denticles Facilitate the Evolutionary Success of Sharks” is a study on structure-property-function relationships in the dermal denticles of sharks, and implications on their evolution. The study includes a massive study on the morphology of the denticles of 158 species based on SEM images and a very clever automated feature recognition approach based on deep learning to categorize the denticles into 10 morphogroups. The study also includes mineral and fluoride content, indentation hardness and modulus, as well as hydrodynamic simulations to estimate drag on 20-30 samples from each of these 10 morphogroups. The authors draw conclusions on structure-property-function, for examples the morphology of some groups is more suited for drag reduction in smaller, faster shark species, while the morphology and mechanical properties of other groups is more suited for protection in slower, large sharks. The authors also draw some conclusions on the evolution of sharks, for example by tying a rapid evolution in denticle with the emergence of the Atlantic Ocean.
The design of this study is extremely strong (although spacing between denticles could also have been considered), the amount of data is massive and as far as I can tell the analysis is solid. My only two concerns are as follows: I feel like this study essentially confirms what is already known about the morphology-function relationships in dermal denticles, and about specialization in shape in relation to ecological niche and lifestyle in different species. While it is important to confirm this knowledge with a strong, comprehensive study, the data and analysis do not provide knowledge that is particularly new. Specialization and differentiations of morphology under ecological pressure has been known since Darwin, and the results reported here in terms of the evolutionary patterns in dermal denticles seem somewhat in line with what should be expected in this context. Is there any oddity, new evolutionary patterns, or an unexpected function of the dermal denticles revealed in this study? Again the results are strong, but I am not sure that the degree of novelty is up to what should be expected in this journal. My second concern is the way the manuscript reads: the manuscript is very well prepared, but at several places the level of details made me feel like a more specialized journal would be a better fit for this report.

Reviewer #2

(Remarks to the Author)
In the manuscript entitled “Diverse Forms of Denticles Facilitate the Evolutionary Success of Sharks”, authors propose a new method to analyze the morphological features of dermal denticles based on deep learning. Utilizing the new method, combined with multiple analyses, they conclude that the diversity of denticles increased at around 170 million years ago, which may have enabled sharks to adapt various environments.
Traditionally, morphological analyses on various forms were conducted by expert human observers, which may tend to depend on personal experiences. The new method proposed in this study enables to conduct objective evaluation, which has a potential to make progress in wide fields of earth science related to observation, such as microfossils, mineral grains and so on. In addition, the conclusion from the proposed observation method is interesting, although I’m not the expert in this field.

Given the wide applicability of the proposed methods and the impact of the conclusions, I would recommend that the manuscript be accepted to Nature Communications after a few questions are cleared up.

Major comments:
[L. 93~] K-means clustering analysis (KCA) was conducted with number of clusters set at 20, and after checking results, class number was manually reduced to 10. Although I understand such a process is practically useful, it reduces the "objectivity" of the results. It would be preferable to try with a smaller number of clusters (K=10 for example) and describe whether the classification is valid or not.

[L.140~] I can't understand how authors determined the timing (geologic age) of evolution. Did they analyze the literature collections of denticles with known age?
Since the dating information significantly affects the conclusions in this manuscript, a more careful description would be desirable.

Minor comments and checks:
[L. 101] What does "consistent" mean in this context? Is it possible to visualize or describe the "consistency"?

[L. 419] Should the full spelling of the PCA be listed here?

[Ref. 11] Duplicated hyphen ("690--700").

[Ref. 58] Since this manuscript is published in a peer-reviewed journal, citation from arXiv should be avoided unless there are special circumstances. Also, the spelling of the first author should be "He, K.".
https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Reviewer #3

(Remarks to the Author)
The paper analyzes dermal denticles of sharks using developed quantitative methods to better understand the impact of dermal denticles on how sharks adapt to various environments. The paper adopts nanoindentation, electron probe micro-analysis, and computational fluid dynamics to establish mechanical properties, elemental composition and swimming performance. Concretely, the paper uses deep learning to learn visual features and clustering them into a predefined number of groups. It further aligns clustering results with the clades of the phylogenetic tree of sharks. Based on these clusters of dermal, the paper conducts a series of mechanical, elemental, and fluid dynamics experiments to investigate the characteristics of these different denticle morphologies. Results lead to multiple conclusions, e.g., denticle ridges are highly variable and present more functional properties beyond drag reduction. Moreover, it suggests a radiation event within the ground sharks around 170MYA, which results in the enormous diversity of the clades where the ridged dermal denticles may be the key.

The paper is generally written well. It has certain merits and contains clear documentations and important details. As a non-expert in life sciences, I enjoy reading this paper. My comments are as below.

- The paper automatically creates some morphotypes of shark dermal denticles by using deep learning (i.e., the variational sparse coding) and clustering algorithms, aligning the clusters with a predefined phylogenetic tree, and analyze these clusters in various aspects such as elemental composition and fluid dynamics. From the writing, my understanding here is that the paper makes an assumption that the dermal denticle clusters are correlated with phylogeny, and analyzing them can reveal evolution events. However, if a goal is to understand evolution through analyzing shark dermal denticles, is it a better clustering strategy to sample and cluster denticles within predefined groups of shark species and do the analysis? Can authors discuss the above?

- Line71. Please cite papers to give reference for "a recently published phylogenetic tree".

- The paper uses Variational Sparse Coding (VSC) to learn visual features. The paper states in Line400 that VSC is an improved version of Variational AutoEncoder (VAE). Yet, VSC is less recognized in the literature. Can authors clarify why VSC is more preferred in the study than VAE and other well-established methods? Moreover, in Line417, can authors explain the regularization of features to the spike-and-slab distribution? How does this regularization help learning the visual features in the study?

- Line412. It is imprecise to write "the decoder, which is symmetrical to the encoder". A more precise statement is that encoder and decoder are symmetrical in terms of network architectures.

- Line428 writes that "we wanted our clustering results to match with the clades of the phylogenetic tree of sharks"; Line446-456 writes how to manually post-process the clustering results, presumably for aligning the results to the clades of the phylogenetic tree of sharks. Does it mean that the clustering results actually do not align with the phylogenetic tree and hence require some manual post-processing? Can authors clarify and discuss? What are the assumptions made here?

- How many shark individuals are used in the study? I can see 2193 SEM images of skin, 53 body locations across 132 extant shark species, 2 ray species and 24 fossil species. Do different individual sharks have distinct denticle patterns? Can

authors clarify?

- Line85. Why should fossil species be included in the training data along with extant ones? Does this assume fossil shark species share the same/similar dermal denticle morphotype? What if training on extant/modern specimens and see if the trained model generalizes to fossil? Can authors clarify?

- Line94. The paper excludes any groups represented by <3 F1-body location skin samples for the purpose of simplifying the representations and for using this particular location to compare denticle morphology of different species. I would like to see more discussions here. For example, does it assume (or is it a fact) that dermal denticles on the whole shark body share similar patterns with F1-body locations? Why is F1-body location so special (e.g., because they are abundant in specimens, or easy to obtain)? Why do some clusters have few examples of F1-body locations -- is it because dermal denticles from other body locations quite different from those at the F1-body location?

Reviewer #4

(Remarks to the Author)
This paper represents a very large effort to quantify shark denticle morphology across 70+ species using deep learning computer vision, measure material properties using nanoindentation of the enameloid layer across multiple body regions of 10 species, measure the drag properties of denticle models of 10 denticle types using multiple regions of the body from 8 species using CFD, integrate functional variables together in morphospace, and run some phylogenetic comparative analyses including trait correlations, BAMM, and ancestral state reconstruction.

I think this manuscript represents a really big and collaborative effort and includes valuable and novel datasets that allow for really interesting progress to be made on understanding denticle morphology, denticle function and performance, and denticle evolution. However, I think the expansive nature of the manuscript makes it very difficult to digest and makes it very unclear for readers – in part due to the lack of explanation of methods and results in many instances, but also because of the many different questions that this manuscript is answering without really clear organization, framing, or justification in many cases.

I think this manuscript could potentially be a great value to the field, but there is enough information lacking that over the course of my review I became increasingly skeptical of both the results, methods, and ultimately therefore some of the conclusions presented here. I have detailed a number of suggestions below but I would also outline a few larger comments here.

1) Methods and results need to be explained in more detail. For example, I had a number of questions while reading the manuscript, many of which are given further detail in the line-numbered comments below.
a) How did you setup your CFD analyses in terms of speed, model size, and Reynolds number? Are you testing across different Reynolds numbers or holding it the same?
b) How did you calculate correlations among your functional measurements (e.g. fluorine content, hardness, drag reduction, stiffness) and where are those data and results?
c) What are the definitions of hardness and stiffness that you are using here? Specifics are needed.
d) How did you decide to deal with variation in your morphological dataset that is due to body size? I'm guessing you didn't correct or account for body size at all so you need to justify that choice and probably also show how much of the variation in your morphological dataset is due to size.
e) Similarly, did you look at the relationship between body size and any of your functional variables?
f) How did you map the functional traits back to your morphospace using functional trait analysis? Also, what is functional trait analysis, how was it employed, and where are its results aside from Fig 6a?
g) Why did you choose 10 morphogroups? The current reading of your manuscript makes this feel arbitrary after you make the point that previous morphological categorizations are based on arbitrary expert opinions so it feels like you are doing a similar thing here. I think more justification would really help.
h) Why is morphogroup 5 called thick ridge but one of the examples in Fig 1 doesn't really have ridges?

2) For the hardness and stiffness measurements, you are measuring these variables on cross sections of the denticles from directions that the material seems unlikely to be loaded on. For example, I would guess that denticles are often scratched or loaded by pushing on the crown towards the skin, but here your measurements are parallel to the skin and therefore orthogonal to my assumed load direction. Please spend some time justifying this. Do you think this is fair and how do you think this might be influencing your results? My assumption is that this tissue is likely to be organized in some way that would create anisotropy among different measurement orientations.

Are there sources from other mineralized tissues (e.g. bone) or morphological studies on the fine structure of the elasmoid layer that would suggest the material has similar properties regardless of direction? My intuition is that the opposite is often true.

3) A very interesting finding in your paper is that you say hardness and mineral content are negatively correlated. My intuition (and the literature) would suggest is that the reverse should be true and would be assumed here – more mineral content would create a harder tissue/material. So on my reading of your paper, this either seems like a really important and novel finding that could use a good deal more time spent on it or perhaps it's some kind of error.

4) What are the major ways your deep learning computer vision is using to separate out denticle 'morphogroups'? The methods here are really cool, but without some idea of what differences these morphogroups represent, it's difficult for readers to get a sense for what is important to distinguish your morphogroups. Also, providing some sort of framework here so that others can adopt and use your morphogroups would be really valuable and would increase the general impact of your results.

5) I noticed that for morphogroup 7, the "smooth" morphogroup, that the S region from Scoliodon laticaudus is used, but in Fig S3 the CT scan images from this sample are obviously ridged. Are the samples used to represent the different morphogroups for minteral content, hardness/stiffness, and drag analyses actually categorized in those morphogroups in your previous exploration of denticle morphospace? If so, then why is the S region from Scoliodon laticaudus in the smooth morphogroup if it is not at all smooth surfaced?

6) It would be valuable in the supplement to include a multipanel figure showing for EACH morphogroup that shows readers the variability for each group. Something like 6-8 examples of each morphogroup would be really valuable to help readers understand what is actually different here among morphogroups and how much disparity is within vs between morphogroups.

7) Is it fair to assume that the functional traits found in 1-3 species in a morphogroup would be characteristic of all denticles from that morphogroup? This is an assumption you are making here that I think you could be open about and describe why you are making it and what limitations are inherent in it. (e.g. species might be different in material properties but similar in morphology (morphogroup)). You could instead frame it as if you are trying to understand the connections between form and function instead of assuming form is representative of function in these morphogroups.

Abstract, line 24: Specious is the incorrect word here – you likely were trying to use speciose.

Abstract, line 25: "although occupying ecologically diverse habitats" is incorrect grammar. Please fix this! Also, it's not clear to the reader why it's interesting/relevant for you to point out that these sharks live in different habitats yet all have denticle ridges. When re-writing this sentence, it might make more sense just to state it as a list of facts about Carcharhiniformes rather than using words like 'although' that imply a relationship between habitat and ridges that you have yet to explain at this point in the paper.

Abstract, line 28: concluding should probably be a word more like 'finding'.

Abstract, line 29: you start the abstract by mentioning just Carcharhiniformes and now have broadened your results to all extant sharks. This is confusing – did you study Carcharhiniformes or all extant sharks? Be specific with your language here to match your methods and results. My suggestions would be to remove or completely rewrite the first sentence (or move it towards the second half of the abstract). Your study is across all sharks, not just Carcharhiniformes.

Line 48: you mention the ecological function of some dermal denticle types are well known and then mention drag reduction. These papers are based on just 1-2 species so I'm not sure I'd really consider them to be statements on denticle 'types', but instead only proof that drag reduction seems possible in these species. Also, I'm not sure ecological function is what you mean. You might want to say something more like biomechanical function or just function.

Line 57-58: This sentence doesn't seem to have any connection to denticles – just about plate tectonics and shark diversity?

Line 60: "their morphological diversity" needs to be more specific. Something like "denticle morphological diversity".

Line 67: What is representation learning? Please elaborate or define this briefly when you mention it.

Line 68: Why do we think 'feature engineering' needs to be reduced? Give a reason or two here to further justify your methods and educate those not familiar with this field.

Line 70: How is this deep learning method more objective? I think you need to explain this a bit more or somewhere in this manuscript and reference that passage here. I feel like someone could argue that experts are not usually selecting features subjectively but rather objectively (based on facts and evidence).

Lines 72-76: Very cool!

Line 94: what are F1- body location skin samples? You haven't explained or introduced this term.

Lines 93-99: Okay so you used deep learning to quantify morphology for you and then k-means clustering to find 20 groups. Why did you initially start with 20 groups? This seems subjective. Provide justification here in your text.

Also, you then reduce 20 groups down to 10 by combining a few groups and then removing a few groups that have low

sample sizes (I think). These decisions seem somewhat arbitrary – especially given your section just above in the introduction that makes a point of how your methods are objective and not subjective – it feels like you are taking your 'objective' dataset and then doing 'subjective' things to change the results. I think you need to justify these decision more in your text.

I would say that the one of the major advancements you have made here with this method is to quantify denticle surface morphology in extreme detail (to the point that you can reconstruct images from your quantified results). And you have done so on images of multiple denticles, not just single ones.

Lines 81-92: An important part of most morphological studies is deciding how to account (or not account) for covariation in body size. Presumably you sampled denticles from an array of different body sizes but it doesn't seem like your analyses are accounting for body size at all – typically body size is responsible for much of variation seen in most morphological datasets, and in those cases researchers will use various methods to account for the variation due to body size and remove it so that they are studying size-independent variation. I think many readers will expect you to justify NOT correcting for size (I'm assuming you didn't because I can't find a mention of it). I think somewhere here might be a good place to add 2-5 sentences justifying NOT accounting for size. A justification might also be really strong if you have analyses of body size vs your morphological variation to perhaps show how much variation in your data is "due" to body size alone.

Lines 104-106: Referencing the figure is great, but your descriptions of the morphology of the denticle groups should be more elaborate. For example, if there are 7 groups with ridges on the crown, why did they get separated into different groups? Same with the smooth denticles (I assume this just refers to have no ridge, which would be a more specific description as smoothness/roughness may not correlate with ridge presence). Also, what is meant by highly structured. Please provide more information here or somewhere else in the paper so that readers get an understanding of your groupings.

Lines 113-114: "variation in denticle morphology across a single individual can be enormous" should have some citations.

Lines 117-119: Here you should give us an idea of how many species and individuals for which you had data on the flank (F1) region? In other words, tell us how much data went into this ancestral state reconstruction.

Line 132: "developed ridged denticles reflection their ecological niches". You make this statement as if there is a connection between ridges and ecology but I don't think you've discussed this at this point in the paper. I think you need to either spend some time in the introduction or here explaining this idea and providing support for it.

Line 180: shark should be sharks

Lines 183-192: As in above comments, you should provide some idea of sample sizes here – for many of these methods you are working with a subset of species and you need to be clear about that here so that readers don't get confused.

Line 195: It might be good to provide a citation for functional trait analysis and spend some time explaining how it was used here. I'm unfamiliar with this analysis and I suspect most other readers would be too. Yet, this functional morphospace is one of the most novel and interesting aspects of this paper so it is really important to clearly communicate to readers what is going on and how the functional trait analysis is performed.

Lines 198-199: Elaborate a bit into the transformation of data into performance ranks! Does this just mean each model was ranked from 1-10? How is this happening…also this ranking also presumably says nothing about statistical significance.

Why did you not also test for statistical significance among groups?

Line 205: You mention resisting demineralization, but I don't follow how you are measuring resistance to demineralization. Please add something about this interpretation either here when you first mention this or just in the paragraph above when you are briefly explaining your measurements. To me, demineralization resistance sounds like you are thinking about how these scales would resist either ocean acidification or some other chemical process that would demineralize the denticles. If you instead are referring to something like abrasion resistance then be clear about that and maybe choose a different term. Regardless, please be more clear here about what measurements this interpretation is related to.

Lines 234: "This concurs with our correlation analysis […]". The only correlation analysis mentioned at this point is correlations between categorical traits. Here you are talking about correlations among the performance/functional variables. These results should be in the methods and in the results – they shouldn't first pop up in the discussion.

Lines 235: Hardness and mineral content are negatively correlated? If true, this is extremely interesting as I believe we typically think of high mineral content producing harder tissue. Please elaborate more on these results and show a scatterplot graph of this data as I think this is counter to expectations and therefore either/both suspicious and also really interesting if true! It would be good to cite some literature that looked at these variables (or similar ones such as hydroxyapatite concentration vs hardness) in other mineralized structures.

Lines 273-275: I don't understand why you are talking about form/pressure drag here? You mention it and then don't really use it for anything in your writing so I think you could remove this to help be clear for readers.

Lines 278-280: The example of how stiffness in octocorals might be useful to reduce drag seems pretty irrelevant to thinking about what is going on in shark denticles. In that study they measured the material properties of pieces of the whole organism and here you are measuring material properties of parts of the small denticles that are embedded in the flexible skin in a variety of overlap patterns. I don't think there is any reason to believe that flow is changing the shape of denticles appreciably so I don't think we would assume there would be a connection between stiffness and flow in denticles, as you are saying here. Denticles are hard, small, and mineralized so why would their stiffness be related to flow and drag in the way you are suggesting? There should be no morphing of denticle shape (which is what is seen in the octocorals paper because they are soft corals – so a much softer and deformable tissue).

Line 282: Related to the ideas and comments above, you mention how bristling might require flexibility of denticles, but again I would state that I don't think there would be much actual bending of denticles during bristling. The denticles simply rotate at their base because they are attached to skin tissues which are soft and deformable…the denticles themselves are likely not bending much.

Lines 287-292: This is an interesting and surprising result! Does damage resistance to a denticle (as measured here) also equate to damage resistance to the shark? They could be related but I could also imagine that sharks might evolve denticles that are more prone to scratching so that they 1) don't break and 2) absorb force to prevent damage to the skin.

Line 284-285: This statement about denticle ridges providing more control is not a well supported statement and is not experimentally tested or supported at all. If anything, it is just a hypothesis that you are here stating as if it were fact. Please be wary of this type of overstatement and consider tempering your language here and making it obvious this has not be experimentally shown or tested.

Line 336: This statement about how ridged denticles and their "ability to optimize swimming performance" have driven the diversification of carcharhiniformes seems like a pretty large overstatement. In part because I don't think there is really good evidence that ridged denticles optimize swimming performance and ridged denticles appear in many other shark groups. Also, how are you connecting the evolution of ridged denticles to this change in diversification rate? Where is the evidence for that? I think it's an interesting point and idea but I think this statement needs to be revised a bit more.

Your Fig 2 shows that pretty much the same denticle morphotypes were present before carcharhiniformes and their increased diversification rate…so I feel like the evidence isn't consistent with saying the ridged denticles or morphotypes you described here are related to that increased diversification rate.

Line 363: Which ancient group of sharks are you referring to? Or maybe you just mean all sharks? (Also I think you can just say marine ecosystems instead of most marine ecosystems).

Line 382: You mention including shark SEM images from the literature. You need to cite what literature here! Also, give a general idea of how many species and indivuduals were sampled using the fish markets, the natural history collections, and the literature. You could even break it down by each market and natural history museum and literature piece. Also please mention somewhere how many individuals per species you sampled.

Lines 385-386: Supplementary table S4 was not attached (at least in the reviewer files) yet this is a very important table. Make sure it is included it includes all references and information.

Lines 389-396: Your SEM preparation methods of drying skin between weighted glass slides almost certainly deformed the surface by modifying the orientation and spacing of denticles. 1) You should be clear about this potential downside and mention/discuss it somewhere. And 2) You should discuss how you think it changed your results?

Line 426: You should also cite this scikit-learn package in your bibliography – not just provide the website.

Line 428: Why did you want clustering results to match clades in the phylogenetic tree? Please justify and explain this.

Line 429: divide should be divided

Lines 465-457: Again, all of these softwares and packages should be cited in your references.

Lines 460-466: You need to add two things here – 1) the names of the species that you sampled, and 2) clarity that a single species is being used for multiple morphogroups in many cases. The way this is written makes it seem like you selected 10 species and each one was a separate morphogroup, but you are actually using different regions from each species as different morphogroups. Just make sure these things are obvious in the main text so that readers have understood what you have done.

Line 514: You need to be specific about what shark skin samples were CT scanned. It also makes sense to state why you did this as well.

Lines 526: Somewhere around here (but probably also in the results) you need to at least briefly mention how you decided to arrange denticles on these 3d models. Fig 5 shows that the spacing between denticles in different models is different so make sure that is obvious here and briefly describe how you determined that. At line 531 I know you say you are "imitating the distribution of denticles" from biological samples but how specifically did you imitate? Did you use actual measurement or just visually decide that the models looked similar to the biological samples.

Line 533-534: How did you create a 3d geometry using only SEM images? You need to explain your methods either here or in the supplement and reference them here.

Lines 535: Somewhere in this section you should describe if you preserved biological denticle size from the CT scans or if you are scaling all the denticle models to the same size. This is an important distinction that allows for proper interpretation of your results.

Lines 533-539: The spadenose shark was used as the example denticle for groups 1, 3, 6, and 7 – presumably though from different regions of the body. This is never really made obvious in the paper and while it is understandable you need to be more forthcoming about this. Having a single species represent more than 40% of your 'morphogroups' is important for interpretation here – samples are potentially less independent than they seem and it's relevant that you are using samples from various regions of the body. This also needs to be clear as some of your analyses are just on the F1 region.

Lines 544 CFD methods: You need to include relevant basic details of the CFD tests here and also in the results. You need to provide information that answers the following questions: 1) What was the size at which models were produced? Did you scale denticle size at all? What was the flow speed at which models were tested? Were all models tested at the same Reynolds number? How did you decide to choose your flow speed(s)?

Line 577: I think ration should be ratio

Lines 582-584: Here you use morphogroups to divide denticles into ridged vs not but Fig 1, category 5 shows SEM of

Lines 587-588: This categorization of habitats needs to be explained more here – this is a very important part of your methods in this portion of your paper and you need to elaborate and describe how your states were categorized. It is not enough to just cite previous work.

Figure 1: for the UMAP analysis, this particular graph shows some separation between these 10 morphogroups, but there is also a large amount of overlap among many groups. I think you need to comment on this in the text when you first introduce this figure.

Fig. 1: Right now this figure has a lot of small text on it…this in theory will be okay with digital versions with zoom but it would be good to make the little label triangles in the corners of your SEM images larger and more legible.

Fig 1: For morphotype 3, the round ridge…I'm guessing that round ridge refers to the denticle crowns being more rounded in morphology? But then the example of the 3d colored model in B has very long and distinct posterior spines or tines…aka it is not rounded. What is going on here? (might want a different name for this group?)

Fig1. The morphogroup 5 aka 'thick ridge' also shows an example of a non-ridged denticle for morphogroup 5 in the SEM of Pseudoginglymostoma. Are not all denticles in this category ridged?

Fig 1. Why does the UMAP graph here look different than the UMAP graph in Fig s8? Simply because there are different groupings?

Fig. 2: I get that you ran BAMM and maybe some of the other comparative methods on this full 500+ species phylogeny but the ancestral state reconstruction shown here probably is not representative of this full phylogeny…especially given that is seems like you denticle morphotypes can often be different among close relatives. I guess what I'm saying is that do you really need this figure and should it be displaying the ancestral state reconstructon which is much more accurately (and

legibly) portrayed in Fig 3 in the pruned phylogeny. My advice would be to get rid of Fig 2 and instead include a different figure instead (like maybe s2, s3, or a modified version of s10).

Fig 3: the figure legend here need to be much larger – currently it's not legible

Fig 4: The lines in the lineages through time plot should be different colors or much different line types – the ones you have chosen are not easy to distinguish.

Fig 5: what are all of the horizontal line on each plot in part A? If these are some representation of significant differences then you need to make that obvous and you also probably need to check your stats because these all likely have way too many significant groups given the number of boxplots. Please either explain these or replace these lines with something more easily interpretable.

Fig 5: For the grad coefficients, what is the drag coefficient in your CFD setup of just a flat skin without features? Are you actually showing drag reduction compared to a flat plate or no?

Fig 6. You need an explanation somewhere of how you are getting these functional axes plotted back onto this PCA morphospace of your denticle data. In the current reading of your paper this is not obvious or discussed enough.

Fig 6. What do the red and black lines refer to in panel A? I'm guessing one must be increasing in performance and the other is decreasing? Why not simplify things and just plot the increasing performance arrows only?

Fig s3: Panel F shows a model that has no surface ridges but then the CT scan shows clear ridges. This seems like a major oversight given that much of your paper is about ridges vs non-ridged denticles.

Fig s10: For me the most interesting and potentially novel part of your paper is the functional interpretations and findings across different shark species (morphotypes). The correlation analyses are mentioned multiple times in this section and include some interesting results but aren't shown. When I went looking for them here in Fig S10 I was confused because these graphs in panels C and D do not show correlations clearly at all. Please make some graphs (scatterplots) that plot hardness vs reduced modulus and fluorine vs damage resistance and include them in a figure or supplementary figure (a figure in the main text would be better). If you want to them split them by morphotype to plot then you can color the points or make a correlation or scatterplot matrix. But really here the points should be species which are your morphotypes in this case (as I understand it, you sampled one species per morphotype for this analysis?).

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
My concerns have been addressed. I now understand the authors' view on Comment 9.
I appreciate the opportunity for reviewing this manuscript.

Reviewer #3

(Remarks to the Author)
I acknowledge the authors for the responses and revision of the manuscript. I read through other reviewers' comments and the authors' responses. The responses and revision have addressed most of my questions and concerns. Yet, I still have questions / concerns. I hope the authors can address them to improve the manuscript. Before presenting concerns / questions, I summarize the work below.

Summary

The paper analyzes dermal denticles of sharks using developed quantitative methods to better understand the impact of dermal denticles on how sharks adapt to various environments. The paper adopts nanoindentation, electron probe micro-analysis, and computational fluid dynamics to establish mechanical properties, elemental composition and swimming performance. Concretely, the paper uses deep learning to learn visual features and clustering them into a predefined number of groups. It further aligns clustering results with the clades of the phylogenetic tree of sharks. Based on these clusters of dermal, the paper conducts a series of mechanical, elemental, and fluid dynamics experiments to investigate the characteristics of these different denticle morphologies. Results lead to multiple conclusions, e.g., denticle ridges are highly variable and present more functional properties beyond drag reduction. Moreover, it suggests a radiation event within the ground sharks around 170MYA, which results in the enormous diversity of the clades where the ridged dermal denticles may be the key.

The paper is generally written well. It has certain merits and contains clear documentations and important details. As a non-expert in life sciences, I enjoy reading this paper while I can't confidently evaluate the significance in the respective discipline. So I leave this evaluation to other expert reviewers. As a research from Computer Science, I think the technical parts (e.g., self-supervised learning by VSC, clustering by K-means) make sense. Yet, there are missing details and inappropriate statements.

Concerns / questions

- In responding to my previous question why the authors choose Variational Sparse Coding (VSC) over Variational Autoencoder (VAE), the authors insert new lines as below. However, the inserted writing mentions "latent vectors" without defining what they are. I don't think readers with little machine learning or deep learning background can follow. So I suggest the authors revise this part to improve the readability.

"We prefer the VSC as it introduces a regularization term in the loss function, which often constraints the latent vectors to approximate a standard normal distribution to ensure that the learned representation does not diverge excessively in pursuit of high similarity in image reconstruction in the training process. Excessively dispersed latent vectors could lead to issues where similar data points, i.e., similar appearances of dermal denticles, do not map to nearby points in the latent space, making the analysing and interpreting of the latent space difficult or unmeaningful."

- In responding to my previous concern about whether "the clustering results actually do not align with the phylogenetic tree and hence require some manual post-processing". The authors added paragraphs in which there are two statements below. However, in my opinion, it does not need to emphasize "in an entirely objective manner" because there is a clear manual intervention as post-process.

"Following the selection of K-means as our clustering algorithm, we conducted the clustering process in an entirely objective manner, without incorporating phylogenetic information."

"we adopted the k-means clustering result of 20 clusters as our final configuration. Out of the 20 morphogroups, we excluded eight clusters that were represented by fewer than three F1 body position denticle samples to ensure a robust sample size for the subsequent ancestral state reconstruction."

- In responding to my previous question why "Why should fossil species be included in the training data along with extant ones", the authors explain and state that "We decided not to go further with the fossil groupings, as the results may have been biased by the images themselves." Does it mean that the authors do not use the clustering results to analyze fossil specimens? But the revised manuscript still states in Line 109 that "We used 2193 SEM images of skin to train the model, representing 204 individual sharks, sampled on 53 different body locations across 132 extant shark species, 2 ray species, and 24 fossil species." This is confusing to me. If clustering results on fossils are not informative, authors can either (1) re-do clustering and the experiments without fossils, or (2) sincerely discuss the limitation of the current setup. Authors can clarify further.

Reviewer #4

(Remarks to the Author)
Thank you for your detailed response to comments and your additions and changes made to this manuscript. I still think this is an impressive paper due in large part to the multiple types of data collected and the large datasets compared to what is typical for similar work on sharks. I think many things that were not clear previously have been made clear and you have generally made many improvements to this work. The multifaceted nature of this project does continue to make this paper hard to interpret and evaluate, but many of the details you have added help clarify your research for readers. However, I still feel as though there are some large issues to be address now that some of the methods and results are more clear.

I appreciate your more detailed explanation of the CFD methods that you used. I think that these methods seem biologically unrealistic for a few reasons:

A) An inflow of 30 m/s to achieve flows over the denticles that are 15-20 m/s seem biologically unreasonable. There are few (if any) sharks that can achieve these kinds of speeds and only then for a moment. Testing at these speeds does not seem biologically relevant. Please justify your choices in the manuscript or consider changing your testing protocol.
B) You tested your different models under the same flow conditions, but also used real denticles sizes based on CT scans for the models. I am sure there are body size differences between the sharks you sampled, so this is the same as saying that all your sharks are swimming at the same speed, regardless of differences in sizes. This means the biological interpretation

for these comparisons is different across different species. For example, assuming a 20 m/s flow speed from your testing, this means 20 body-lengths/second for a 1 m shark versus 5 body-lengths/second for a 4 m shark. Another way of thinking of this same issue is that when discussing swimming speeds in sharks and fishes across species, we often standardize speed by body length and use body-lengths/second. In part this is because most species have routine, cruising swimming speeds of around 1-2 body-lengths/second, and correcting speed by body-length in this way helps allow comparisons across different sizes. Your CFD test is then effectively having some species go much faster in body-length relative units compared to others, making this comparison not as relevant to biological reality. Again, please justify these choices or consider changing your protocol.
C) Lastly, I don't really see a developed boundary layer in the setup of your CFD tests and this general testing configuration doesn't seem particularly realistic – where a small patch of denticles are subject to high incoming flow with nothing upstream or downstream. The boundary layer seems exceptionally thin compared to denticle sizes (perhaps also due to the high speed flow). I grant that it may be overly difficult to get closer to biological reality, but the apparent lack of a developed boundary layer above the denticles seems different than what is likely happening on real sharks. If these physics do not match expectations for what might be happening on real shark, then how should we interpret your results?

In part d) of my first comment (your Comment and Author response 18), I asked about how you dealt with variation due to body size, which is always present in morphological datasets across species. However, in your response you told me to look at a different response above that is all about your process of k-means clustering. Perhaps this is a mistake, but I think this is a very significant issue and deserves a response as well as clarification in the manuscript. Based on your response elsewhere, it seems you have not corrected anything based on body size because you state that all of your samples are from sexually mature individuals (which is also difficult to determine and prove, but that is a separate issue). It is very uncommon to choose not to correct for size in some way in a comparative morphological dataset – I strongly suspect that variation due to body size is a large part of the morphological variation in your study simply because it typically is for most morphological characters. If you have any information on body size, you should be able to look at this relatively easily to verify or deny this prediction. It is also possible that variation due to body size is not such a large component of variability here, but regardless, some justification needs to be provided for your methods in this case and it may be necessary to see how body size is impacting your data if you have fork length or other size information. In addition, even if all the sharks were sexually mature, there is still a large range of sizes for mature sharks even within the same species (not to mention between species).

After learning more about your decisions on how to choose ten morphogroups and how your deep learning approach created and defines these morphogroups, a few things seem like they are still potentially issues. First, your descriptions for how you decided on 10 clusters/morphogroups is very helpful but none of it sounds particularly 'objective' despite you referring to it as such. I'm not an expert on these methods but my reading of this description generally gives me the impression that you did some filtering based on clusters that were small and then some that were close together (all of which has an arbitrary or subjective feel to it). Then you also looked at the results on the phylogeny and decided that ten was just about the same as twelve groups (again, kind of arbitrary feeling). I think it is great to explain all of this to readers but at the very least I would not describe this as an objective process.

Second, I deeply appreciate you posting your code in case someone wants to recreate your analyses, but it seems to me that these morphogroups you define may be nearly impossible for anyone else to use without data outside of your own. If another author wanted to also characterize their own data on some different species or body regions using your morphogroup framework, it seems like they would need to fully recreate your analyses (including all your data) and then assign their new data to one of above morphogroups. I tend to think this is asking a lot and it seems unlikely other authors will be able to utilize your contributions in this way. The current denticle literature very often uses morphotypes (same idea as morphogroups) based on work by Reif and I believe these morphotypes have been so pervasive because he provided clear descriptions for determining denticle morphotype and connected these types to function using logical arguments. If other researchers have difficulty in making use of the morphogroups you have defined here, it seems unlikely that your morphogroups will be able to supplant the status quo.

I also believe that the methods you use to categorize morphogroups are very interesting, but that it is unfortunate you are unable to understand what features are most important to define and separate morphogroups and that there is substantial morphological variation in some morphogroups. The addition of Figure S3 is excellent and really demonstrates that some morphogroups have some shared features but in many of these groups it is very difficult to describe shared features of all the provided images. In many cases it therefore seems unpredictable why a particular image was assigned to one morphogroup instead of another. For me these issues lower my confidence in these results.

The variation within some of the morphogroups makes it difficult to have confidence that using a few representatives from a morphogroup is a good model for all the species with a morphogroup. This makes me more cautious in assuming your results on various denticle functions will be generalizable for a given morphogroup.

Version 2:

Reviewer comments:

Reviewer #4

(Remarks to the Author)

I appreciate the lengthy responses to my previous comments and the efforts that the authors have made here. I still continue to think that this is by far the most ambitious treatment of shark denticles to date and it is very cool that the authors combine comparative data with multiple functional measurements.

Despite a number of different comments and multiple rounds of revisions, you continue to describe your methods as 'objective' and you also describe experts' opinions on selecting features to quantify as 'subjective'…and in doing both these things you denote that subjectivity is a bad thing and objectivity is a good thing. I totally understand what you mean with these statements, but it is frustrating to see you still use this language despite multiple comments. I think you risk immediately offending anyone who studies and measures morphology. I also think that your responses in the 1st review also revealed that not all of the methods you use for creating 10 morphogroups of denticles are wholly 'objective'. I think you could deliver the same message you have here but without some of these terms. I also think that by using these terms you are opening a semantic debate about the 'objectivity of science' (many folks would argue that science is never really objective), and a debate about if subjectivity is an inherently bad thing (again, I think many people would argue for or against it). These are interesting topics, but not really the focus of your paper.

Somewhere – probably around lines 236-238 – please list the total samples (specimens) that were measured for each of the functional testing methods (nanoindentation, mineral protection, CFD). You mention "2-4 species from multiple body locations within each morphogroup" for nanoindentation (and I think also mineral protection?) and then "1-2 samples per morphogroup" for the CFD, but you should also just provide actual numbers. 1-2 samples per morphogroup could be 11 samples or it could be 19…. Similarly, I don't actually know what you mean by "2-4 species from multiple body locations within each morphogroup" means because the language isn't clear. Perhaps this means something between 21-39 samples or perhaps more if each species was sampled at multiple body locations. Both the total number of unique species+locations should be provided, and the range of individuals per species should be given – basically, it would be great to briefly give readers an understanding of what your sample size is here.

These ranges of numbers you're providing also don't line up with the samples sizes given in the boxplots in Figure 5. This must mean multiple measurements were made on each sample (I think you seem to indicate 10-30 per sample in the methods?). If so, you should not be treating each individual measurement as a statistically independent sample – your methods mention using an ANOVA, but your data seems to have a nested structure with multiple measurements per denticle and per individual. You should be accounting for this data structure in your tests of statistical significant differences in functon(nanoindentation, mineral protection, CFD) among the different morphogroups. There is the additional complexity here of sampling multiple distinct regions of the same individual and the fact that species are related and thus no-independent, but without even worry about those you should probably be at least using a nested ANOVA or an ANOVA with some covariates.

I appreciate your attention to looking at the effects of body size on your data. I think I understand some of what you have done and although it isn't ideal to have found differences in body size across your morphogroups, you seem to suggest that it has little effect. However, I'm not sure your results really support that idea. As I'm understanding it, your models have only body length and sex in them? And just with two parameters you gained some improvement over the null model (I don't quite understand what the null model is here)? A classification accuracy of over 50% seems impressive to me with only body size and sex as predictors, but I suppose it is all relative. Is there anything more in this section that you can say to give readers a sense that body length is a minor effect here? It isn't surprising that it might have a minor effect on such a high-dimensional dataset, but the fact that it does seem to separate groups is interesting but also could be an issue.

I also should have been a bit more prescriptive in my previous review. I would really like to see data on how body size relates (correlates) to the various UMAP axes. Significance values and correlations (r or R-squared).

Reviewer #5

(Remarks to the Author)
The paper employed the VSC model and K-means clustering algorithm to quantify and group the geometry features of shark denticles by analyzing SEM images. Combing with nanoindentation, electron probe micro analysis and CFD simulations, the authors tried to establish the form-function relationship of the dermal denticles. Also, the tempo and model of shark denticle evolution has been studied through phylogenetic tree analysis. The topic is interesting, the analysis is comprehensive, and the paper is well written. However, I have a few questions for the authored to be addressed.
General concerns:
1. The authors already got the shark denticle 3D models using SEM methods. But in the CFD part, the authors used micro 3D-scan to obtain the 3D models. Could the authors explain why they didn't directly use the 3D models from SEM and the differences between these two methods?
2. Although the external morphology features determine the functions of dermal denticles, the hydrodynamic function also depends on the fluid environment, the locomotion gait of sharks and the arrangement of denticles. With the same flow conditions and same setup, the conclusions about the hydrodynamics of different groups denticles may not be very solid.
Concerns in Supplementary Information:
3. The dimensions of shark denticles in simulations? It seems the authors enlarged the denticles, which is okay if the authors can keep the dimensional parameters, such as the Reynolds number, are the same. However, the local Reynolds number (calculated based on the model length) the authors used is around 2*10^4, which is much larger than that for shark

swimming. The local Re of shark denticles changed around 1,000[1].
4. Could the authors explain why you set the outflow boundary condition as zero pressure? Have you tried other outflow boundary conditions? I am not sure whether there is not enough space between the denticles and the outflow boundary. Also, for the bottom where the denticles are placed, should it be set up as no-slip boundary condition to mimic the shark body?
5. The temperature where sharks live varied widely, which greatly affects the ocean water viscosity and is not negligible.
6. The surface drag Fd should both include viscous drag and pressure drag. At least at Re=10^4, the pressure drag is not negligible.
7. Typo: Line 171 in Supplementary Information, “element”.
8. Line 216: “L = 3”. What is the unit?
9. Supplementary Fig. 15: the discrepancy between the simulation and experiment of flow past sphere is more than 30% at Re=10^4, which is not a slight difference for a validation case. If the authors included the pressure force, the discrepancy might disappear. The pressure force is more sensitive to the morphology changes.
10. For the convergence study, I don’t think the size differences between the coarse, medium and fine meshes are not large enough, which might affect the choose of mesh size. Also, please provide the quantitative comparison of three meshes instead of flow plots which can’t tell whether the simulation is converged or not.

[1] Graybill, M. T., & Xu, N. W. (2024). Experimental Studies of Bioinspired Shark Denticles for Drag Reduction. Integrative and Comparative Biology, 64(3), 742-752.

Version 3:

Reviewer comments:

Reviewer #4

(Remarks to the Author)
I think the authors have done a nice job of responding-to and addressing my comments. And of course I continue to think that this is the most detailed treatment of shark denticle morphology and potential functions to date. I think this paper will be very popular for students and other scientists interested in this and similar topics.

I have only two last comments and they are again about clarifying sample sizes. Supplementary table 2 is really helpful, as are the additions you have made in your text and figure captions. Thanks for those additions. I suspect that your samples for NI and EPMA all come from one individual shark per species. I am also unsure about how many repeated measurements are made on a single denticle. I have two requests:

1) Can you confirm that all your sharks are represented by one individual per species and state that clearly somewhere in your methods (perhaps when you introduce the sampling of denticles for NI and EPMA). This should be clearly stated.
2) When you make NI and EPMA measurements on sections, how many different denticles do your measurements come from? I’m guessing it might be one denticle for each method but I’m not sure. This should be clarified multiple places (Supplemental Table 2, Lines 245-270, Material and methods lines 691+). For example, there are 32 NI datapoints for Scoliodon laticaudus F1. How many of these are from one denticle? Or how many data points are collected for each denticle if more than one denticle is sampled? This should be clearly stated in some way…again likely in multiple parts of your paper.

Thanks for all your patience here and with my reviews. I think your paper has continued to improve in quality and clarity, and I think you have done an excellent job here.

Reviewer #5

(Remarks to the Author)

Reviewer comments

Reviewer #1 (Remarks to the Author):

The manuscript “Diverse Forms of Denticles Facilitate the Evolutionary Success of Sharks” is a study on structure-property-function relationships in the dermal denticles of sharks, and implications on their evolution. The study includes a massive study on the morphology of the denticles of 158 species based on SEM images and a very clever automated feature recognition approach based on deep learning to categorize the denticles into 10 morphogroups. The study also includes mineral and fluoride content, indentation hardness and modulus, as well as hydrodynamic simulations to estimate drag on 20-30 samples from each of these 10 morphogroups. The authors draw conclusions on structure-property-function, for examples the morphology of some groups is more suited for drag reduction in smaller, faster shark species, while the morphology and mechanical properties of other groups is more suited for protection in slower, large sharks. The authors also draw some conclusions on the evolution of sharks, for example by tying a rapid evolution in denticle with the emergence of the Atlantic Ocean.

Comment 1

The design of this study is extremely strong (although spacing between denticles could also have been considered), the amount of data is massive and as far as I can tell the analysis is solid.

Author response 1

We really appreciate the recognition given for our collected dataset and analyses. Thank you for noticing the spacing between denticles as a morphological characteristic. We did consider

spacing, but we certainly agree to clarify the mentioning of it in the text. See below, **corrections are in bold**.

Lines 140-144: “This VSC model extracted sparse and essential visual features of dermal denticles from the standardized SEM images without human labelling, capable of accounting for denticle density, **i.e., the spacing between denticles**, overlap, size, and external morphology, and represented the input images with feature vectors in a 512-dimensional morpho-space.”

Comment 2

My only two concerns are as follows: I feel like this study essentially confirms what is already known about the morphology-function relationships in dermal denticles, and about specialization in shape in relation to ecological niche and lifestyle in different species. While it is important to confirm this knowledge with a strong, comprehensive study, the data and analysis do not provide knowledge that is particularly new. Specialization and differentiations of morphology under ecological pressure has been known since Darwin, and the results reported here in terms of the evolutionary patterns in dermal denticles seem somewhat in line with what should be expected in this context. Is there any oddity, new evolutionary patterns, or an unexpected function of the dermal denticles revealed in this study? Again the results are strong, but I am not sure that the degree of novelty is up to what should be expected in this journal.

Author response 2

We are thankful for the reviewer’s recognition of our work establishing the morphology-function relationships of dermal denticles, as it was one of our main goals. Regarding the

novelty of the study, we understand that this can be highlighted much better and have added a section describing this in the beginning of the manuscript.

Lines 66-71 […] 83-98: "**The novelty of this study relies on the methodology used to examine and quantify the morphology-function relationship of dermal denticles. We establish quantifiable parameters describing each denticle morphogroup, and subject these to analyses on a macroevolutionary scale. For the first time,** we use a deep learning approach of artificial intelligence to group, quantify and classify the images of shark dermal denticles. **This study objectively defines ten dermal denticle morphogroups across 132 extant shark species.** […] In addition, we designed a range of mechanical, elemental, and fluid dynamic experiments and simulations that provided insights into the mechanisms behind abrasion strength, defence, and drag reduction functions (Fig. 5). **Here, we found that the internal properties add to the overall functional capacity of dermal denticles. Among the carpet sharks, which are primarily demersal and reef-associated sharks, we found dermal denticles with low damage resistance - a result disagreeing with previous hypotheses suggesting that reef-associated and benthic sharks carry denticles with high damage resistance to prevent abrasions.** We performed a functional trait analysis to provide an integrated understanding of the possible functional trade-offs in different types of denticles (Fig. 6). **For instance, denticles with low hardness have a higher mineral protection, in terms of a higher fluorine content, to potentially compensate for damage repair when broken. These results show the complexity of the functional properties of dermal denticles.** Finally, we conducted a correlation analysis of trait and habitat evolution to understand whether a relationship exists between ecological niche and denticle morphogroup (Supplementary Fig. S4A-C). **Here we can conclude a strong correlation between denticle type, ridged vs. non-ridged, and the ecological niches of the sharks.**"

Comment 3

My second concern is the way the manuscript reads: the manuscript is very well prepared, but at several places the level of details made me feel like a more specialized journal would be a better fit for this report.

Author response 3

We understand that some parts of the manuscripts have a higher level of detail, however, this was a decision based on the complex nature of the content. We felt the need for expansive writing on some parts of the manuscript, as a simplified version could result in misunderstandings or lack of comprehension from the reader's side. However, we are certainly willing to rewrite these parts if they are pointed out to us, if that becomes relevant.

Reviewer #2 (Remarks to the Author):

In the manuscript entitled "Diverse Forms of Denticles Facilitate the Evolutionary Success of Sharks", authors propose a new method to analyze the morphological features of dermal denticles based on deep learning. Utilizing the new method, combined with multiple analyses, they conclude that the diversity of denticles increased at around 170 million years ago, which may have enabled sharks to adapt various environments.

Traditionally, morphological analyses on various forms were conducted by expert human observers, which may tend to depend on personal experiences. The new method proposed in this study enables to conduct objective evaluation, which has a potential to make progress in wide fields of earth science related to observation, such as microfossils, mineral grains and so on. In addition, the conclusion from the proposed observation method is interesting, although I'm not the expert in this field.

Given the wide applicability of the proposed methods and the impact of the conclusions, I would recommend that the manuscript be accepted to Nature Communications after a few questions are cleared up.

Comment 4

Major comments:

[L. 93~] K-means clustering analysis (KCA) was conducted with number of clusters set at 20, and after checking results, class number was manually reduced to 10. Although I understand such a process is practically useful, it reduces the "objectivity" of the results. It would be preferable to try with a smaller number of clusters (K=10 for example) and describe whether the classification is valid or not.

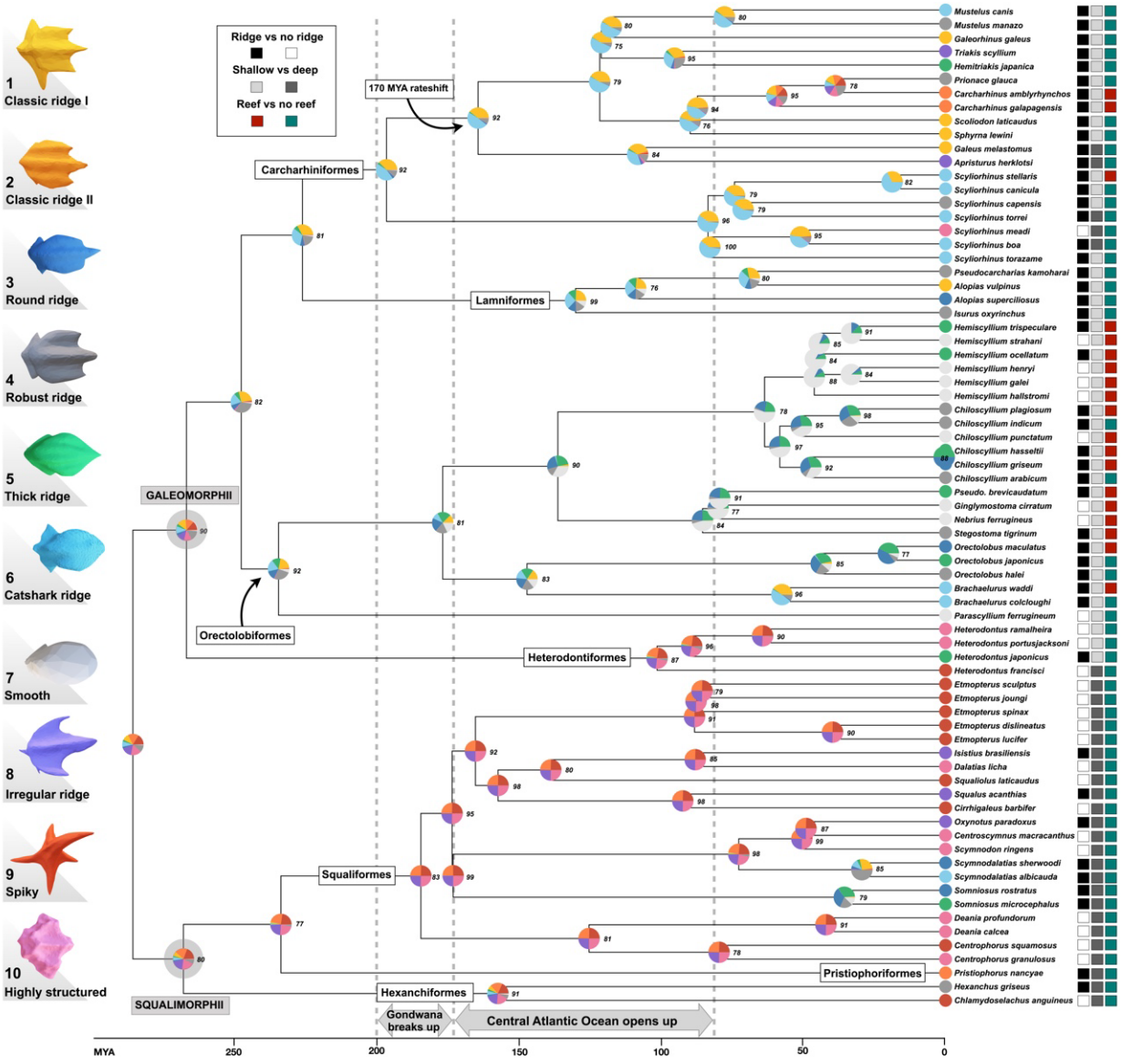
Author response 4

Thank you for the comment. We understand the need to outline more clearly why K=20 was optimal. **See improvement to the Methods section below, highlighted in bold.**

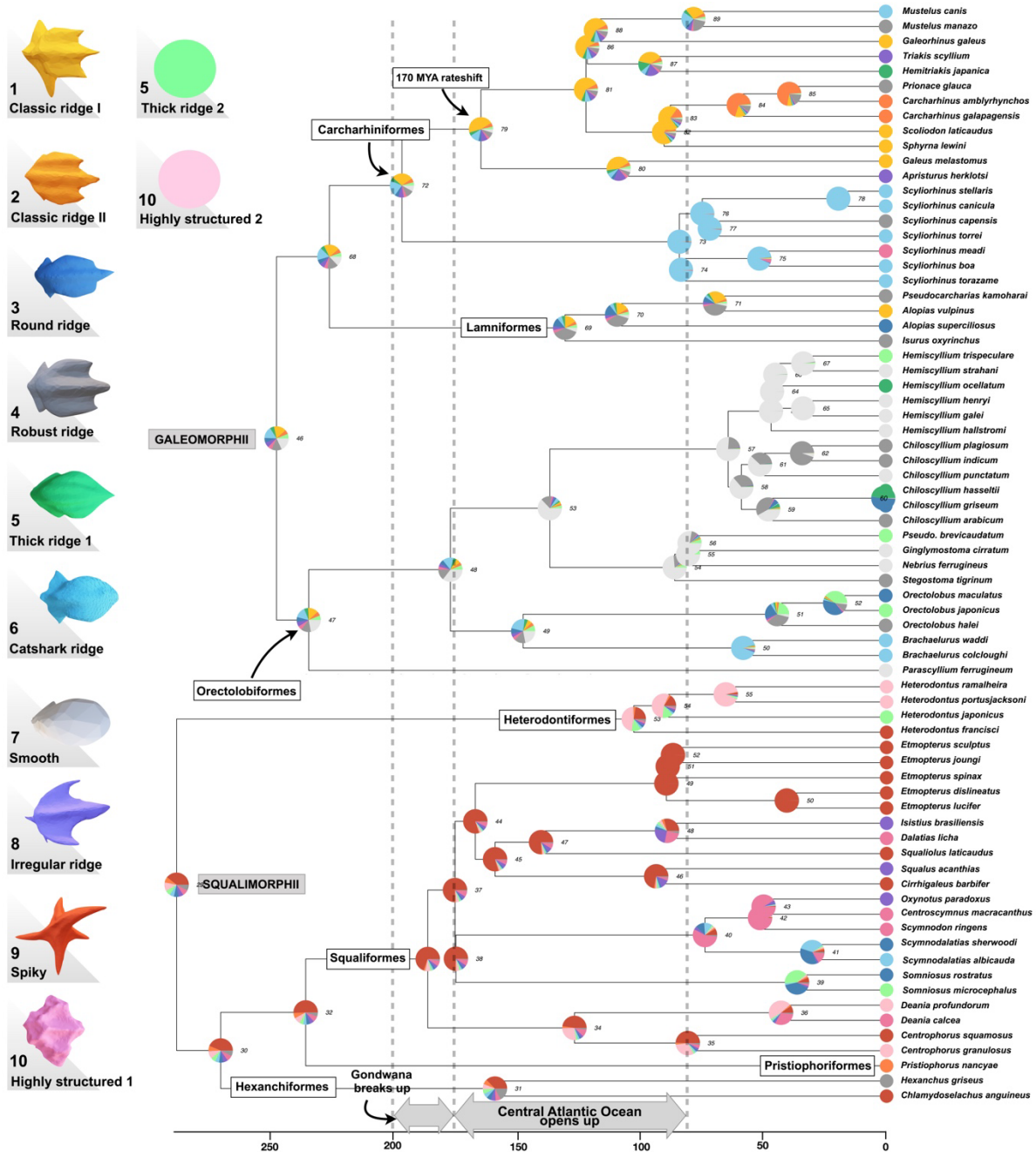
Additionally, to settle the reviewer's uncertainty, we ran the ancestral state reconstruction using K=12, meaning we did not merge any groups (the new Supplementary Fig. S2). The results are similar, yet with lower bootstrap values compared with the original Fig. 3. As the K=12 tree was made in two parts, split into Galeomorphii and Squalimorphii, it cannot account for the fact that these two subclasses share a common ancestor. The K=12 tree is thus less representative regarding the ancestral states of the morphogroups. In the end, the two trees seen in Fig. 3 and Supplementary Fig. S2 show the same trends, and we thus accept the K=10 morphogroups for the sake of simplicity and deeper ancestral resolution.

Methods lines 557-570: "**Following the selection of K-means as our clustering algorithm, we conducted the clustering process in an entirely objective manner, without incorporating phylogenetic information. First, utilizing the K-means function from scikit-learn, we partitioned the dataset into 20 and 50 clusters. These clusters corresponded with the taxonomic groups delineated by the phylogenetic tree. Second, due to the scarcity of representative samples for F1 body positions, we adopted the k-means clustering result of 20 clusters as our final configuration. Out of the 20 morphogroups, we excluded eight clusters that were represented by fewer than three F1 body position denticle samples to ensure a robust sample size for the subsequent ancestral state reconstruction. Third, from the remaining 12 morphogroups, we further reduced the number to 10 by merging clusters with the lowest pairwise centroid distances, using a dimensionless threshold of 1.6. This step, which only merged the closest groups, was designed to simplify further analyses and did not alter the grouping patterns mapped on the phylogenetic tree. Additionally, merging these morphogroups did not impact the results of the ancestral state reconstruction. (Fig. 1A, Supplementary Fig. S2 and S10).**"

Original ancestral state results using K=10, where "thick ridge" and "highly structured" consists of two merged groups, respectively, based on their close proximity in the trait space (Fig. 3):



Ancestral state reconstruction with unmerged morphogroups (K=12):



Supplementary Fig. 2. Ancestral state reconstruction of K=12 morphogroups. Thick ridge (grp 5) and highly structured (grp 10) are treated as the original, unmerged groups concluded by the K-means clustering method (see Methods section for details). Original green colour (seen in Fig. 3) refers to “thick ridge 1” and light green colour refers to “thick ridge 2”, and original pink colour refers to “highly structured 1” and light pink colour refers to “highly structured 2”, depicting which species belong to either group. Unmerging the two morphogroups does not impact the ASR results as presented in Fig. 3.

Comment 5

[L.140~] I can't understand how authors determined the timing (geologic age) of evolution. Did they analyse the literature collections of denticles with known age?
Since the dating information significantly affects the conclusions in this manuscript, a more careful description would be desirable.

Author response 5

To avoid confusion on the topic of the geological timeline used in our study, we have added a sentence clarifying this early on in the mentioned section.

Lines 148-153: "To understand the tempo and mode of the evolution of shark dermal denticles in macroscopic terms, we mapped and analysed the morphology using a recently published phylogenetic tree of shark relatives[11]. This **fossil-calibrated** phylogeny represents the most taxonomically comprehensive to date, however, we acknowledge that other phylogenies may present alternative hypotheses. **The geological dating of fossil specimens and thus the geological timeline referred to in this section is based on the phylogeny by Stein et al. (2018).**"

11) Stein, R. W. *et al.* Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nat Ecol Evol* **2**, 288-298, doi:10.1038/s41559-017-0448-4 (2018).

Comment 6

Minor comments and checks:

[L. 101] What does "consistent" mean in this context? Is it possible to visualize or describe

the “consistency”?

Author response 6

Thank you for needing clarification on this. In this context “consistent” means that the grouping pattern of the different K sized groups match the groups already formed in the phylogenetic tree of sharks. We have clarified this in the text, see below.

Lines 134-135: “We compared the grouping patterns of different K with the **taxonomically defined** clades of a phylogenetic tree of sharks[11], and found **that these groupings were similar**.”

Comment 7

[L. 419] Should the full spelling of the PCA be listed here?

Author response 7

Correct. We have inserted the full spelling of PCA on Line 514. Thank you pointing this out.

Comment 8

[Ref. 11] Duplicated hyphen (“690--700”).

Author response 8

We have corrected the mistake of duplicated hyphen.

Comment 9

[Ref. 58] Since this manuscript is published in a peer-reviewed journal, citation from arXiv

should be avoided unless there are special circumstances. Also, the spelling of the first author should be “He, K.”.

https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Author response 9

We agree that this needs clarification. In the computer science discipline, the conference papers are as important (if not more) as peer-reviewed papers because they develop very fast. In recent years, open accessible papers and results are mostly required to be reproducible by third parties. We have corrected the spelling of the first author as suggested. Further, this conference paper is a milestone marking the first time in history that machines surpassed human performance in image recognition. To date, it has been cited over 193,280 times by AI research and application domains. Thus, we prefer to cite the paper as in our original submission.

Reviewer #3 (Remarks to the Author):

The paper analyzes dermal denticles of sharks using developed quantitative methods to better understand the impact of dermal denticles on how sharks adapt to various environments. The paper adopts nanoindentation, electron probe micro-analysis, and computational fluid dynamics to establish mechanical properties, elemental composition and swimming performance. Concretely, the paper uses deep learning to learn visual features and clustering them into a predefined number of groups. It further aligns clustering results with the clades of the phylogenetic tree of sharks. Based on these clusters of dermal, the paper conducts a series of mechanical, elemental, and fluid dynamics experiments to investigate the characteristics of these different denticle morphologies. Results lead to multiple conclusions, e.g., denticle

ridges are highly variable and present more functional properties beyond drag reduction. Moreover, it suggests a radiation event within the ground sharks around 170MYA, which results in the enormous diversity of the clades where the ridged dermal denticles may be the key.

The paper is generally written well. It has certain merits and contains clear documentations and important details. As a non-expert in life sciences, I enjoy reading this paper. My comments are as below.

Comment 10

- The paper automatically creates some morphotypes of shark dermal denticles by using deep learning (i.e., the variational sparse coding) and clustering algorithms, aligning the clusters with a predefined phylogenetic tree, and analyze these clusters in various aspects such as elemental composition and fluid dynamics. From the writing, my understanding here is that the paper makes an assumption that the dermal denticle clusters are correlated with phylogeny, and analyzing them can reveal evolution events. However, if a goal is to understand evolution through analyzing shark dermal denticles, is it a better clustering strategy to sample and cluster denticles within predefined groups of shark species and do the analysis? Can authors discuss the above?

Author response 10

We apologize for not making this part clear. Our approach to sampling was to stay objective and observe how denticle morphologies were grouped together, without any predefined phylogenetic grouping. See changes to the text below.

We refer to **Author Response 14**, which clarifies this in much detail.

Lines 104-110: “We wanted to objectively categorize dermal denticles based on key morphological features, and for this we used the Variational Sparse Coding model (VSC), a branch of deep learning autoencoders to quantify the visual features of the shark's dermal denticles[12,13]. **Prior to morphogroup analysis, we carried out broad taxonomic sampling in order to represent the most morphological variation of dermal denticles in the dataset.** We used 2193 SEM images of skin to train the model, representing 204 individual sharks, sampled on 53 different body locations across 132 extant shark species, 2 ray species, and 24 fossil species.”

Comment 11

- Line71. Please cite papers to give reference for "a recently published phylogenetic tree".

Author response 11

We have added a reference to support "a recently published phylogenetic tree".

11) Stein, R. W. *et al.* Global priorities for conserving the evolutionary history of sharks, rays and chimaeras. *Nat Ecol Evol* **2**, 288-298, doi:10.1038/s41559-017-0448-4 (2018).

Comment 12

- The paper uses Variational Sparse Coding (VSC) to learn visual features. The paper states in Line400 that VSC is an improved version of Variational AutoEncoder (VAE). Yet, VSC is less recognized in the literature. Can authors clarify why VSC is more preferred in the study than VAE and other well-established methods? Moreover, in Line417, can authors explain

the regularization of features to the spike-and-slab distribution? How does this regularization help learning the visual features in the study?

Author response 12

Thank you for need for clarification, we will be happy to explain. VSC is basically a simple modification of VAE, and VAEs performance has been proved in the original paper and by third parties (the open reviews and the reproducibility validation can be found here: https://openreview.net/forum?id=SkeJ6iR9Km). Since the VSC is an upgrade of standard VAE and easy to implement, we chose to use it in our research. **See inserted explanations in the text below.**

Methods lines 485-494 […] 510-513: "We use the Variational Sparse Coding (VSC) autoencoder, which is an improved version of the Variational Autoencoder (VAE)[12,13,64]. The VSC is a self-supervised model, and it can extract visual features of images by reconstructing the inputs with constricted conditions and revising the model without predetermined labelling. **We prefer the VSC as it introduces a regularization term in the loss function, which often constraints the latent vectors to approximate a standard normal distribution to ensure that the learned representation does not diverge excessively in pursuit of high similarity in image reconstruction in the training process. Excessively dispersed latent vectors could lead to issues where similar data points, i.e., similar appearances of dermal denticles, do not map to nearby points in the latent space, making the analysing and interpreting of the latent space difficult or unmeaningful.** […] The feature vectors are regularized in the training process to the spike-and-slab distribution for the model, **which increases the latent space sparsity, i.e., the model extracts less numeric features for representing the whole visual feature, and provides better**

disentanglement performance where each extracted feature is more independent."

Comment 13

- Line412. It is imprecise to write "the decoder, which is symmetrical to the encoder". A more precise statement is that encoder and decoder are symmetrical in terms of network architectures.

Author response 13

We have fully complied with this comment and altered the text to the following:

Lines 504-506: "The decoder, which is symmetrical to the encoder **in terms of network architecture**, takes the 512-dimension feature vectors as inputs, and learns to decode them back to reconstruct the 256x256 images (output images)."

Comment 14

- Line428 writes that "we wanted our clustering results to match with the clades of the phylogenetic tree of sharks"; Line446-456 writes how to manually post-process the clustering results, presumably for aligning the results to the clades of the phylogenetic tree of sharks. Does it mean that the clustering results actually do not align with the phylogenetic tree and hence require some manual post-processing? Can authors clarify and discuss? What are the assumptions made here?

Author response 14

We have revised the manuscript to provide a clearer explanation of our rationale for selecting the K-means algorithm and the steps taken to objectively cluster the dermal denticle images

into morphogroups. We emphasised that the choice of clustering method was based on two main principles: 1) natural clustering, which favours algorithms that follow naturally occurring patterns such as clades in a phylogenetic tree, and 2) limited variance within each group to control for visual differences in dermal denticles within a group. We tested four different clustering methods from two categories: distance-based algorithms (BIRCH and K-means) and density-based algorithms (DBSCAN and OPTICS). In the end, K-means was chosen as it most closely adhered to the two principles mentioned above. Furthermore, we clarified that the K-means clustering was performed objectively, without reference to the phylogenetic information. The specific steps taken to obtain the final 10 morphogroups are detailed in the revised manuscript. The updated explanation of our clustering method and the steps taken to obtain the morphogroups is as follows:

Lines 518-572: "**Morphogroup formation: We tested different combinations of clustering algorithms and group sizes to divide dermal denticle images into different morphogroups and evaluate their performance. We considered two main principles to evaluate the clustering results of different algorithms. The first is natural clustering, which is a preference for a clustering algorithm that better aligns naturally occurring patterns, such as the clades in a phylogenetic tree. The second is a limited number of variants in each clade, thus controlling for visual differences in dermal denticles within a clade.**

Based on these two principles, we tested four different clustering methods from two categories: distance-based algorithms (BIRCH and K-means) and density-based algorithms (DBSCAN and OPTICS). All these algorithms are well documented and implemented in the scikit-learn package of Python[66] (https://scikit-

learn.org/stable/modules/classes.html#module-sklearn.cluster61). The first method we tested was a distance-based algorithm, BIRCH, combined with hierarchical clustering using Birch and AgglomerativeClustering from the scikit-learn package. Birch automatically divided our dataset into 500 seed groups, which were then merged into 15, 20, 25 and 30 groups by AgglomerativeClustering. The group numbers were the only parameters assigned. The second method we tested was K-means using KMeans from scikit-learn, which is also a distance-based algorithm that tends to divide the data points into groups with small variates, which in the case of dermal denticles refers to controlling the visual differences within each group. The controlled group differences make the clustering results easier to interpret. The only parameter assigned is the number of groups K. We tested 20 and 50 groups and found that these also corresponded to the taxonomic groups of the phylogenetic tree. The third and fourth methods we tested were the density-based clustering algorithms DBSCAN and OPTICS using scikit-learn. These methods do not require the assignment of a number of groups. However, they do require information to determine the parameters related to the density and connectivity of the points to form meaningful groups. We used the default parameters provided by scikit-learn. In the case of density-based algorithms, they produce very unbalanced groups, i.e. one large group with a variety of visual features, many small ones, and an addition of groupless outliers. The unbalanced grouping results look random on the phylogenetic tree and can be difficult to explain, e.g., members of a very large group were scattered all over the tree, while smaller groups matched only a few species.

Ultimately, we chose K-means because it most closely followed the two principles of natural clustering and controlled for visual differences within groups. K-means

produced more visually controlled clustering results, had a simpler implementation for reproducing results, and its result was more consistent with natural clustering, with the alignments being consistent between different numbers of groups.

Following the selection of K-means as our clustering algorithm, we conducted the clustering process in an entirely objective manner, without incorporating phylogenetic information. First, utilizing the K-means function from scikit-learn, we partitioned the dataset into 20 and 50 clusters. These clusters corresponded with the taxonomic groups delineated by the phylogenetic tree. Second, due to the scarcity of representative samples for F1 body positions, we adopted the k-means clustering result of 20 clusters as our final configuration. Out of the 20 morphogroups, we excluded eight clusters that were represented by fewer than three F1 body position denticle samples to ensure a robust sample size for the subsequent ancestral state reconstruction. Third, from the remaining 12 morphogroups, we further reduced the number to 10 by merging clusters with the lowest pairwise centroid distances, using a dimensionless threshold of 1.6. This step, which only merged the closest groups, was designed to simplify further analyses and did not alter the grouping patterns mapped on the phylogenetic tree. Additionally, merging these morphogroups did not impact the results of the ancestral state reconstruction. (Fig. 1A, Supplementary Fig. S2 and S10). All data operations were executed using Python 3.6, with the following libraries: NumPy 1.19.5[67], Pandas 1.0.5[68,69], SciPy 1.4.1[70], Scikit-learn 0.22.2.post1[66] and UMAP 0.4.6[71].”

Comment 15

- How many shark individuals are used in the study? I can see 2193 SEM images of skin, 53

body locations across 132 extant shark species, 2 ray species and 24 fossil species. Do different individual sharks have distinct denticle patterns? Can authors clarify?

Author response 15

Thank you for pointing this out, **we have clarified these changes in the text below**.

Lines 108-114: "We used 2193 SEM images of skin to train the model, **representing 204 individual sharks**, sampled on 53 different body locations across 132 extant shark species, 2 ray species, and 24 fossil species. Each SEM image represents a single body location from a single species and should be regarded as a unique entity, which means that skin samples across the body of the same species may fall into several morphogroups. **We only include sexually mature specimens, as denticle morphology varies with ontogeny**[14]"

Comment 16

- Line85. Why should fossil species be included in the training data along with extant ones? Does this assume fossil shark species share the same/similar dermal denticle morphotype? What if training on extant/modern specimens and see if the trained model generalizes to fossil? Can authors clarify?

Author response 16

The initial assumption was that fossil denticles would group among modern denticles, and we included them to see how they would distribute. However, in most of the clustering runs the fossil denticles grouped with themselves, in spite of having very different morphologies. The fossil denticles are often presented as singles on a black/white background in low resolution quality (lack of detail), whereas the modern denticle images show a complete skin sample

with rows of denticles in high resolution. The amount of information in the two examples are very different, and thus affects how the fossil denticles are grouped. **We decided not to go further with the fossil groupings, as the results may have been biased by the images themselves.**

Comment 17

- Line94. The paper excludes any groups represented by <3 F1-body location skin samples for the purpose of simplifying the representations and for using this particular location to compare denticle morphology of different species. I would like to see more discussions here. For example, does it assume (or is it a fact) that dermal denticles on the whole shark body share similar patterns with F1-body locations? Why is F1-body location so special (e.g., because they are abundant in specimens, or easy to obtain)? Why do some clusters have few examples of F1-body locations -- is it because dermal denticles from other body locations quite different from those at the F1-body location?

Author response 17

Thank you for pointing out this need for clarification. **We have elaborated on the use of the F1-body location in the text, see below.**

Lines 153-164: “It has been well established that variation in denticle morphology across a single individual can be enormous[15,16]. **Dermal denticles on different body regions, i.e., the snout, the flank region above the pectoral fin (F1 location), the caudal fin tip, etc., can vary as much as denticles between species. In this study, we observed that the denticles of the F1 region were the most variable and thus a good indicator for interspecific variation. Additionally,** the flank region above the pectoral fins (F1-body location) is the

most commonly sampled in literature for understanding the relationship between dermal denticles and shark lifestyle[17-21] (see Fig. 1B and Supplementary Fig. S1 for position of the F1-body location). **For many studies, it may have been the sheer proportional size of the F1 body location that made it ideal for sampling skin, in which a larger body area with similar denticle morphology may indicate a representative morphology for the sharks' lifestyle."**

Reviewer #4 (Remarks to the Author):

This paper represents a very large effort to quantify shark denticle morphology across 70+ species using deep learning computer vision, measure material properties using nanoindentation of the enameloid layer across multiple body regions of 10 species, measure the drag properties of denticle models of 10 denticle types using multiple regions of the body from 8 species using CFD, integrate functional variables together in morphospace, and run some phylogenetic comparative analyses including trait correlations, BAMM, and ancestral state reconstruction.

I think this manuscript represents a really big and collaborative effort and includes valuable and novel datasets that allow for really interesting progress to be made on understanding denticle morphology, denticle function and performance, and denticle evolution. However, I think the expansive nature of the manuscript makes it very difficult to digest and makes it very unclear for readers – in part due to the lack of explanation of methods and results in many instances, but also because of the many different questions that this manuscript is answering without really clear organization, framing, or justification in many cases.

I think this manuscript could potentially be a great value to the field, but there is enough

information lacking that over the course of my review I became increasingly skeptical of both the results, methods, and ultimately therefore some of the conclusions presented here. I have detailed a number of suggestions below but I would also outline a few larger comments here.

Comment 18

1) Methods and results need to be explained in more detail. For example, I had a number of questions while reading the manuscript, many of which are given further detail in the line-numbered comments below.

a) How did you setup your CFD analyses in terms of speed, model size, and Reynolds number? Are you testing across different Reynolds numbers or holding it the same?

b) How did you calculate correlations among your functional measurements (e.g. fluorine content, hardness, drag reduction, stiffness) and where are those data and results? Fig 6?

c) What are the definitions of hardness and stiffness that you are using here? Specifics are needed.

d) How did you decide to deal with variation in your morphological dataset that is due to body size? I'm guessing you didn't correct or account for body size at all so you need to justify that choice and probably also show how much of the variation in your morphological dataset is due to size.

e) Similarly, did you look at the relationship between body size and any of your functional variables?

f) How did you map the functional traits back to your morphospace using functional trait analysis? Also, what is functional trait analysis, how was it employed, and where are its results aside from Fig 6a?

g) Why did you choose 10 morphogroups? The current reading of your manuscript makes this feel arbitrary after you make the point that previous morphological categorizations are based

on arbitrary expert opinions so it feels like you are doing a similar thing here. I think more justification would really help.

h) Why is morphogroup 5 called thick ridge but one of the examples in Fig 1 doesn't really have ridges?

Author response 18

a)

We appreciate the comments from the reviewer, please find our addition to the methods below:

Methods lines 700-721: "**The fluid flow domain selected is a cuboid measuring $L_x = 6$ mm, $L_y = 3$ mm, and $L_z = 3$ mm, housing shark denticles arranged in a three-by-three matrix at the bottom, depicted in Fig. S6A. The denticles are positioned 3 mm away from the inlet surface (see Fig. S6), with a constant inlet velocity (flow speed) set at 3×10^4 mm/s perpendicular to the inlet surface. This corresponds to 30 m/s (~108 km/h), which ensures a flow speed of 15-20 m/s (~54-72 km/h) over the dermal denticles, as it decreases due to water viscosity and the sharkskin surface no-slip condition. From a numerical standpoint, a high inflow speed is advantageous for understanding the flow pattern, as the drag force under low flow speed (low Reynolds numbers) will lead to laminar flow and thus minimal drag conditions. For the model size, as shown in the Fig. S5, real denticle sizes were chosen from the Micro-CT images of the sharkskin samples and reconstructed into 3D patterns for flow simulation in CFD. Then, as shown in Fig. S6B, & C, taking lantern shark denticles as an illustrative example, the size of the denticles is approximate at $D_x = 2.08$ mm, $D_y = 0.45$ mm, and $D_z = 2.21$ mm (Note: these dimensions may vary for denticles of different shark species used, and their sizes are detailed in Supplementary Data section 3.3). For the Reynolds number, we conducted an extensive analysis of drag coefficients**

on the lantern shark (Supplementary Data section 3.4, lines 765 to 768, also illustrated in Supplementary Fig. S8A and S8B) to verify the CFD simulation stability, accuracy and mesh dependency. Then, for the comparison between different shark species, we maintain a constant Reynolds number of 1×10^4 for all cases using the specified CFD setup, as shown in Supplementary section 3.4. Further details of the CFD model construction, simulation, numerical algorithm and the verification can be found in the supplementary materials."

b)

We have added the following to the legend of Fig. 6a, as requested by the reviewer:

"Fig. 6. Spectrum of denticle form and function. (**a**) Projection of functional variables (dots) defined by principal component axes (PC) 1 and 2, with PC1 describing hardness and stiffness and PC2 damage resistance, drag reduction and stiffness. Arrows demonstrate weighing and direction of the five variables. The direction of the red arrows indicates higher values for that function, while the blue arrows represent lower values. The colour gradient illustrates zones of high (red) and low (yellow) occurrence probability of functions in the trait space defined by the PC axes. **See Methods for calculations of PC1 and PC2** […]"**.**

c)

We appreciate the reviewer's inquiry regarding the definitions of hardness and stiffness as applied in our study of nanoindentation mechanical property measurements of shark denticles. **See revised text below.**

Lines 240-245: “For internal properties, we used nanoindentation (NI) to measure the enameloid hardness, **which is the material’s resistance to localized plastic deformation** (H, 137 indents per morphogroup on average), reduced elastic modulus (E_r, 137 datapoints per morphogroup on average), i.e., the material’s ability to withstand elastic deformation **when subjected to an applied force**, here referred to as the stiffness, and we calculated the H/E_r, also known as an indicator of the damage resistance[29].”

d)

Thank you for thinking of this question. It is a valid consideration. We will refer to **Author Response 14** as the reply will be the same in this case.

e)

We did not perform a correlation analysis between body size and functional properties. We include only sexually mature specimens in the study, and since the functional properties are describing a morphogroup, and thus not a species, we consider the ontogenetic stage of the individuals within each morphogroup similar and comparable.

f)

The functional trait analysis results are to be found in Fig. 6A. We refer to Supplementary Table S14 (new) for the raw data used to create Fig. 6A. **We clarify what the functional trait analysis is in the revised text below.**

Lines 258-263: ”We then used functional trait analysis (similar to a PCA) to integrate the mechanical properties, mineral composition, and fluid dynamic properties of the various denticle morphogroups (see methods section for further detail, Fig. 6, **Supplementary Table**

S14). **The functional trait analysis is essentially a PCA, which is able to minimize the dimensions of the projected trait data, and test the significance of each dimension in order to illustrate differences and correlations between the traits**[33]."

g)

We kindly refer to **Author Response 4** as a response to this comment.

h)

Thank you for pointing out this variation. **We have clarified the presence of individual variation of morphological characters of each morphogroup in the legend of Fig. 1, see below.** The new Fig. S3 that we refer to has been created in context with **Author response 24**.

"Fig. 1. The 10 denticle morphogroups from SEM image to 3D model. (**a**) deep learning workflow for producing the ten morphogroups. * See methods and Supplementary Fig. S10. (**b**) SEM images of the skin from the F1-body location, marked on shark vignettes, of 16 shark species (representing five orders and 10 families) with 3D models of the ten morphogroups. Coloured triangles with numbers 1-10 refer to denticle morphogroups. **Individual variation of morphological characters within each morphogroup is visible, see Supplementary Fig. S3 for more detail**. All scale bars represent 200µ."

Comment 19

2) For the hardness and stiffness measurements, you are measuring these variables on cross sections of the denticles from directions that the material seems unlikely to be loaded on. For example, I would guess that denticles are often scratched or loaded by pushing on the crown

towards the skin, but here your measurements are parallel to the skin and therefore orthogonal to my assumed load direction. Please spend some time justifying this. Do you think this is fair and how do you think this might be influencing your results? My assumption is that this tissue is likely to be organized in some way that would create anisotropy among different measurement orientations.

Author response 19

It is indeed true that in real-world scenarios, denticles are likely subjected to loading or scratching forces applied to the crown towards the skin. However, due to the geometric complexity of denticle ridges and the thinness of the enameloid layer, preparing samples to expose the enameloid layer from the top and conducting indentations in that orientation presents challenges. **See revised text below.**

Methods Lines 603-614: "NI: For each section, 10-30 indents were performed along the enameloid (using several denticles from the same sample), from which the average hardness (H), reduced elastic modulus (E_r), and damage resistance (H/E_r) were calculated (Fig. 5, Supplementary Fig. S11, Supplementary Note S1, Supplementary Tables S5-8). **We opted to conduct measurements parallel to the skin, which corresponds to lateral and cross-sections of the denticle. This orientation allowed for the enameloid layers to be well exposed after polishing, facilitating measurements across all ridges within the same plane. While it is acknowledged that this orientation may not perfectly mimic real-world loading conditions, we observed insignificant differences in mechanical properties between cross and lateral sections of denticles from the same body location of the same species. This suggests that any potential anisotropy in the material due to orientation variations is minimal and unlikely to significantly influence our results.**"

Comment 20

Are there sources from other mineralized tissues (e.g. bone) or morphological studies on the fine structure of the elasmoid layer that would suggest the material has similar properties regardless of direction? My intuition is that the opposite is often true.

Author response 20

Yes, Enax et al. (2012) found that nanohardness of shark teeth enameloid is the same regardless of direction. They performed nanoindentation on shark teeth cut in both axial and transversal directions, but found no significant difference in hardness values between the two. **We have added this reference to the section revised above in Author Response 19.**

35) Enax, J., Prymark, O., Raabe, D., and Epple, M. (2012). Structure, composition, and mechanical properties of shark teeth. Journal of Structural Biology 178:290-299.

Comment 21

3) A very interesting finding in your paper is that you say hardness and mineral content are negatively correlated. My intuition (and the literature) would suggest is that the reverse should be true and would be assumed here – more mineral content would create a harder tissue/material. So on my reading of your paper, this either seems like a really important and novel finding that could use a good deal more time spent on it or perhaps it's some kind of error.

Author response 21

This may be a misunderstanding. We do not state that hardness and mineral content are negatively correlated. We state that hardness and mineral protection (fluorine content) are negatively correlated, and discuss the idea of there being two ways to protect denticles from abrasions, 1) consisting of hard tissue (measured from nanoindentation) or preventing demineralization after the damage is done (measured from the fluorine content).

It is important to note that the hardness (mechanical properties) of enameloid is influenced by several minerals such as calcium, phosphorus, fluorine, and magnesium. The ratios of these elements, particularly Ca/P and Ca/F, provide a more comprehensive understanding of the mechanical properties than any single element alone.

We conducted correlation analyses between nanoindentation hardness and elemental compositions by calculating the correlation coefficients (Fig. 6D). Our results show that hardness is negatively correlated with fluorine, whereas calcium, phosphorus, the Ca/P ratio, and the Ca/F ratio are positively correlated with hardness. We have revised our text as follows:

Lines 302-314: "**Our results show that the Ca/P ratios of denticle morphogroups range from 1.89 wt% to 2.0 wt%, which are close to the ideal stoichiometry of the apatite structure (2.13 wt%), and hardness ranges from 4.2 to 5.9 GPa in these morphogroups. Comparatively, shortfin mako shark teeth, with a near-perfect stoichiometric Ca/P ratio (2.14 wt%), exhibit the highest hardness values (6.2 to 7.5 GPa), while port jackson shark teeth with the lowest Ca/P values (1.77-1.82 wt%) show the lowest hardness (3.7-4.6 GPa),[34, 35] indicating that the hardness of teeth enameloid is directly related to the Ca/P ratio. Furthermore, our samples indicate substantial fluorine content, ranging from approximately 1.82 wt% to 4.42 wt%, suggesting sufficient fluoride ions for the formation of fluorapatite compounds, which may provide an extra layer of protection against demineralization in denticles more prone to damage, as evidenced by the**

negative correlation between hardness and mineral protection in our functional trait analysis (Fig. 6A, D, see also supplementary material for additional discussion).”

Furthermore, the relationship between hardness and mineral content is discussed in detail in **Author response 53.**

Comment 22

4) What are the major ways your deep learning computer vision is using to separate out denticle ‘morphogroups’? The methods here are really cool, but without some idea of what differences these morphogroups represent, it’s difficult for readers to get a sense for what is important to distinguish your morphogroups. Also, providing some sort of framework here so that others can adopt and use your morphogroups would be really valuable and would increase the general impact of your results.

Author response 22

Our deep learning approaches are quite simple as described in Figure 1A and the methods: we trained a self-supervised VSC model, and extracted the latent vectors as quantified visual features of each input image of dermal denticles using the encoder of VSC, and then used K-means to separate out K morphogroups (K=20 in our research). Although we described many details of the VSC model in the methods, due to the nature of self-supervised learning, no labels were needed in the end-to-end training process, and all the augmentations depend on the variations of the input image dataset. The UMAP was only used for visualization and did not affect other results.

The extracted vectors are trained to learn representations of the full visual features of input images, and not to recognize the difference between human-assigned categories. And so, it is

difficult to say what exact features help us distinguish morphogroups because each of the morphogroups further summarized their member features of dermal denticles. However, we can directly observe the representative images selected near the centroid of each morphogroup in Figure 1B to get an idea.

We have uploaded a spreadsheet containing the name and number of species included in the morphogroup-formation-analysis (Table S4) on Figshare (https://figshare.com/s/ae0a3f288287d5afedc8), and also the essential codes used to create the denticle morphogroups (https://doi.org/10.6084/m9.figshare.22683577.v3).

Comment 23

5) I noticed that for morphogroup 7, the "smooth" morphogroup, that the S region from Scoliodon laticaudus is used, but in Fig S3 the CT scan images from this sample are obviously ridged. Are the samples used to represent the different morphogroups for minteral content, hardness/stiffness, and drag analyses actually categorized in those morphogroups in your previous exploration of denticle morphospace? If so, then why is the S region from Scoliodon laticaudus in the smooth morphogroup if it is not at all smooth surfaced?

Author response 23

Similar to Comment 18h), each morphogroup consists of morphological variation of the traits defining the group. The Smooth morphogroup also contains variations of the average denticle morph, and this was evidently reflected by the micro-CT scan in Fig. S5. **To make matters less complicated, we have micro-CT scanned another denticle sample belonging to the Smooth morphogroup (shortfin mako shark *Isurus oxyrinchus* snout location, panel F), and we have inserted it in Fig. S5. See below for results.**

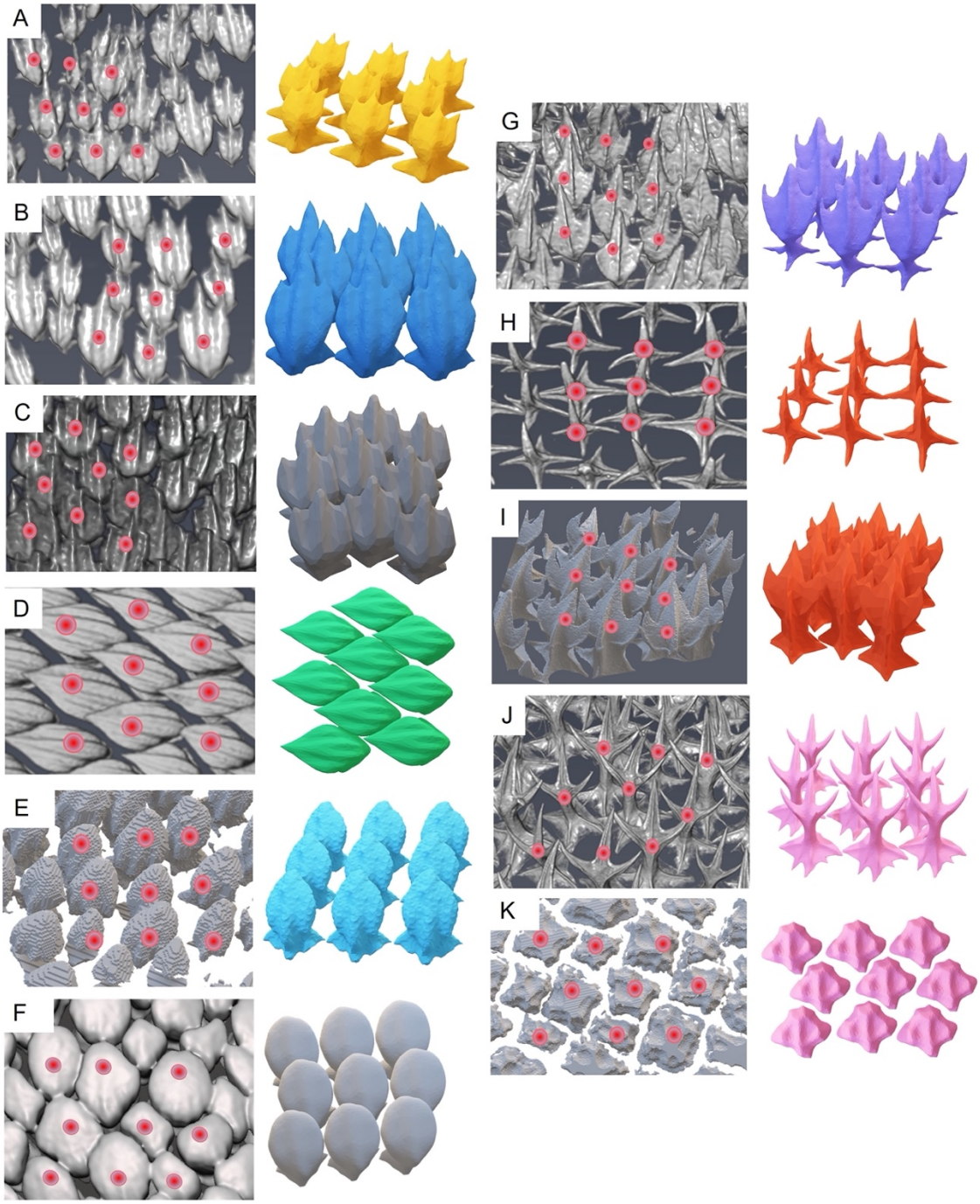


Fig. S5. Micro-CT reconstructed images of the sharkskin samples and their mimicked 3D patterns for flow simulation. (**A**) Spadenose shark C1 (**B**) Spadenose shark F1 (**C**) shortfin mako shark F1 (**D**) Topeshark C2 (**E**) Spadenose shark preANL (**F**) shortfin mako shark S (**G**) Longfin Catshark F1 (**H**) Lantern shark F1 (**I**) Mandarin dogfish F1 (**J**) Birdbeak dogfish B (**K**) Frill shark B. (**H**) and (**I**) represent the variation we observed in the spiky denticle

morphogroup (grp 10), as (**J**) and (**K**) represent that of the highly structured morphogroup (grp 9).

Comment 24

6) It would be valuable in the supplement to include a multipanel figure showing for EACH morphogroup that shows readers the variability for each group. Something like 6-8 examples of each morphogroup would be really valuable to help readers understand what is actually different here among morphogroups and how much disparity is within vs between morphogroups.

Author response 24

Thank you for this input. We already display part of this variation in Fig 1, however, we agree that it would be very helpful to show further examples. **We have created a new figure like this in the Supplementary Data, named "Fig. S3.", see below.**

We have added a reference to the new figure on line 138:

"The features of each morphogroup were expressed as an "average morphology", which we derived from the mean feature values at the centroid of each morphogroup in the 512-dimension morpho-space (see detailed examples in **Supplementary Fig. S3**)."

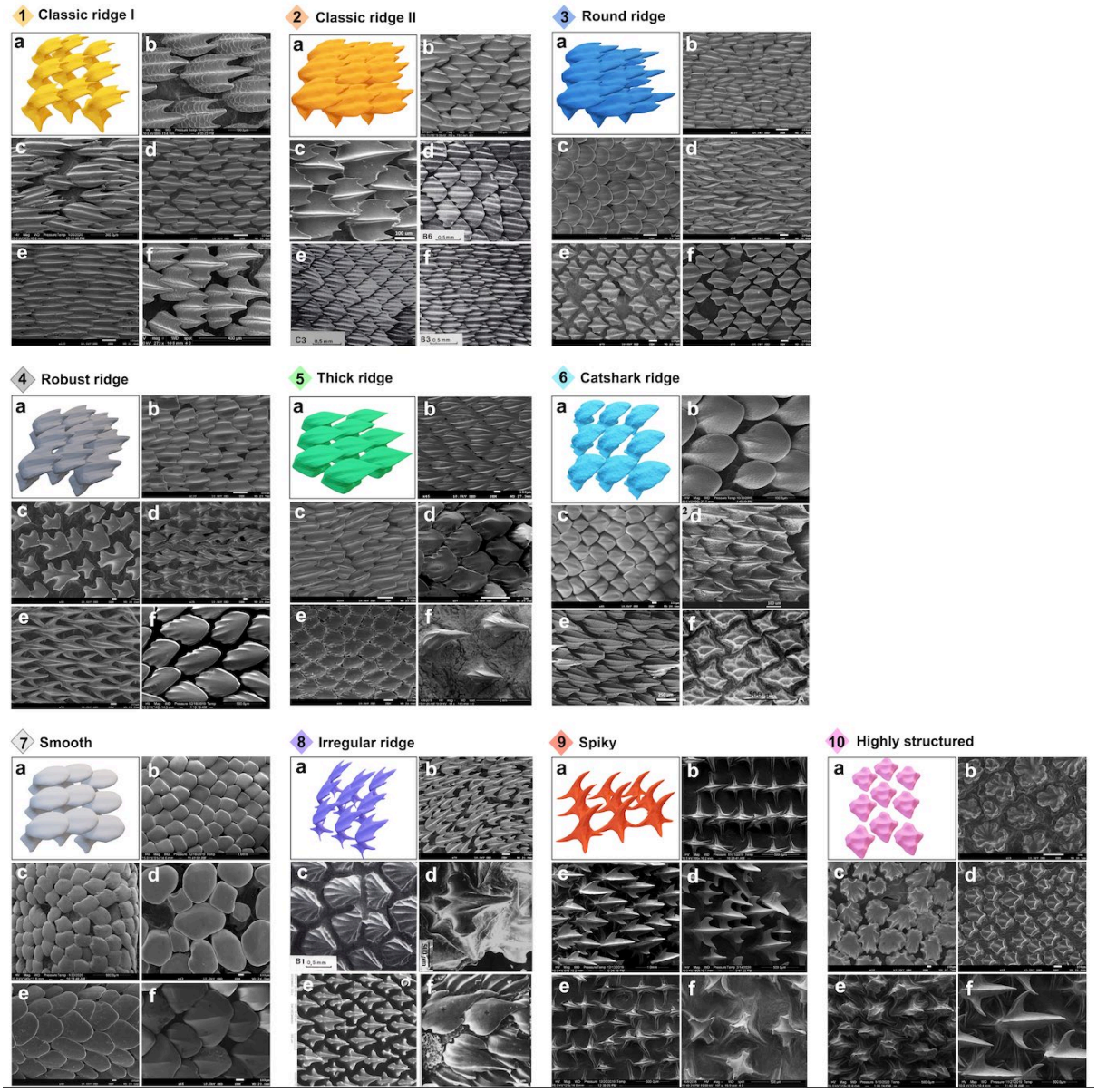


Fig. S3. Morphological variation within each morphogroup. a) displays the average morphology as a 3D-model. Grp 1b-f) spadenose shark *Scoliodon laticaudus* F1, spadenose shark C2, bigeye thresher shark *Alopias superciliosus* F2, scalloped hammerhead shark *Sphyrna lewini* F2, tope shark *Galeorhinus galeus* F1. Grp 2b-f) blue shark *Prionace glauca* F1, dwarf sawshark *Pristiophorus nancyae* F1, sandbar shark *Carcharhinus plumbeus* B6*, blacktip reef shark *C. melanopterus* C3*, Galapagos shark *C. galapagensis* B3*. Grp 3b-f) shortfin mako shark *Isurus oxyrinchus* F3, shortfin mako shark G, bluntnose sixgill shark *Hexanchus griseus* CP, crocodile shark *Pseudocarcharias kamoharai* F3, blue shark F3. Grp

4b-f) shortfin mako shark F2, whitespotted bullhead shark *Heterodontus ramalheira* B, kitefin shark *Dalatias licha* PEC1, birdbeak dogfish *Deania calcea* CP, whitespotted bamboo shark *Chiloscyllium plagiosum* C2. Grp 5b-f) Japanese topeshark *Hemitriakis japanica* C2, shortfin mako shark ID, zebra shark *Stegostoma tigrinum* 1DF, scalloped hammerhead shark PEC1, Greenland shark *Somniosus microcephalus* F1. Grp 6b-f) spadenose shark preANL, blind shark *Brachaelurus waddi* F1, Atlantic ghost catshark *Apristurus laurissoni* F2*, yellowspotted catshark *Scyliorhinus capensis* F2*, little sleeper sharks *Somniosus rostratus* H*. Grp 7b-f) whitespotted bamboo shark S, spadenose shark S, zebra shark M, whitespotted bullhead shark ANL, tawny nurse shark *Nebrius ferrugineus* F1. Grp 8b-f) longfin catshark *Apristurus herklotsi* F1, gulper shark *Centrophorus granulosus* B1*, sailfin roughshark *Oxynotus paradoxus* F1*, Taiwan spurdog *Squalus formosus* F2*, bull shark *Carcharhinus leucas* ANL*. Grp 9b-f) whitespotted bullhead shark ID, leafscale gulper shark *Centrophorus squamosus* S, frill shark *Chlamydoselachus anguineus* PEL2, blackbelly lanternshark *Etmopterus lucifer* S, arrowhead dogfish *Deania profundorum* F1. Grp 10b-f) blackbelly lanternshark F1, leafscale gulper shark F1, mandarin dogfish *Cirrhoscyllium barbifer* ID, sculpted lanternshark *Etmopterus sculptus* F1, horn shark *Heterodontus francisci* F1. Asterisk refers to images found in literature, see Table S4 for references.

Comment 25

7) Is it fair to assume that the functional traits found in 1-3 species in a morphogroup would be characteristic of all denticles from that morphogroup? This is an assumption you are making here that I think you could be open about and describe why you are making it and what limitations are inherent in it. (e.g. species might be different in material properties but similar in morphology (morphogroup)). You could instead frame it as if you are trying to

understand the connections between form and function instead of assuming form is representative of function in these morphogroups.

Author response 25

Thank you for the need of clarification. The reviewer is exactly right in the notion that we examine the form-function relationship, we do not assume it. **We would gladly accommodate the concerns from the reviewer, and we have added an explanation to state our assumptions:**

Lines 234-240: "**To explore the potential relationships between external morphology, internal structure and material properties, we quantified a variety of such characteristics for representative species within each denticle morphogroup (Fig. 5). While we acknowledge that species within a morphogroup may exhibit differences in material properties despite similarities in external morphology, we believe that sampling 2-4 species from multiple body locations within each morphogroup provides valuable insights into the relationship between form and function (see Supplementary Table S1 for sampling details).**"

Comment 26

Abstract, line 24: Specious is the incorrect word here – you likely were trying to use speciose.

Author response 26

Yes, thank you. We meant to use speciose, and we have corrected it.

Comment 27

Abstract, line 25: “although occupying ecologically diverse habitats” is incorrect grammar. Please fix this! Also, it’s not clear to the reader why it’s interesting/relevant for you to point out that these sharks live in different habitats yet all have denticle ridges. When re-writing this sentence, it might make more sense just to state it as a list of facts about Carcharhiniformes rather than using words like ‘although’ that imply a relationship between habitat and ridges that you have yet to explain at this point in the paper.

Author response 27

Thank you for pointing out the grammatical error in this opening sentence. **We have corrected it** as suggested, however, **the revision of the abstract can be seen in Author Response 29.**

Comment 28

Abstract, line 28: concluding should probably be a word more like ‘finding’.

Author response 28

Thank you for suggesting a different word use. We would like to use ‘discerning’ instead.

Comment 29

Abstract, line 29: you start the abstract by mentioning just Carcharhiniformes and now have broadened your results to all extant sharks. This is confusing – did you study Carcharhiniformes or all extant sharks? Be specific with your language here to match your methods and results. My suggestions would be to remove or completely rewrite the first

sentence (or move it towards the second half of the abstract). Your study is across all sharks, not just Carcharhiniformes.

Author response 29

Thank you for this comment. **We agree on the rewriting and have inserted the changes below**.

Lines 26-38:

”**Sharks have roamed the oceans since the Devonian, surviving multiple mass extinctions. Today, sharks are agile, streamlined swimmers, ambush predators hiding beneath the sand, and giants cruising in slow motion. Impressively, they are all covered in an armour of tiny skin teeth (dermal denticles).** For the first time, we objectively quantify dermal denticle disparity using deep learning on scanning electron microscopy images, discerning ten morphogroups across modern sharks (Selachii). Seven of the ten morphogroups display ridged denticles. Through nanoindentation, electron probe micro-analysis, and computational fluid dynamics, we establish mechanical properties, elemental composition, and swimming performance, and find that denticle ridges are highly variable and present a spectrum of functional properties beyond drag reduction. Further, we found a radiation event within the ground sharks **Carcharhiniformes** around 170 MYA, resulting in the enormous diversity of the clade, for which ridged dermal denticles may have been key.”

Comment 30

Line 48: you mention the ecological function of some dermal denticle types are well known and then mention drag reduction. These papers are based on just 1-2 species so I’m not sure I’d really consider them to be statements on denticle ‘types’, but instead only proof that drag

reduction seems possible in these species. Also, I'm not sure ecological function is what you mean. You might want to say something more like biomechanical function or just function.

Author response 30

We have rewritten the sentence to be more specific to the references given, and we will use ´function´ instead of ´ecological function´.

Lines 50-53: "Today, the functions of some dermal denticles are well known, such as how ridged denticles **of some fast swimming sharks** interact with the turbulent boundary layer to reduce the surface drag[5-7]."

Comment 31

Line 57-58: This sentence doesn't seem to have any connection to denticles – just about plate tectonics and shark diversity?

Author response 31

Thank you for pointing this out. We had omitted the word ´denticle´ in the sentence by mistake, and have **rewritten the sentence to make this statement clearer**.

Lines 60-62: "For instance, how ecological opportunities, **such as those** created by plate tectonics, **i.e., the opening of novel oceans,** have influenced the evolution of shark **denticl**e diversity."

Comment 32

Line 60: “their morphological diversity” needs to be more specific. Something like “denticle morphological diversity”.

Author response 32

We have complied fully, and added ´dermal denticles´ to the sentence for clarification.

Lines 62-64: “These important yet understudied questions are likely due to the difficulty in objectively quantifying and classifying **the** morphological disparity **of dermal denticles**.”

Comment 33

Line 67: What is representation learning? Please elaborate or define this briefly when you mention it.

Author response 33

We have added an explanation to representation learning at the first mention, **see below.**

Lines 74-80: “The process of subjectively defining important features is known as "feature engineering" in the field of artificial intelligence, and an important goal of representation learning is to reduce the need for feature engineering relying on human labours and decisions[10]. **Specifically, representation learning is a machine learning approach that aims to automatically discover the most effective way to represent data prior to recognizing differences between pre-assigned data categories**.”

Comment 34

Line 68: Why do we think ‘feature engineering’ needs to be reduced? Give a reason or two here to further justify your methods and educate those not familiar with this field.

Author response 34

We apologize for not being clear. We aimed to reduce the reliance on feature engineering that involves human labour and judgment. **We have revised the text to clarify this.**

Lines 75-76: "…and an important goal of representation learning is to reduce the need for feature engineering **relying on human labours and decisions**[10]."

Comment 35

Line 70: How is this deep learning method more objective? I think you need to explain this a bit more or somewhere in this manuscript and reference that passage here. I feel like someone could argue that experts are not usually selecting features subjectively but rather objectively (based on facts and evidence).

Author response 35

We agree on the need for this clarification. As a continuation of the revised text in **Author Response 34**, we have revised the next sentence as well, to accommodate the reviewer’s comment.

Lines 80-82: “Thus, by applying the deep learning method, we are able to solve the problem of experts having to subjectively select features to quantify complex morphologies.

Therefore, our study **approaches an unprecedented level of** objectively classif**ied** shark dermal denticles […]"

Comment 36

Lines 72-76: Very cool!

Author response 36

Thank you!

Comment 37

Line 94: what are F1- body location skin samples? You haven't explained or introduced this term.

Author response 37

Thank you for pointing this out, we have added a clarification to the text, see below.

Lines 125-129: "Further, we excluded any morphogroups represented by less than three F1-body location skin samples (**the flank region above the pectoral fins**), omitting eight morphogroups, for the purpose of simplifying the representations and for using this particular location to compare denticle morphology of different species in the ancestral state reconstruction (Figs. 2-3)."

Comment 38

Lines 93-99: Okay so you used deep learning to quantify morphology for you and then k-

means clustering to find 20 groups. Why did you initially start with 20 groups? This seems subjective. Provide justification here in your text.

Author response 38

We would like to kindly refer to **Author Response 4.**

Comment 39

Also, you then reduce 20 groups down to 10 by combining a few groups and then removing a few groups that have low sample sizes (I think). These decisions seem somewhat arbitrary – especially given your section just above in the introduction that makes a point of how your methods are objective and not subjective – it feels like you are taking your 'objective' dataset and then doing 'subjective' things to change the results. I think you need to justify these decision more in your text.

Author response 39

Author Response 4 also responds to this particular concern, thank you.

Comment 40

I would say that the one of the major advancements you have made here with this method is to quantify denticle surface morphology in extreme detail (to the point that you can reconstruct images from your quantified results). And you have done so on images of multiple denticles, not just single ones.

Author response 40

Thank you very much for appreciating our approach!

Comment 41

Lines 81-92: An important part of most morphological studies is deciding how to account (or not account) for covariation in body size. Presumably you sampled denticles from an array of different body sizes but it doesn't seem like your analyses are accounting for body size at all – typically body size is responsible for much of variation seen in most morphological datasets, and in those cases researchers will use various methods to account for the variation due to body size and remove it so that they are studying size-independent variation. I think many readers will expect you to justify NOT correcting for size (I'm assuming you didn't because I can't find a mention of it). I think somewhere here might be a good place to add 2-5 sentences justifying NOT accounting for size. A justification might also be really strong if you have analyses of body size vs your morphological variation to perhaps show how much variation in your data is "due" to body size alone.

Author response 41

The reviewer is correct in their understanding of how denticle morphology changes with body size. We did indeed consider body size in terms of ontogenetic stages (juvenile, transitional and mature) and how it affects dermal denticle morphology. We apologize for not making it clear that these considerations already exist in the Methods section. **We have thus revised the section the reviewer refers to, see below.**

Lines 108-115: "We used 2193 SEM images of skin to train the model, representing 53 different body locations across 132 extant shark species, 2 ray species, and 24 fossil species. Each SEM image represents a single body location from a single species and should be

regarded as a unique entity, which means that skin samples across the body of the same species may fall into several morphogroups. **We only include sexually mature specimens, as denticle morphology varies with ontogeny[14] (see Sample Collection in the methods section for more detail**, Supplementary Table S4, Supplementary Fig. S1)."

Comment 42

Lines 104-106: Referencing the figure is great, but your descriptions of the morphology of the denticle groups should be more elaborate. For example, if there are 7 groups with ridges on the crown, why did they get separated into different groups? Same with the smooth denticles (I assume this just refers to have no ridge, which would be a more specific description as smoothness/roughness may not correlate with ridge presence). Also, what is meant by highly structured. Please provide more information here or somewhere else in the paper so that readers get an understanding of your groupings.

Author response 42

This is a fair point. **We have elaborated on this as suggested, please see below.**

Lines 139-145: "The resulting morphogroups show that seven of the 10 groups include denticles with a variety of ridges on the crown (grp 1-6, 8), **each identified with different ridge-characteristics.** The remaining three groups include denticles that are smooth, **i.e., without prominent ridges** (grp 7), spiky (grp 9) or highly structured (grp 10). **The latter referring to denticles with elaborate structures neither fitting in the groups of ridged or spiky denticles (see Supplementary Fig. S3 for morphological variation within the 10 morphogroups)**."

Comment 43

Lines 113-114: “variation in denticle morphology across a single individual can be enormous” should have some citations.

Author response 43

We have added the following reference to support our claim:

16) Ankhelyi, M. V., Wainwright, D. K. and Lauder, G. V. Diversity of dermal denticle structure in sharks: Skin surface roughness and three-dimensional morphology. Journal of Morphology 1-23. (2018).

Comment 44

Lines 117-119: Here you should give us an idea of how many species and individuals for which you had data on the flank (F1) region? In other words, tell us how much data went into this ancestral state reconstruction.

Author response 44

We agree and have complied, see below.

Lines 164-167: “We thus focused on the F1-body location when examining the ancestral states, mapping the ten morphogroups **for 73 extant shark species representing seven orders and 23 families**, and when calculating the diversification rates of different clades (Figs. 2-4).”

Comment 45

Line 132: “developed ridged denticles reflection their ecological niches”. You make this statement as if there is a connection between ridges and ecology but I don’t think you’ve

discussed this at this point in the paper. I think you need to either spend some time in the introduction or here explaining this idea and providing support for it.

Author response 45

We are grateful for this comment, as it is a mistake on our part to phrase it as we did. **We have altered the sentence to the one below.**

Lines 177-180: "Squalomorph sharks are mainly found in cold, deeper waters, with a variation of pelagic-oceanic, bathydemersal and benthopelagic lifestyles[22,23], however, some have diverged to shelf-related habitats and **developed novel lifestyles reflecting these** ecological niches."

Comment 46

Line 180: shark should be sharks

Author response 46

We have changed it as suggested, thank you.

Comment 47

Lines 183-192: As in above comments, you should provide some idea of sample sizes here – for many of these methods you are working with a subset of species and you need to be clear about that here so that readers don't get confused.

Author response 47

We have complied and added sample sizes, **see below.**

Lines 240-246 […] 249-250 […] 254-258: "For internal properties, we used nanoindentation (NI) to measure the enameloid hardness (H, **137 indents per morphogroup on average**), reduced elastic modulus (E_r, **137 datapoints per morphogroup on average**), i.e., the material's ability to withstand elastic deformation, here referred to as the stiffness. From these, we calculated the H/E_r, also known as an indicator of the damage resistance[29] (Fig. 5A, Supplementary Note S1, Supplementary Tables S4-7). […] Additionally, we use the F wt% as an indicator of mineral protection **(89 datapoints per morphogroup on average)**[30,31]. […] Regarding external properties, we further scanned the samples with micro-CT and recreated 3D models of representative morphogroups to analyse their capacity for drag reduction with computational fluid dynamics, **using 1-2 samples per morphogroup** (Fig. 5C, Supplementary Figs. S5 -S7, Supplementary Note S3, Supplementary Table S1 and S10)."

Comment 48

Line 195: It might be good to provide a citation for functional trait analysis and spend some time explaining how it was used here. I'm unfamiliar with this analysis and I suspect most other readers would be too. Yet, this functional morphospace is one of the most novel and interesting aspects of this paper so it is really important to clearly communicate to readers what is going on and how the functional trait analysis is performed.

Author response 48

We have complied and added an explanation to the functional trait analysis and a reference using this method as a main tool, **see below.**

Lines 258-263: “We then used functional trait analysis (similar to a PCA) to integrate the mechanical properties, mineral composition, and fluid dynamic properties of the various denticle morphogroups (see methods section for further detail, Fig. 6, Supplementary Fig. S6A, B). **The functional trait analysis is essentially a PCA, which is able to minimize the dimensions of the projected trait data, and test the significance of each dimension in order to illustrate differences and correlations between the traits[33]**.”

33) Diaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167-171, doi:10.1038/nature16489 (2016).

Comment 49

Lines 198-199: Elaborate a bit into the transformation of data into performance ranks! Does this just mean each model was ranked from 1-10? How is this happening…also this ranking also presumably says nothing about statistical significance.

Author response 49

We have complied and revised the text as follows:

Lines 263-267: “We transformed the **functional** data **from Fig. 5A** into performance ranks between 0-10, with 0 indicating a low value (Supplementary Table S11), **in order to standardize the various properties and facilitate comparison across different metrics (Fig. 5D). This ranking system serve as a method for organizing and comparing relative performance and does not inherently address statistical significance.”**

Comment 50

Why did you not also test for statistical significance among groups?

Author response 50

We apologize for not being clear about the statistical significance among the groups. We did perform an ANOVA test, and we have now revised the text to clarify.

Lines 240-241 […] 242-243 […] 245-249 […] 252-254: "For internal properties, we used nanoindentation (NI) to measure the enameloid hardness […] reduced elastic modulus […] and we calculated the H/E_r, also known as an indicator of the damage resistance[29] (Fig. 5A, Supplementary Note S1, Supplementary Tables S4-7). With electron probe microanalysis (EPMA) we confirm the biomineral of the enameloid as fluorapatite ($Ca_5(PO_4)_3F$) from the Ca/P ratio and F wt% (Fig. 5A, B, Supplementary Note S2, Supplementary Table S8). […] **We tested for statistical significance of internal properties among the morphogroups using ANOVA (Fig. 5A, see methods for details).**"

Comment 51

Line 205: You mention resisting demineralization, but I don't follow how you are measuring resistance to demineralization. Please add something about this interpretation either here when you first mention this or just in the paragraph above when you are briefly explaining your measurements. To me, demineralization resistance sounds like you are thinking about how these scales would resist either ocean acidification or some other chemical process that would demineralize the denticles. If you instead are referring to something like abrasion resistance then be clear about that and maybe choose a different term. Regardless, please be more clear here about what measurements this interpretation is related to.

Author response 51

We understand the confusion here. We have revised the text at the first mention of F wt% to clarify what is meant later on, and added a reference. **See below.**

Lines 246-252: "With electron probe microanalysis (EPMA) we confirm the biomineral of the enameloid as fluorapatite ($Ca_5(PO_4)_3F$) from the Ca/P ratio and F wt% (Fig. 5A, B, Supplementary Note S2, Supplementary Table S8). Additionally, we use the F wt% as an indicator of mineral protection, not to be confused with mineral content (89 datapoints per morphogroup on average)[30,31]. **As fluorapatite is commonly known for its protective properties against tooth caries, also known as demineralization, we will use these terms in later sections[32]."**

32) Pajor K, Pajchel L, Kolmas J. Hydroxyapatite and Fluorapatite in Conservative Dentistry and Oral Implantology-A Review. Materials (Basel). 2019 Aug 22;12(17):2683. doi: 10.3390/ma12172683.

Comment 52

Lines 234: "This concurs with our correlation analysis […]". The only correlation analysis mentioned at this point is correlations between categorical traits. Here you are talking about correlations among the performance/functional variables. These results should be in the methods and in the results – they shouldn't first pop up in the discussion.

Author response 52

We have revised the text to avoid confusion, **see below.**

Lines 366-369: "This concurs with our **functional trait** analysis, where we found hardness and mineral protection to be negatively correlated, indicating that denticles more prone to damage have a secondary defence, i.e., high fluorine content (Fig. 6A, D)."

Comment 53

Lines 235: Hardness and mineral content are negatively correlated? If true, this is extremely interesting as I believe we typically think of high mineral content producing harder tissue. Please elaborate more on these results and show a scatterplot graph of this data as I think this is counter to expectations and therefore either/both suspicious and also really interesting if true! It would be good to cite some literature that looked at these variables (or similar ones such as hydroxyapatite concentration vs hardness) in other mineralized structures.

Author response 53

We kindly refer to **Author response 21.** Regarding the latter suggestion, we have added the discussion in the Supplementary Data and updated Supplementary Table S2, as follows:

Supplementary Data Lines 872-895: "**The comparison of Ca/P ratios and hardness values among the denticle morphogroups, shark teeth enameloid, and human teeth enamel provides further insights into the relationship between chemical composition and mechanical properties (Supplementary Table S2). The Ca/P ratios of the denticle morphogroups in our study, ranging from 1.89 wt% to 2.0 wt%, are close to the ideal stoichiometry of the apatite structure (2.13 wt%). The hardness of these morphogroups ranges from 4.2 to 5.9 GPa. In comparison, shortfin mako shark teeth enameloid, with a near-perfect stoichiometric Ca/P ratio (2.14 wt%), exhibits the highest hardness values (6.2 to 7.5 GPa), while port jackson shark teeth with the lowest Ca/P values (1.77-1.82 wt%) show the lowest hardness (3.7-4.6 GPa)[35]. This comparison indicates that the hardness of teeth enameloid is directly related to the Ca/P ratio, with near-ideal stoichiometry resulting in higher hardness and deviations leading to lower hardness values.**

Furthermore, the substantial presence of fluorine in our samples, ranging from approximately 1.82 wt% to 4.42 wt%, suggests sufficient fluoride ions for the formation of fluorapatite compounds. The incorporation of fluorine into the apatite structure has been shown to enhance the stability and resistance to demineralization[90,91]. This extra layer of protection provided by fluorapatite may be particularly beneficial for denticle morphogroups more prone to damage, as evidenced by the negative correlation between hardness and mineral protection in our functional trait analysis (Fig. 6A, D).

The relationship between Ca/P ratios and hardness observed in our study is consistent with findings from studies on the effect of Ca/P ratios on the hardness of bioceramics. These studies have shown that ratios either lower (1.57 at%) or higher (1.87 at%) than the stoichiometric value (1.67 at%) result in comparatively lower hardness due to the occurrence of a secondary phase in the apatite structure[92,93]. The presence of a secondary phase can disrupt the crystalline structure and lead to reduced mechanical properties."

Updated version of Supplementary Table S2:

Morphogroup	Hardness (H) GPa	Fluorine (F) wt%	Calcium (Ca) wt%	Phosphorus (P) wt%	Ca/P	Ca/F	Instrument
1 Classic ridge I	4.8 ± 1.3	4.0 ± 1.0	33.5 ± 0.4	16.5 ± 0.1	2.0 ± 0.1	8.4 ± 1.9	EPMA
3 Round ridge	4.5 ± 0.7	2.9 ± 1.2	32.6 ± 1.4	16.5 ± 0.5	1.97 ± 0.1	11.1 ± 4.5	EPMA
4 Robust ridge	4.2 ± 0.8	3.9 ± 1.3	31.3 ± 1.0	16.2 ± 0.5	1.93 ± 0.3	7.9 ± 2.6	EPMA
5 Thick ridge	5.0 ± 0.6	2.3 ± 0.7	33.7 ± 0.2	16.9 ± 0.3	1.99	14.8 ± 4.5	EPMA
6 Catshark ridge	4.6 ± 0.5	1.8 ± 0.4	32.4 ± 0.5	17.0 ± 0.1	1.89 ± 0.1	17.8 ± 3.9	EPMA
7 Smooth	4.6 ± 1.3	3.7 ± 1.4	32.5 ± 1.1	16.4 ± 0.4	1.97 ± 0.1	8.8 ± 3.3	EPMA
8 Irregular ridge	5.9 ± 1.0	2.2 ± 0.3	33.4 ± 0.1	17.0 ± 0.3	1.97	15.3 ± 2.1	EPMA
9 Spiky	4.7 ± 1.0	4.4 ± 1.1	33.4 ± 0.2	16.7 ± 0.5	1.97 ± 0.1	7.4 ± 1.7	EPMA

10 Highly structured	5.7 ± 1.4	3.1 ± 1.1	33.4 ± 0.3	16.7 ± 0.3	2.0 ± 0.1	11.0 ± 3.6	EPMA
Shortfin mako shark (*Isurus oxyrinchus*) teeth[35] *	6.2-7.5	3.08	-	-	2.14	-	IR, XRD**
Port Jackson shark (*H. portusjacksoni*) teeth[34]	3.7-4.6	-	-	-	1.77-1.82	-	EDS
Human teeth[35] *	6.3-7.0	0.01	-	-	2.03	-	EDS
Puffadder shyshark (*H. edwardsii*) denticle[91]	-	0.59 ± 0.1	24.0 ± 6.1	11.6 ± 2.0	2.0 ± 0.2	-	EDS

Table S2. **Nanoindentation hardness and e**lemental values of morphogroups. **Hardness (H) measured in GPa, fluorine content (F),** calcium (Ca) and phosphorous (P) weight percent (wt%), and Ca/P ratio with standard deviations of nine morphogroups (present study), **oral teeth of shortfin mako shark**, denticles of the puffadder shyshark, oral teeth of humans, and oral teeth of the Port Jackson shark. The present study is the only study using EPMA (electron probe micro-analyzer) where other studies use EDS (energy dispersive X-ray spectroscopy**), **IR (infrared spectroscopy) and XRD (X-ray powder diffraction). Ca/P ratios are calculated from weight percentages (wt%) of calcium and phosphorus. *For easier comparison, the atomic % of Ca/P ratios in Enax et al. (2012) were converted to (wt%).**

Comment 54

Lines 273-275: I don't understand why you are talking about form/pressure drag here? You mention it and then don't really use it for anything in your writing so I think you could remove this to help be clear for readers.

Author response 54

We have deleted this sentence to clarify the text.

Deleted lines: “In addition, the thrust of the caudal-fin causes longitudinal tip vortices along the body creating ‘drag due to lift’, which combined with ‘form drag’, i.e., the difference in pressure around the body, creates pressure drag[29,30].”

Comment 55

Lines 278-280: The example of how stiffness in octocorals might be useful to reduce drag seems pretty irrelevant to thinking about what is going on in shark denticles. In that study they measured the material properties of pieces of the whole organism and here you are measuring material properties of parts of the small denticles that are embedded in the flexible skin in a variety of overlap patterns. I don’t think there is any reason to believe that flow is changing the shape of denticles appreciably so I don’t think we would assume there would be a connection between stiffness and flow in denticles, as you are saying here. Denticles are hard, small, and mineralized so why would their stiffness be related to flow and drag in the way you are suggesting? There should be no morphing of denticle shape (which is what is seen in the octocorals paper because they are soft corals – so a much softer and deformable tissue).

Author response 55

We concur. We have found a more appropriate reference that correlates structural stiffness to maintaining streamlining during flow.

Lines 354-355: ”Further, **a stiff structure creates less turbulence compared to a soft structure[38]**, however, we found stiffness to be independent of drag reduction in shark denticles (Fig. 6A).”

38) Luo, Y. *et al.* A fluid-structure interaction solver for the study on a passively deformed fish fin with non-uniformly distributed stiffness. *Journal of Fluids and Structures* **92** (2020).

Comment 56

Line 282: Related to the ideas and comments above, you mention how bristling might require flexibility of denticles, but again I would state that I don't think there would be much actual bending of denticles during bristling. The denticles simply rotate at their base because they are attached to skin tissues which are soft and deformable…the denticles themselves are likely not bending much.

Author response 56

We have deleted out statement about bristling.

Comment 57

Lines 287-292: This is an interesting and surprising result! Does damage resistance to a denticle (as measured here) also equate to damage resistance to the shark? They could be related but I could also imagine that sharks might evolve denticles that are more prone to scratching so that they 1) don't break and 2) absorb force to prevent damage to the skin.

Author response 57

It is indeed an interesting result. Following this, we discuss that the external morphology, such as the thickness and sheer size of the denticle, may influence the damage resistance. However, this is not something we tested for. Also, a secondary defence in terms of a high

Fluorine content may protect the denticles when damaged.

Comment 58

Line 284-285: This statement about denticle ridges providing more control is not a well supported statement and is not experimentally tested or supported at all. If anything, it is just a hypothesis that you are here stating as if it were fact. Please be wary of this type of overstatement and consider tempering your language here and making it obvious this has not be experimentally shown or tested.

Author response 58

We have altered the sentence to make it much more specific to certain results within this field of research. We agree that the original sentence was too general in meaning. We stand by the four different references supporting the statement in this sentence.

Lines 356-359: "Lastly, although we measured the highest drag reduction for the smooth denticles (grp 7) (Fig. 5D, Fig. 6D), most studies **reach the conclusion that specific spacing and depth of denticle ridges affect the ability to interact with the boundary layer and thus provide control** while moving through water at different speeds and angles, being more efficient for the shark[5,39-41]."

Comment 59

Line 336: This statement about how ridged denticles and their "ability to optimize swimming performance" have driven the diversification of carcharhiniformes seems like a pretty large overstatement. In part because I don't think there is really good evidence that ridged denticles optimize swimming performance and ridged denticles appear in many other shark groups.

Also, how are you connecting the evolution of ridged denticles to this change in diversification rate? Where is the evidence for that? I think it's an interesting point and idea but I think this statement needs to be revised a bit more.

Author response 59

We thank the reviewer for this comment. It has enhanced the clarity of our results to revise the text as suggested. We have changed the language to directly reflect the results of the transition rate and clarified both the results and the discussion.

Results, lines 216-231:

"**We further compared the diversification rates of ridged and non-ridged dermal denticles using the binary state speciation and extinction model (BiSSE)[27] implemented in the R package diversitree[28]. We found that clades with ridged dermal denticles generally have a higher diversification rate compared with the non-ridged denticles, although with no significant difference (P=0.3026, Supplementary Fig. S4D). We further examined whether denticle ridges were related to specific ecological niches.** To understand the relationship between shark dermal denticles and ecological niches, we conducted a correlation analysis of trait and habitat evolution. Our results show that when analyzing all clades together, dermal denticles **of sharks** in shallow water environments, i.e., species carrying out most of their life in the photic zone above 200 meters depth, tend to evolve from non-ridged (grps 7, 9, 10) to ridged (**P <0.05**, grps 1-6, 8, Supplementary Fig. S4A). However, in deep-sea environments, i.e., species living below 200 meters depth, there is no significant evolutionary direction between non-ridged and ridged denticles (Supplementary Fig. S4A). **We also found a higher transition rate from a reef to a non-reef environment for sharks carrying ridged denticles (P <0.05, Supplementary Fig. S4B-C). Our results suggest a**

strong correlation between denticle type, ridged vs. non-ridged, and the ecological niches of the sharks."

Discussion, lines 399-407 […] 410-417:

"Furthermore, by examining the relationship between the phylogenetic tree and morphological evolution of shark dermal denticles, our study indicates that a rapid species diversification occurred around 170 million years ago (MYA), significantly increasing the taxonomic diversity of the Carcharhiniformes **(Fig. 2). The species within this diversified clade** primarily inhabit shallow waters, **are associated with a non-reef environment, and they all carry ridged denticles (Fig. 3).** We observed all seven ridged morphogroups (grps 1-6, 8) **within this clade** (Figs. 2-3). Our correlation analysis of traits and ecological niches also revealed that in shallow-water **and non-reef associated** environments, shark denticles are more likely to evolve from non-ridged to ridged forms (Supplementary Fig. S4A**-C**).

[…]

These significant marine ecological opportunities, **in combination** with **existing** ridged denticle **morphogroups** may have been the primary driving force behind the rapid diversification of the Carcharhiniformes, around 170 MYA **(Fig. 2)**. The dramatic changes in the marine environment caused by plate tectonics during this time also led to a substantial increase in the taxonomic diversity of corals[57,58] and benthic animal groups[55]. **The ridged denticles' of the Carcharhiniformes are generally optimized for drag reduction (Fig. 6A, D), among other related functions[60], and may have provided an advantage for the sharks to explore shallow habitats** (Fig. S4A-B)**.**"

Comment 60

Your Fig 2 shows that pretty much the same denticle morphotypes were present before

carcharhiniformes and their increased diversification rate…so I feel like the evidence isn't consistent with saying the ridged denticles or morphotypes you described here are related to that increased diversification rate.

Author response 60

As we mentioned and revised above, we believe that the success of sharks can be attributed to a combination of ecological opportunities (the expansion of the Atlantic Ocean and the increase in shallow habitats) and two other factors, such as ridged denticles. Therefore, it is reasonable to assume that ridged denticles already existed. We kindly refer to **Author response 59.**

Comment 61

Line 363: Which ancient group of sharks are you referring to? Or maybe you just mean all sharks? (Also I think you can just say marine ecosystems instead of most marine ecosystems).

Author response 61

Thank you for pointing this out, we have revised the text.

Lines 439-441: "Ultimately, our study showcases how integrating deep learning, materials science, and evolutionary analysis provides a comprehensive understanding of how **an** ancient group**, such as sharks,** has managed to maintain its dominant position in marine ecosystems."

Comment 62

Line 382: You mention including shark SEM images from the literature. You need to cite what literature here! Also, give a general idea of how many species and indivuduals were sampled using the fish markets, the natural history collections, and the literature. You could even break it down by each market and natural history museum and literature piece. Also please mention somewhere how many individuals per species you sampled.

Author response 62

We have already provided this information in Supplementary Table S4. Accordingly, we have modified this section above to more clearly refer to Supplementary Table S4. See below.

Lines 463-465: "**A list of sampled species and body locations from literature, museum collections, and fish markets in Taiwan, with references to literature, can be found in detail in** Supplementary Table S4 **and** Supplementary Fig. S1."

Comment 63

Lines 385-386: Supplementary table S4 was not attached (at least in the reviewer files) yet this is a very important table. Make sure it is included it includes all references and information.

Author response 63

Thank you, and we agree. The Supplementary tables may not have been available to reviewers and we understand the lack of information as a result. It contains all the sample

information and has been carefully setup to include as many details as possible. **It is available through Figshare through the links at the end of the Manuscript file.**

Data and materials availability: Data are provided with this paper in Supplementary Data and can be accessed on Figshare.

Supplementary Data Tables S4-S14: https://figshare.com/s/ae0a3f288287d5afedc8

Supplementary Data Movies S1-S12: https://figshare.com/s/b8f2d28e8599f98b2b3d

Supplementary Data Movies S13-S34: https://figshare.com/s/307292d3169b0af4ad0f

R code: https://doi.org/10.6084/m9.figshare.22683577.v3

Comment 64

Lines 389-396: Your SEM preparation methods of drying skin between weighted glass slides almost certainly deformed the surface by modifying the orientation and spacing of denticles. 1) You should be clear about this potential downside and mention/discuss it somewhere. And 2) You should discuss how you think it changed your results?

Author response 64

Thank you for pointing this out. **We have revised the text below.**

Lines 468-476: “Skin was washed and cleansed with water and a 0.5 mm brush, placed between two glass slides with added weights at 45ºC for at least 1 week to evaporate retained fluids. **The weights were used to flatten the skin while drying in order to allow for a large, consistent area for photography (SEM). The same technique is used by Reif (1985) who created the most comprehensive image database of shark denticles, and we follow the same preparations in order to facilitate comparison. Additionally, the**

flattened pieces of skin are close to their original shape when they are harvested from the sharks. As sharks are often in constant motion, it would be impractical to establish a "natural" position of dermal denticles from different species, and we thus work with the denticle positions from the flattened pieces of skin."

Comment 65

Line 426: You should also cite this scikit-learn package in your bibliography – not just provide the website.

Author response 65

We have added the reference for the Scikit-Learn package in our bibliography.

66) Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* **12**, 2825-2830 (2011).

Comment 66

Line 428: Why did you want clustering results to match clades in the phylogenetic tree? Please justify and explain this.

Author response 66

The manuscript has been updated to clarify the rationale behind the selection of the K-means algorithm and the objective steps taken to cluster the dermal denticle images into morphogroups. The choice of clustering method was guided by two key principles: natural clustering and limited within-group variance. After testing four clustering methods, K-means was chosen because it best adhered to these principles. We also emphasized that the

clustering was performed objectively, without consideration of phylogenetic information. See **Author Response 14** for a detailed explanation of the updated clustering method and the steps taken to obtain the final 10 morphogroups.

Comment 67

Line 429: divide should be divided

Author response 67

We have corrected this.

Comment 68

Lines 455-457: Again, all of these softwares and packages should be cited in your references.

Author response 68

We have added references to the softwares and packages as suggested. See below.

NumPy:

67) Harris, C. R., Millman, K. J., van der Walt, S. J. & al., e. Array programming with NumPy. *Nature* **585**, 357-362, doi:10.1038/s41586-020-2649-2 (2020).

Pandas:

68) McKinney. in *Proceedings of the 9th Python in Science Conference* Vol. 445 (2010).

69) pandas-dev/pandas v. 2021 (Zenodo, 2020).

SciPy:

70) Virtanen, P. *et al.* SciPy 1.0: Fundamental algorithms for scientific computing in python. *Nature Methods* **17**, 261-272 (2020).

Scikit-Learn:

66) Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research* **12**, 2825-2830 (2011).

UMAP:

71) McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *Journal of Open Source Software* **3**, 861, doi:https://doi.org/10.21105/joss.00861 (2018).

Comment 69

Lines 460-466: You need to add two things here – 1) the names of the species that you sampled, and 2) clarity that a single species is being used for multiple morphogroups in many cases. The way this is written makes it seem like you selected 10 species and each one was a separate morphogroup, but you are actually using different regions from each species as different morphogroups. Just make sure these things are obvious in the main text so that readers have understood what you have done.

Author response 69

We understand the need for clarification and we have complied. See revised text below.

Lines 575-591: "**Ten** shark **species were found** to represent the morphological variation in nine morphogroups, excluding group 2 due to lack of physical material (see species

Supplementary Table S1). **Species include the spadenose shark *Scoliodon laticaudus*, whitespotted bamboo shark *Chiloscyllium plagiosum,* blue shark *Prionace glauca*, shortfin mako shark *Isurus oxyrinchus,* Japanese topeshark *Hemitriakis japonica*, longfin catshark *Apristurus herklotsi*, blackbelly lanternshark *Etmopterus lucifer*, mandarin dogfish *Cirrhigaleus barbifer*, frill shark *Chlamydoselachus anguineus,* birdbeak dogfish *Deania calcea.* The species and sampled body locations were chosen based on taxonomic breath, prior morphogroup formation. The NI and EPMA data of the collective samples were allocated the AI-defined morphogroups post data collection. See Supplementary Table S1 to see how NI and EPMA data has been distributed between the morphogroups.** From these **10 species** we sampled **a total of** 10 different body locations **concluding 37 samples**: tip of snout (S), abdomen between pectoral fins (B), anterior edge of the pectoral fin (PEC1), flank region above pectoral fin (F1), below second dorsal fin (F3), laterally on the caudal peduncle (CP), lateral flank of dorsal caudal lobe (C2), pre-anal fin region (preANL), anterior edge of dorsal caudal lobe (C1), and posterior region of the pectoral fin (P3) (Supplementary Fig. S1). **The NI and EPMA data representing each morphogroup is thus composed of data from multiple species and body locations.**"

Comment 70

Line 514: You need to be specific about what shark skin samples were CT scanned. It also makes sense to state why you did this as well.

Author response 70

We agree and have added the following to the section:

Lines 647-651: "**We used the following samples for micro-CT: spadenose shark C1, F1, S, preANL, shortfin mako shark F1 and S, topeshark C2, longfin catshark F1, lantern shark F1, mandarin dogfish F1, birdbeak dogfish B, and frill shark B (see overview and pictures in Supplementary Fig. S1 and S5). These samples were chosen post morphogroup formation to represent the average morphology of all groups, and to use for the computational fluid dynamics analyses.**"

Comment 71

Lines 526: Somewhere around here (but probably also in the results) you need to at least briefly mention how you decided to arrange denticles on these 3d models. Fig 5 shows that the spacing between denticles in different models is different so make sure that is obvious here and briefly describe how you determined that. At line 531 I know you say you are "imitating the distribution of denticles" from biological samples but how specifically did you imitate? Did you use actual measurement or just visually decide that the models looked similar to the biological samples.

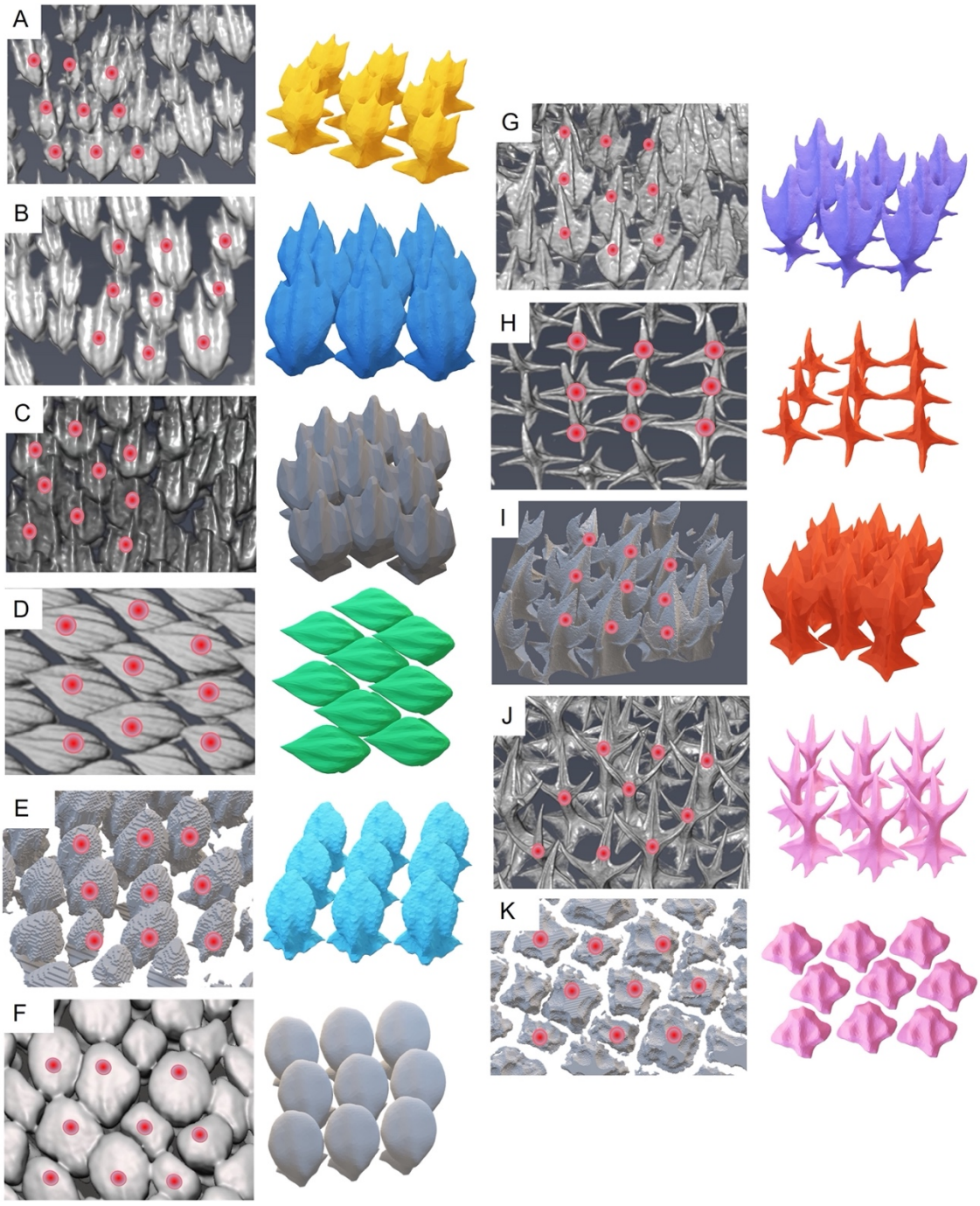
Author response 71

This is a good point. We have complied and revised the text, see below. Further, we have modified Supplementary Figure S5 (previous Fig. S3) to illustrate the patterns observed in reconstructed skin, which were translated into 3D models for the drag study.

Lines 668-675: "Finally, the isolated denticle model was exported to Solidworks software (Dassault system) to create denticle patterns imitating the distribution of the denticles in a particular body location of the shark species (Supplementary Fig. S5). **In order to ensure computational feasibility, we opted for patterns featuring nine denticles on a rectangular**

base plate. To simplify the representation, we aimed to retain the overall pattern similarity by overlaying the constructed model onto the real patterns and adjusting the positions accordingly. Real denticle sizes were preserved in the 3D models used for the computational fluid dynamics modelling."

Alterations to Figure S5 are seen as a red-dotted pattern on each denticle crown:



Comment 72

Line 533-534: How did you create a 3d geometry using only SEM images? You need to explain your methods either here or in the supplement and reference them here.

Author response 72

We have complied fully, see added text to the Methods below.

Lines 675-684: "**The 3D model of morphogroup 2 was based on SEM images only, using ImageJ software to detect the edges and outlines of an individual denticle in the SEM image. The resulting outlines were exported as Scalable Vector Graphics (SVG) files, which were then imported into SolidWorks. Within SolidWorks, the SVG files were imported to create 2D sketches. Employing features such as Extrude and Revolve, the 2D sketches were transformed into 3D geometry, adjusting parameters such as thickness and height to match the dimensions observed in the SEM images. Additional features, including grooves or ridges, were incorporated into the 3D model as needed, using functions such as curved surface tools. This model was not subjected to fluid dynamic simulations or used for quantitative analysis in drag studies. Its sole purpose was for representation as a 3D model.**"

Comment 73

Lines 535: Somewhere in this section you should describe if you preserved biological denticle size from the CT scans or if you are scaling all the denticle models to the same size. This is an important distinction that allows for proper interpretation of your results.

Author response 73

We agree and have added this detail to the mentioned section:

Lines 674-675: "**Real denticle sizes were preserved in the 3D models used for the computational fluid dynamics modelling**."

Comment 74

Lines 533-539: The spadenose shark was used as the example denticle for groups 1, 3, 6, and 7 – presumably though from different regions of the body. This is never really made obvious in the paper and while it is understandable you need to be more forthcoming about this. Having a single species represent more than 40% of your 'morphogroups' is important for interpretation here – samples are potentially less independent than they seem and it's relevant that you are using samples from various regions of the body. This also needs to be clear as some of your analyses are just on the F1 region.

Author response 74

Thank you for pointing this out. We agree and have revised the suggested text, see below.

Lines 108-115: "We used 2193 SEM images of skin to train the model, representing 204 individual sharks, sampled on 53 different body locations across 132 extant shark species, 2 ray species, and 24 fossil species. **Each SEM image represents a single body location from a single species and should be regarded as a unique entity, which means that skin samples across the body of the same species may fall into several morphogroups. We only include sexually mature specimens, as denticle morphology varies with ontogeny[14] (see Sample Collection in the methods section for more detail**, Supplementary Table S4, Supplementary Fig. S1)."

Comment 75

Lines 544 CFD methods: You need to include relevant basic details of the CFD tests here and also in the results. You need to provide information that answers the following questions: 1) What was the size at which models were produced? Did you scale denticle size at all? What

was the flow speed at which models were tested? Were all models tested at the same Reynolds number? How did you decide to choose your flow speed(s)?

Author Response 75

We kindly refer to **Author response 18a**.

Comment 76

Line 577: I think ration should be ratio

Author response 76

Yes, thank you. We have corrected this.

Comment 77

Lines 582-584: Here you use morphogroups to divide denticles into ridged vs not but Fig 1, category 5 shows SEM of

Author response 77

It looks as though the reviewer's comment is incomplete, but we guess that it is referring to the previously mentioned observation of denticles with subtle/no ridges which appear in SEM images from group 5 "Thick ridge denticles". We kindly refer to **Author Response 18h.**

Comment 78

Lines 587-588: This categorization of habitats needs to be explained more here – this is a very important part of your methods in this portion of your paper and you need to elaborate and describe how your states were categorized. It is not enough to just cite previous work.

Author response 78

We have accordingly elaborated:

Lines 758-761: "We categorized the binary states of deep versus shallow habitats, and reef versus non-reef habitats following **the categorization in Fishbase. We use the same vocabulary to describe these binary states, which have been defined based on depth and geographic location of capture[24,52,81]"**

Comment 79

Figure 1: for the UMAP analysis, this particular graph shows some separation between these 10 morphogroups, but there is also a large amount of overlap among many groups. I think you need to comment on this in the text when you first introduce this figure.

Author response 79

We have added further clarification, see below.

Lines 119-123: "The embeddings of all the input images were plotted in 2-D coordinates reduced from the 512-D morpho-space with the UMAP algorithm (https://umap-learn.readthedocs.io/en/latest/) for visualization only (Fig. 1A). **The UMAP in Fig. 1A thus shows a fragment of dimensions when illustrating the placement of morphogroups in the morpho-space."**

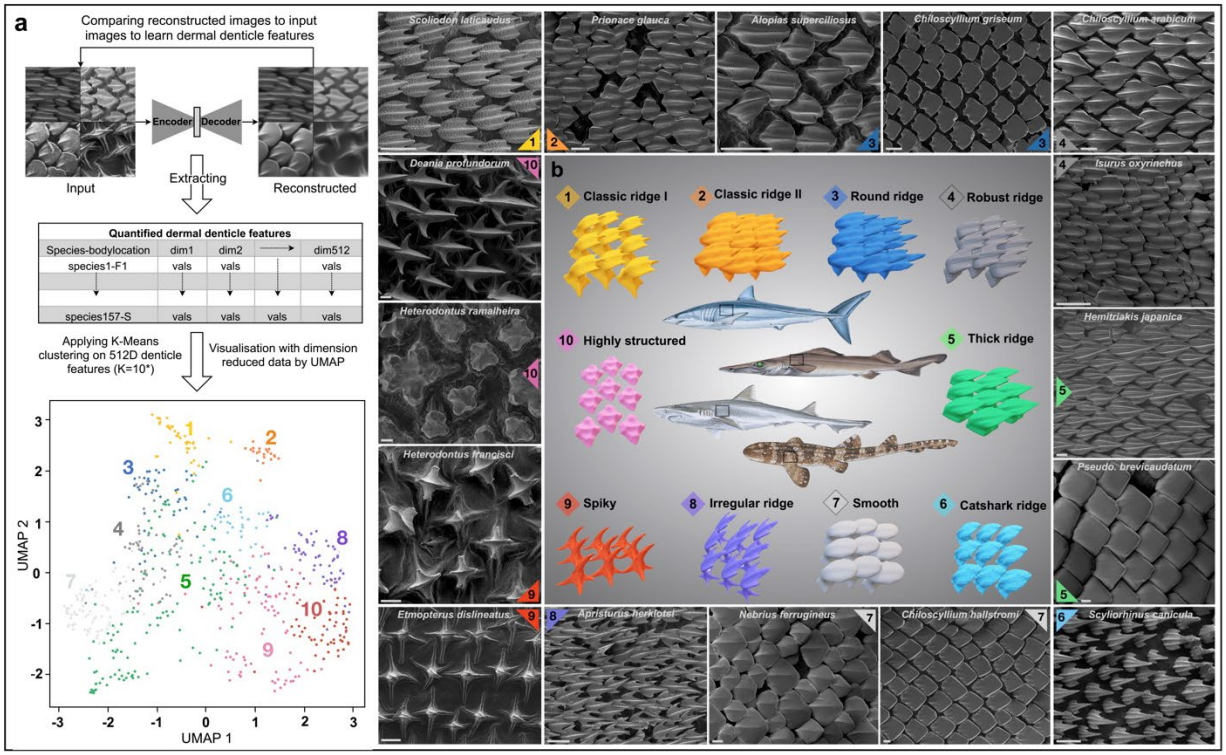
Comment 80

Fig. 1: Right now this figure has a lot of small text on it…this in theory will be okay with

digital versions with zoom but it would be good to make the little label triangles in the corners of your SEM images larger and more legible.

Author response 80

We have enlarged the numbers in the SEM triangles in Fig 1.



Comment 81

Fig 1: For morphotype 3, the round ridge…I'm guessing that round ridge refers to the denticle crowns being more rounded in morphology? But then the example of the 3d colored model in B has very long and distinct posterior spines or tines…aka it is not rounded. What is going on here? (might want a different name for this group?)

Author response 81

We appreciate the comment and refer to **Author response 24** where we introduce the individual variation within the different morphogroups. The names of the morphogroups reflect the average morphology, and we would very much like to keep this name.

Comment 82

Fig1. The morphogroup 5 aka ‘thick ridge’ also shows an example of a non-ridged denticle for morphogroup 5 in the SEM of Pseudoginglymostoma. Are not all denticles in this category ridged?

Author response 82

We kindly refer to **Author response 18h and 24**. Like all other ridged morphogroups, the ‘thick ridge’ morphogroup will display variations in denticle ridges. We have created Supplementary Fig S3 to visualize this variation.

Comment 83

Fig 1. Why does the UMAP graph here look different than the UMAP graph in Fig S8? Simply because there are different groupings?

Author response 83

Yes, exactly. Fig S10 depicts the original morphospace (shown in 2D) of the 20 groups, whereas the UMAP in Fig. 1 depicts the final 10 morphogroups on their own. To avoid confusion for our readers, we have added a clarification to the figure legend headline of Fig S10, see below (in underline).

"Fig. S10. Deep learning UMAP of the original 20 morphogroups. Circled groups are the 10 morphogroups kept for further analysis. Two of the groups (2+3, 7+8) are comprised of merged groups based on the lowest pairwise distances of group centroids (a dimensionless value of 1.6). The remaining eight groups were excluded as they were represented by less than three F1-body location denticle samples."

Comment 84

Fig. 2: I get that you ran BAMM and maybe some of the other comparative methods on this full 500+ species phylogeny but the ancestral state reconstruction shown here probably is not representative of this full phylogeny…especially given that is seems like you denticle morphotypes can often be different among close relatives. I guess what I'm saying is that do you really need this figure and should it be displaying the ancestral state reconstructon which is much more accurately (and legibly) portrayed in Fig 3 in the pruned phylogeny. My advice would be to get rid of Fig 2 and instead include a different figure instead (like maybe s2, s3, or a modified version of s10).

Author response 84

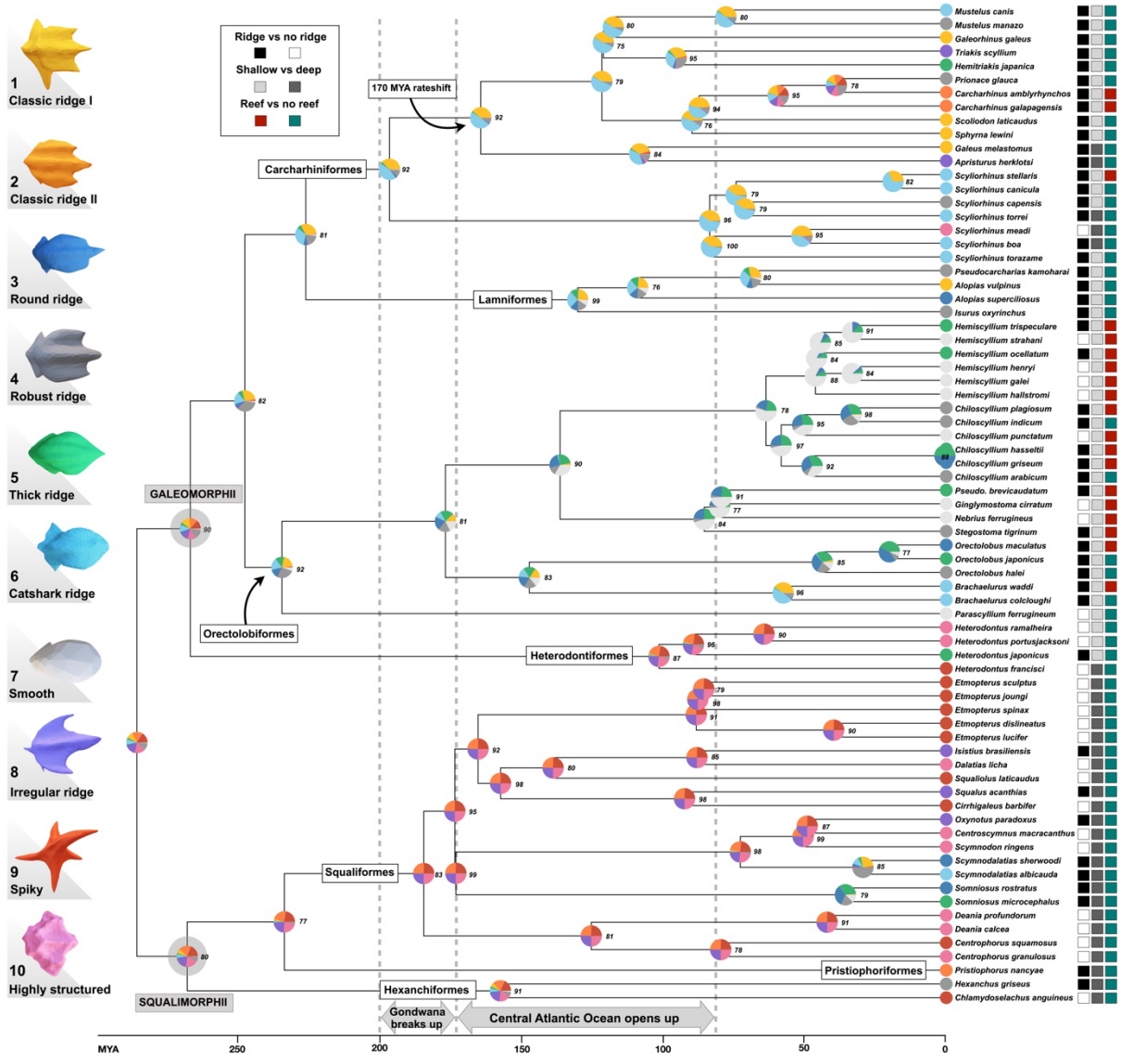
Thank you for your valuable suggestion regarding Figure 2. We agree that Figure 3 presents the ancestral state reconstruction results more clearly. However, Figure 2 serves as a more comprehensive phylogenetic tree that includes the diversification rate results. In addition, Figure 2 better illustrates the positions of our samples within the overall phylogeny, highlighting the strengths of our sampling (comprehensive coverage of different taxa) and areas where our samples are relatively deficient. Therefore, we would like to keep Figure 2 in the manuscript.

Comment 85

Fig 3: the figure legend here need to be much larger – currently it's not legible

Author response 85

We believe the reviewer is referring to the figure text, and not the legend, and we have thus increased the size of the figure text.

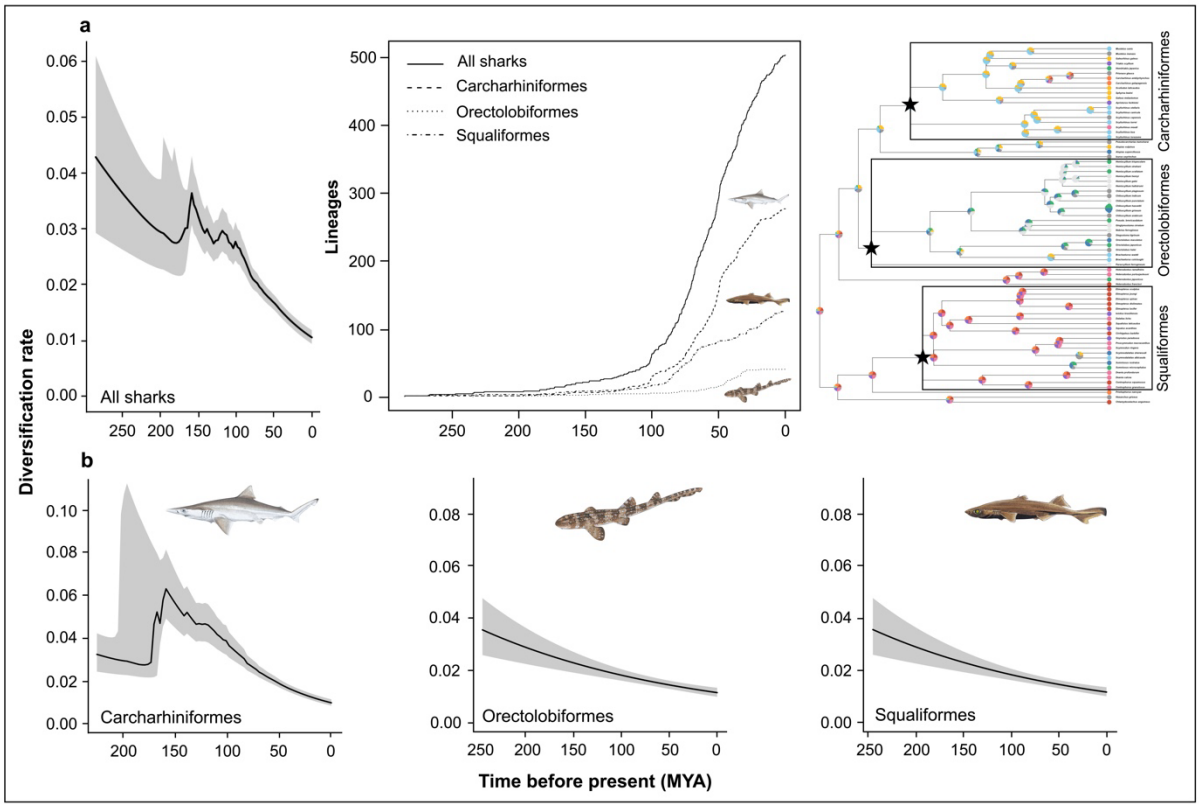


Comment 86

Fig 4: The lines in the lineages through time plot should be different colors or much different line types – the ones you have chosen are not easy to distinguish.

Author response 86

We have added little shark icons on the graph to better differentiate which lines represent which lineage. See below for reference (middle upper box).



Comment 87

Fig 5: what are all of the horizontal line on each plot in part A? If these are some representation of significant differences then you need to make that obvous and you also probably need to check your stats because these all likely have way too many significant groups given the number of boxplots. Please either explain these or replace these lines with something more easily interpretable.

Author response 87

We have added to the existing description of the error bars in the legend of Fig. 5, see below.

"…**horizontal bars in (A) are** error bars **displaying** significant differences between morphogroups with significance level $\boldsymbol{\alpha}$ = 0.05 (Tukey's HSD test)."

Comment 88

Fig 5: For the drag coefficients, what is the drag coefficient in your CFD setup of just a flat skin without features? Are you actually showing drag reduction compared to a flat plate or no?

Author Response 88

We appreciate the comments from the reviewer, and have allowed an explanation below. To summarize: **Since the purpose of this research is to compare the functionality of shark denticles between different shark species and parts of the body, the flat plate analysis will not affect the conclusion. To not distract the reader, we do not put the flat plate studies in the main manuscript.**

For the drag coefficient in our CFD setup, we have conducted a drag coefficient analysis for a flat plate with no-slip boundary condition. All other numerical settings are consistent with the shark denticle cases employed. The analysis reveals the drag coefficient of approximately 4.5×10^{-3}, consistent with the literature value of 4.33×10^{-3} (see Munson et al. 2013: B. R. Munson, T. H. Okiishi, W. W. Huebsch and A. P. Rothmayer, "Fluid mechanics," 2013.).

In the flat plate analysis, we believe such a small drag coefficient is mainly due to a streamlined laminar flow feature gone through the smooth surface of the plate.

As suggested by the reviewer, we originally expected to see a higher drag coefficient than the denticle cases, and we observed that it is the opposite. This seemingly counterintuitive outcome could result from several reasons. Firstly, the complex interaction between the flow and the irregular surface of shark denticles might introduce localized turbulence or alter flow patterns (see B. Dean and B. Bhushan, "Shark-skin surfaces for fluid-drag reduction in turbulent flow: a review," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, pp. 4775-4806, 2010.). This influences the overall drag characteristics differently than expected. Secondly, the specific geometric features of the shark denticles could potentially induce flow separation or other effects contributing to increased drag despite the observed drag reduction at certain Reynolds numbers. Lastly, as the flat plate is modelled to be perfectly smooth, we hypothesize that the absence of roughness might lead to this counterintuitive outcome, suggesting that shark denticles might possess superior drag reduction capabilities even under conditions of similar smoothness.

Comment 89

Fig 6. You need an explanation somewhere of how you are getting these functional axes plotted back onto this PCA morphospace of your denticle data. In the current reading of your paper this is not obvious or discussed enough.

Author response 89

We have **added a new supplementary table** (Table S14), which is uploaded to Figshare (https://figshare.com/s/ae0a3f288287d5afedc8), with the raw input data for the functional

trait analysis (Fig 6). We have **added reference to this table in the methods, and the results**. See below:

Results:

Lines 258-260: "We then used functional trait analysis (**similar to a PCA**) to integrate the mechanical properties, mineral composition, and fluid dynamic properties of the various denticle morphogroups (**see methods section for further detail**, Fig. 6, **Supplementary Table 14**).

Methods:

Lines 815-818. "We performed PCA on our trait values (stiffness, hardness, mineral protection, damage resistance and drag reduction, see raw data in Supplementary **Table S14**) using the function dudi.pca in ade4 R package[86]."

Comment 90

Fig 6. What do the red and black lines refer to in panel A? I'm guessing one must be increasing in performance and the other is decreasing? Why not simplify things and just plot the increasing performance arrows only?

Author response 90

We have revised the legend of Fig. 6, see below. We believe this way of illustrating it makes it easy to view the different relationships. This was similarly illustrated in the reference we refer to for the functional trait analysis, see below.

Fig. 6. Spectrum of denticle form and function. (**a**) Projection of functional variables (dots) defined by principal component axes (PC) 1 and 2, with PC1 describing hardness and stiffness and PC2 damage resistance, drag reduction and stiffness. Arrows demonstrate weighing and direction of the five variables. **The direction of the red arrows indicates higher values for that function, while the blue arrows represent lower values.** […]"

33) Diaz, S. *et al.* The global spectrum of plant form and function. *Nature* **529**, 167-171, doi:10.1038/nature16489 (2016).

Comment 91

Fig S3: Panel F shows a model that has no surface ridges but then the CT scan shows clear ridges. This seems like a major oversight given that much of your paper is about ridges vs non-ridged denticles.

Author response 91

We kindly refer to **Author responses 18h, 23 and 24.**

Comment 92

Fig s10: For me the most interesting and potentially novel part of your paper is the functional interpretations and findings across different shark species (morphotypes). The correlation analyses are mentioned multiple times in this section and include some interesting results but aren't shown. When I went looking for them here in Fig S10 I was confused because these graphs in panels C and D do not show correlations clearly at all. Please make some graphs (scatterplots) that plot hardness vs reduced modulus and fluorine vs damage resistance and include them in a figure or supplementary figure (a figure in the main text would be better). If

you want to split them by morphotype to plot then you can color the points or make a correlation or scatterplot matrix. But really here the points should be species which are your morphotypes in this case (as I understand it, you sampled one species per morphotype for this analysis?).

Author response 92

We acknowledge that the term "**correlation**" in Supplementary Fig. S10 (now Supplementary Fig. S12) may have caused confusion. We have revised this terminology to "**comparison**" to accurately reflect the purpose of the figure.

Supplementary Fig. S12 serves to compare hardness and modulus, as well as fluorine content and damage resistance. We aimed to demonstrate the similarity in trends between these pairs of properties. Typically, hardness and modulus exhibit a positive relationship, which is evident in the graph. This coherence underscores the integrity of our mechanical property data.

Furthermore, considering the anticipated mechanical advantages of fluorine content, we examined the relationship between fluorine wt% and damage resistance. Our observations revealed a positive trend in most of the morphogroups between these properties (Supplementary Fig. S12D). In addition, we have conducted a correlation matrix analysis that clearly shows a positive relation between mechanical properties and most elemental compositions (Fig. 6D).

Additionally, we have also added scatterplots of hardness vs reduced modulus and fluorine vs damage resistance in the Fig. 6E-F.

Lastly, to clarify our sampling approach, we want to emphasize that we sampled 2 to 4 species and 1 to 5 body locations per species within each morphogroup to calculate mechanical properties and compositions. The detailed information regarding our sampling is provided in Supplementary Table S1 in the Supplementary Data accompanying the manuscript.

Updated version of Fig. 6:

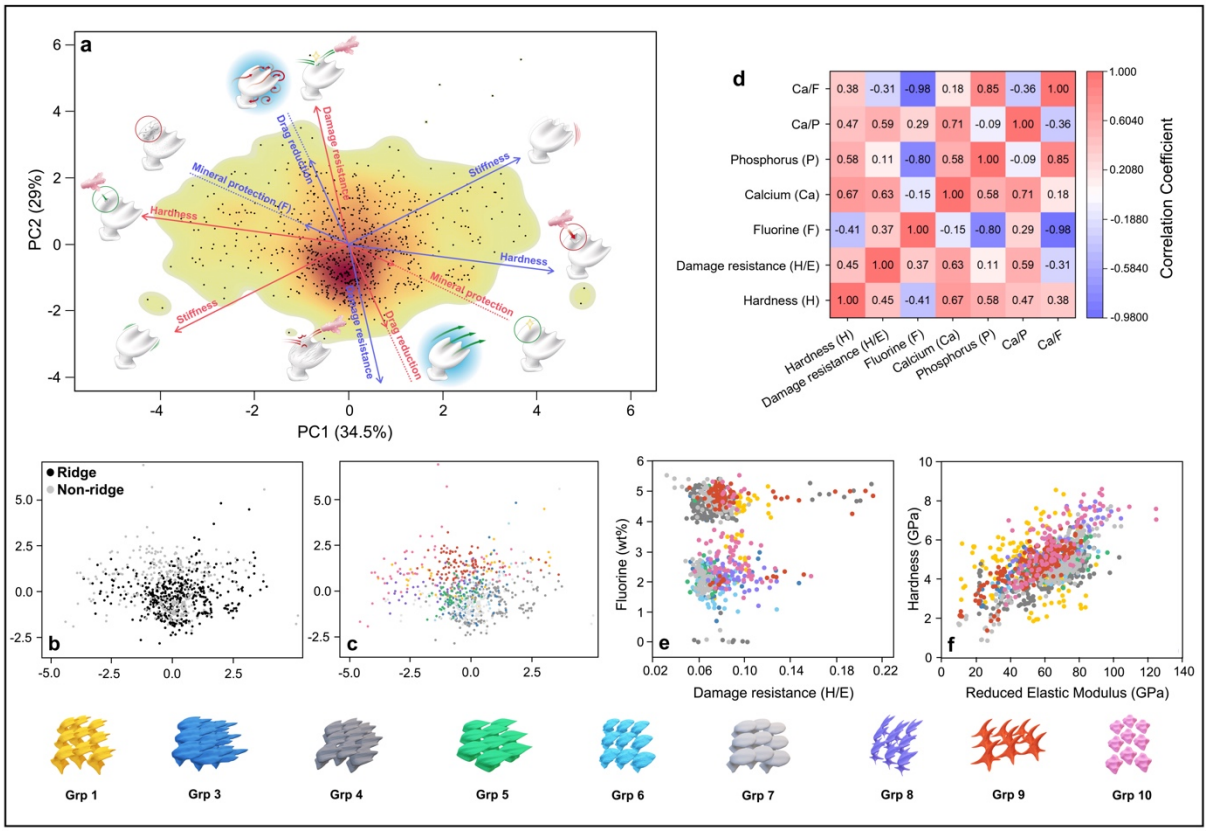


Fig. 6. Spectrum of denticle form and function. (**a**) Projection of functional variables (dots) defined by principal component axes (PC) 1 and 2, with PC1 describing hardness and stiffness and PC2 damage resistance, drag reduction and stiffness. Arrows demonstrate weighing and direction of the five variables. The direction of the red arrows indicates higher values for that function, while the blue arrows represent lower values. The colour gradient illustrates zones of high (red) and low (yellow) occurrence probability of functions in the trait space defined by the PC axes. See Methods for calculations of PC1 and PC2. (**b**) Location of the **ridged (groups 1,3-6, 8) vs. non-ridged (groups 7, 9-10) morphogroups, and (c) all** nine morphogroups in the spectrum. **(d) Heat map of correlation matrix displaying the correlation between hardness, damage resistance, and elemental composition. Inset numbers are correlation coefficients. Pearson correlation was adopted to calculate the coefficients. Scatter plot of**

(e) hardness vs reduced elastic modulus and (f) fluorine content vs damage resistance. Mineral Protection and Fluorine (F) represent the same function in (a) and (d), respectively.

Responses to reviewers' comments

Reviewer #2 (Remarks to the Author):

My concerns have been addressed. I now understand the authors' view on Comment 9.

I appreciate the opportunity for reviewing this manuscript.

Response to remark from Reviewer #2

Thank you very much for going through the manuscript once again, and for accepting our responses to your previous review.

Reviewer #3 (Remarks to the Author):

I acknowledge the authors for the responses and revision of the manuscript. I read through other reviewers' comments and the authors' responses. The responses and revision have addressed most of my questions and concerns. Yet, I still have questions / concerns. I hope the authors can address them to improve the manuscript. Before presenting concerns / questions, I summarize the work below.

Summary

The paper analyzes dermal denticles of sharks using developed quantitative methods to better understand the impact of dermal denticles on how sharks adapt to various environments. The paper adopts nanoindentation, electron probe micro-analysis, and computational fluid dynamics to establish mechanical properties, elemental composition and swimming performance. Concretely, the paper uses deep learning to learn visual features and clustering them into a predefined number of groups. It further aligns clustering results with the clades of the phylogenetic tree of sharks. Based on these clusters of dermal, the paper conducts a series of mechanical, elemental, and fluid dynamics experiments to investigate the characteristics of

these different denticle morphologies. Results lead to multiple conclusions, e.g., denticle ridges are highly variable and present more functional properties beyond drag reduction. Moreover, it suggests a radiation event within the ground sharks around 170MYA, which results in the enormous diversity of the clades where the ridged dermal denticles may be the key.

The paper is generally written well. It has certain merits and contains clear documentations and important details. As a non-expert in life sciences, I enjoy reading this paper while I can't confidently evaluate the significance in the respective discipline. So I leave this evaluation to other expert reviewers. As a research from Computer Science, I think the technical parts (e.g., self-supervised learning by VSC, clustering by K-means) make sense. Yet, there are missing details and inappropriate statements.

Comment 1

In responding to my previous question why the authors choose Variational Sparse Coding (VSC) over Variational Autoencoder (VAE), the authors insert new lines as below. However, the inserted writing mentions "latent vectors" without defining what they are. I don't think readers with little machine learning or deep learning background can follow. So I suggest the authors revise this part to improve the readability.

"We prefer the VSC as it introduces a regularization term in the loss function, which often constraints the latent vectors to approximate a standard normal distribution to ensure that the learned representation does not diverge excessively in pursuit of high similarity in image reconstruction in the training process. Excessively dispersed latent vectors could lead to issues where similar data points, i.e., similar appearances of dermal denticles, do not map to nearby points in the latent space, making the analysing and interpreting of the latent space

difficult or unmeaningful."

Author response 1

We fully understand the need to explain latent vectors, and we have revised the section (see below, text in **bold** has been added).

Lines 503-513: ”We prefer the VSC because it introduces a **stronger** regularization term in the loss function, which constrains the latent vectors **(the underlying numerical representations that encode the essential features of the input data)** to approximate **a sparse (spike-and-slab)** distribution, rather than a normal distribution. **The sparse distribution regularization helps improve dimension independence and prevents excessive divergence while still achieving high similarity in image reconstruction. If the latent vectors become too dispersed, similar data points—such as dermal denticles with similar appearances—might not map to nearby points in the lower-dimensional latent space, where the latent vectors are positioned to capture the core structure of the original feature space formed by the raw features of the input data. This would make it difficult, or even meaningless, to effectively analyze and interpret the latent space.”**

Comment 2

In responding to my previous concern about whether "the clustering results actually do not align with the phylogenetic tree and hence require some manual post-processing". The authors added paragraphs in which there are two statements below. However, in my opinion, it does not need to emphasize "in an entirely objective manner" because there is a clear manual intervention as post-process.

"Following the selection of K-means as our clustering algorithm, we conducted the clustering process in an entirely objective manner, without incorporating phylogenetic information."

"we adopted the k-means clustering result of 20 clusters as our final configuration. Out of the 20 morphogroups, we excluded eight clusters that were represented by fewer than three F1 body position denticle samples to ensure a robust sample size for the subsequent ancestral state reconstruction."

Author response 2

We understand the reviewer's concern and have complied fully. See corrections to the text below.

Lines 575-576 "Following the selection of K-means as our clustering algorithm, we conducted the clustering process ~~in an entirely objective manner,~~ without incorporating phylogenetic information."

Comment 3

In responding to my previous question why "Why should fossil species be included in the training data along with extant ones", the authors explain and state that "We decided not to go further with the fossil groupings, as the results may have been biased by the images themselves." Does it mean that the authors do not use the clustering results to analyze fossil specimens? But the revised manuscript still states in Line 109 that "We used 2193 SEM images of skin to train the model, representing 204 individual sharks, sampled on 53 different body locations across 132 extant shark species, 2 ray species, and 24 fossil species." This is confusing to me. If clustering results on fossils are not informative, authors can either (1) re-

do clustering and the experiments without fossils, or (2) sincerely discuss the limitation of the current setup. Authors can clarify further.

Author response 3

To answer this question, we have created a UMAP highlighting the fossil samples (red dots) among the extant samples (blue dots). This shows that most fossil samples (108 of 112) cluster together, adding a new group into the feature space. They do not affect the current results, i.e., the groupings based on extant samples.

We have introduced the figure below into the Supplementary Information as Fig. 1, and referenced it in the Results.

Result Lines 106-110 "We used 2193 SEM images of skin to train the model, representing 204 individual sharks, sampled on 53 different body locations across 132 extant shark species, 2 ray species, and 24 fossil species. **The fossil samples are not among the final groupings, as they clustered together as their own group. Excluding them did not affect the current result (Supplementary Fig. 1)**.".

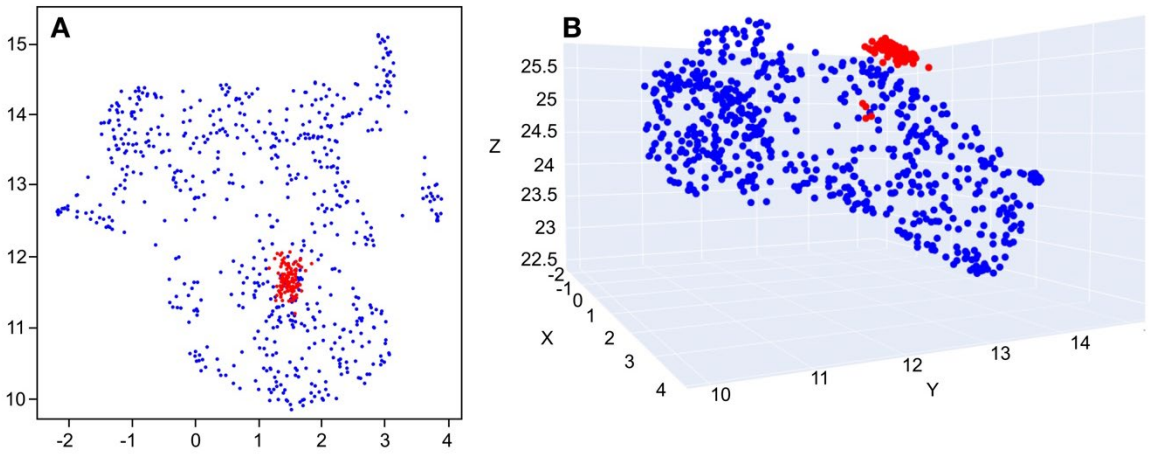


Fig. 1. UMAP of fossil samples among extant samples. UMAP of dermal denticles in the feature space after grouping, including both extant (blue dots) and fossil (red dots) samples.

(**A**) corresponds to the UMAP in Fig. 1A, slightly rotated, and (**B**) shows the fossil cluster in a different angle of the UMAP. Fossil samples cluster as an individual group and do not affect the current grouping results.

Reviewer #4 (Remarks to the Author):

Thank you for your detailed response to comments and your additions and changes made to this manuscript. I still think this is an impressive paper due in large part to the multiple types of data collected and the large datasets compared to what is typical for similar work on sharks. I think many things that were not clear previously have been made clear and you have generally made many improvements to this work. The multifaceted nature of this project does continue to make this paper hard to interpret and evaluate, but many of the details you have added help clarify your research for readers. However, I still feel as though there are some large issues to be address now that some of the methods and results are more clear.

I appreciate your more detailed explanation of the CFD methods that you used. I think that these methods seem biologically unrealistic for a few reasons:

Comment 4

An inflow of 30 m/s to achieve flows over the denticles that are 15-20 m/s seem biologically unreasonable. There are few (if any) sharks that can achieve these kinds of speeds and only then for a moment. Testing at these speeds does not seem biologically relevant. Please justify your choices in the manuscript or consider changing your testing protocol.

Author response 4

We acknowledge that the inflow speed of 30 m/s is higher than those typically observed in most shark species. However, the purpose of using this speed was to investigate extreme flow conditions, which could maximize the hydrodynamic effects of the denticles and provide insights into potential fluid dynamic benefits at higher velocities. The 30 m/s inflow speed ensures that the flow patterns over the denticles transition through laminar, vortex, and potential turbulence phases at various Reynolds numbers. This approach allows us to measure the drag coefficient in a steady state, which aligns more accurately with its theoretical definition. Moreover, theoretically, the drag coefficient is independent of inflow speed once the flow reaches a steady state. Thus, changing the inflow speed to 15-20 m/s would not affect the conclusions of this study, as the drag coefficient primarily depends on the geometry of the body. From a numerical perspective, we selected 30 m/s (the peak speed of some sharks) as a characteristic speed to measure the drag coefficient.

Corrections to the text in added in **bold**, below.

Methods: Lines 737-744 ”This corresponds to 30 m/s (~108 km/h), which ensures a flow speed of 15-20 m/s (~54-72 km/h) over the dermal denticles, as it decreases due to water viscosity and the sharkskin surface no-slip condition. **The high flow speed used does not reflect the actual swimming speed of the sharks involved in the study, but** from a numerical standpoint, a high inflow speed is advantageous for understanding the flow pattern, as the drag force under low flow speed (low Reynolds numbers) will lead to laminar flow and thus minimal drag conditions. **We emphasize that these flow conditions are selected based on the examination of the denticle geometry rather than the whole body of the sharks**.”

Comment 5

You tested your different models under the same flow conditions, but also used real denticles sizes based on CT scans for the models. I am sure there are body size differences between the sharks you sampled, so this is the same as saying that all your sharks are swimming at the same speed, regardless of differences in sizes. This means the biological interpretation for these comparisons is different across different species. For example, assuming a 20 m/s flow speed from your testing, this means 20 body-lengths/second for a 1 m shark versus 5 body-lengths/second for a 4 m shark. Another way of thinking of this same issue is that when discussing swimming speeds in sharks and fishes across species, we often standardize speed by body length and use body-lengths/second. In part this is because most species have routine, cruising swimming speeds of around 1-2 body-lengths/second, and correcting speed by body-length in this way helps allow comparisons across different sizes. Your CFD test is then effectively having some species go much faster in body-length relative units compared to others, making this comparison not as relevant to biological reality. Again, please justify these choices or consider changing your protocol.

Author response 5

This is a critical point, and we appreciate the observation. However, this concern is more relevant if we were measuring the drag coefficient of the entire shark, where the overall geometry (including fin size and shape) would play a role. In our study, we are focusing on the effect of denticle geometry, which is quite small in scale. Therefore, to eliminate the size effect, we standardized the inflow speed across all models to focus solely on the denticle geometry's influence. We agree that standardizing the speed by body length (body-lengths/second) is a valid approach when measuring performance across entire shark

specimens, but in this case, our goal is to isolate the other effects and just focus on denticle geometry.

Methods: Lines 753-757: ”Then, for the comparison between **the denticle geometry of** different shark species, we maintain a constant Reynolds number of 1×10^4 for all cases using the specified CFD setup, as shown in Supplementary Note 3.4.”

Comment 6

Lastly, I don’t really see a developed boundary layer in the setup of your CFD tests and this general testing configuration doesn’t seem particularly realistic – where a small patch of denticles are subject to high incoming flow with nothing upstream or downstream. The boundary layer seems exceptionally thin compared to denticle sizes (perhaps also due to the high speed flow). I grant that it may be overly difficult to get closer to biological reality, but the apparent lack of a developed boundary layer above the denticles seems different than what is likely happening on real sharks. If these physics do not match expectations for what might be happening on real shark, then how should we interpret your results?

Author response 6

We appreciate your comment regarding the boundary layer in our CFD simulations. The boundary layer is indeed thinner than what would be expected in natural settings, mainly due to the high inflow speeds. We have added an explanation of this in the Supplementary Information to avoid confusion about the thin boundary layer of our samples.

Supplementary Information: Lines 1059-1069: “The result is similar to the drag analysis for flat plate in the literature[103]: the attached flow (Stokes flow) and steady separated flow transit

to separated unsteady flow with laminar flow boundary layer upstream of the separation, and possible vortex street. Same trends are found in other types of shark denticles**. The primary aim of our setup was to study the direct interaction of flow with the denticles, rather than fully replicate the complex boundary layer behavior seen in nature. The relatively undeveloped boundary layer may also result from the smaller size of the denticle matrix in comparison to the larger domain required to produce a fully developed boundary layer. Given equipment and computational constraints, reproducing such a setup is currently not feasible. However, we believe that this limitation does not affect the overall trend observed in the drag coefficient, which is primarily driven by the denticle geometry itself.**"

Comment 7

In part d) of my first comment (your Comment and Author response 18), I asked about how you dealt with variation due to body size, which is always present in morphological datasets across species. However, in your response you told me to look at a different response above that is all about your process of k-means clustering. Perhaps this is a mistake, but I think this is a very significant issue and deserves a response as well as clarification in the manuscript. Based on your response elsewhere, it seems you have not corrected anything based on body size because you state that all of your samples are from sexually mature individuals (which is also difficult to determine and prove, but that is a separate issue). It is very uncommon to choose not to correct for size in some way in a comparative morphological dataset – I strongly suspect that variation due to body size is a large part of the morphological variation in your study simply because it typically is for most morphological characters. If you have any information on body size, you should be able to look at this relatively easily to verify or deny this prediction. It is also possible that variation due to body size is not such a large

component of variability here, but regardless, some justification needs to be provided for your methods in this case and it may be necessary to see how body size is impacting your data if you have fork length or other size information. In addition, even if all the sharks were sexually mature, there is still a large range of sizes for mature sharks even within the same species (not to mention between species).

Author response 7

We apologize if we referred to a wrong comment in the previous response letter, that must have been a mistake. To comply with this comment, we have analyzed the data we have on total body length and sex (M/F) to see if body size affects the grouping results.

We made a violin plot to see the body length distribution within each group and a plot of the significant differences between the groups (Supplementary Fig. 6. A and B, respectively, see below). This new Figure can be found in the Supplementary Information. A new table, Supplementary Table 12, with body length information has been uploaded to Figshare.

Further, we have described the statistical tests and results thereof in the Methods section on Lines 462-480:

"**We tested the effect of body size on morphological variation (Supplementary Fig. 6, Supplementary Table 12). The ANOVA showed a significant mean body length difference between the groups (F=33.182, p=0.000). The Tukey's Honest Significant Difference test then showed us where those differences lay. To explore the role of body length in the grouping, we fitted both a linear (logistic regression) and a non-linear (random forest) model to predict groups, using body length and sex as predictors. The data were split into a 70:30 train-test ratio. The logistic regression model showed relatively poor performance, with a mean accuracy of 28.5%, a mean F1-score of 0.16,**

and a McFadden's pseudo R2 of 0.062, suggesting only a modest improvement of 6.2% over the null model. Additionally, the absolute coefficients of body length and sex were similar in magnitude (ANOVA F=1.97, p=0.099), indicating that neither feature had a noticeably greater contribution compared to the other in predicting groups. The random forest classifier performed better, with a mean accuracy of 54.5%, a mean F1-score of 0.55, and a McFadden's pseudo R2 of 0.142, reflecting a 14.2% improvement over the null model. Although body length was assigned a high importance score of 97.7%, the overall model performance suggests that its contribution might not be as impactful as the importance score implies. The moderate accuracy indicates that body length alone may not sufficiently explain group differences, even in a non-linear context. Based on these findings, while body length appears to have some relevance, particularly in non-linear models like the random forest, its overall impact on grouping may be limited."

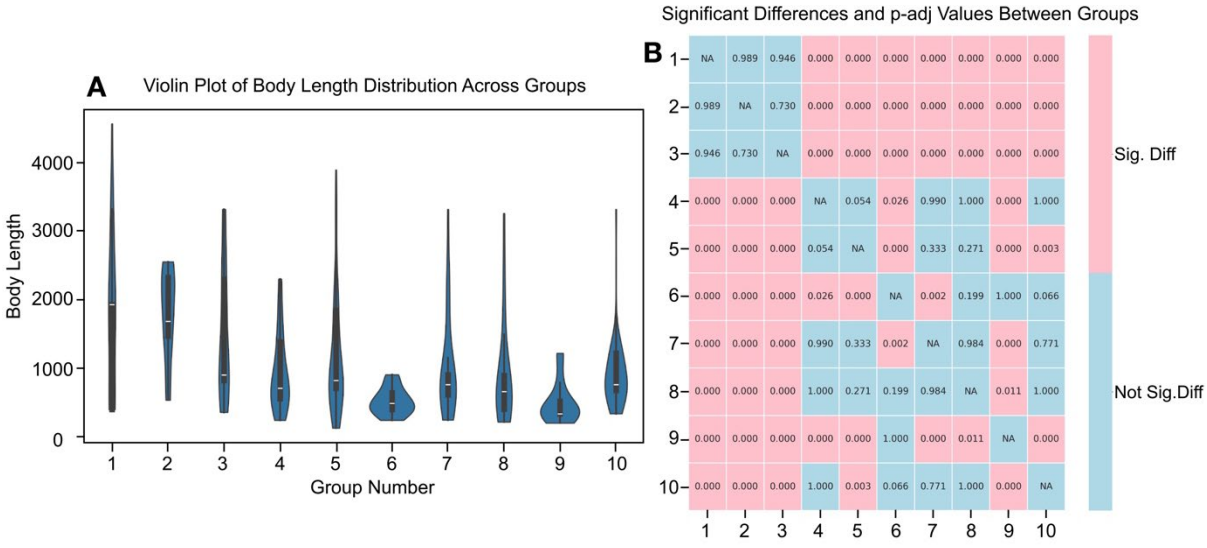


Fig. 6. Effect of body size on denticle morphogroups. (**A**) Violin plot of body length between the ten morphogroups shows a difference between groups (ANOVA F=33.182, p=0.000). However, due to poor accuracy in the logistic regression model (28.5%) and the

random forest classifier (54.5%) reflecting a 6.2 and 14.2% improvement over the null model, respectively, body length may have a limited impact on grouping. (**B**) A plot of significant differences and p-adj values between morphogroups based on the Tukey's Honest significant difference test. See Sample Collection in the Methods for more detail.

Comment 8

After learning more about your decisions on how to choose ten morphogroups and how your deep learning approach created and defines these morphogroups, a few things seem like they are still potentially issues. First, your descriptions for how you decided on 10 clusters/morphogroups is very helpful but none of it sounds particularly 'objective' despite you referring to it as such. I'm not an expert on these methods but my reading of this description generally gives me the impression that you did some filtering based on clusters that were small and then some that were close together (all of which has an arbitrary or subjective feel to it). Then you also looked at the results on the phylogeny and decided that ten was just about the same as twelve groups (again, kind of arbitrary feeling). I think it is great to explain all of this to readers but at the very least I would not describe this as an objective process.

Author response 8

This was also a concern for Reviewer #3 in Comment 2. We understand the concern from both reviewers, and we have decided to rephrase this part of the manuscript to suggest a less objective approach. See corrections below, (~~deleted~~, **added**).

Abstract: Lines 29-31: “For the first time, we ~~objectively~~ quantify dermal denticle disparity using deep learning on scanning electron microscopy images, discerning ten morphogroups across modern sharks (Selachii).”

Introduction: Lines 70-71 “This study ~~objectively~~ defined ten dermal denticle morphogroups across 132 extant shark species.”

Results: Lines 102-103 “We wanted to ~~objectively~~ categorize dermal denticles based on key morphological features **as objectively as possible**, and for this we used the Variational Sparse Coding model (VSC),[…]”

Discussion: Lines 330-331 “By utilizing deep learning methods, we were able to ~~objectively~~ classify the complex dermal denticles of sharks into ten major groups based on morphological similarities.”

Comment 9

Second, I deeply appreciate you posting your code in case someone wants to recreate your analyses, but it seems to me that these morphogroups you define may be nearly impossible for anyone else to use without data outside of your own. If another author wanted to also characterize their own data on some different species or body regions using your morphogroup framework, it seems like they would need to fully recreate your analyses (including all your data) and then assign their new data to one of above morphogroups. I tend to think this is asking a lot and it seems unlikely other authors will be able to utilize your contributions in this way. The current denticle literature very often uses morphotypes (same idea as morphogroups) based on work by Reif and I believe these morphotypes have been so

pervasive because he provided clear descriptions for determining denticle morphotype and connected these types to function using logical arguments. If other researchers have difficulty in making use of the morphogroups you have defined here, it seems unlikely that your morphogroups will be able to supplant the status quo.

I also believe that the methods you use to categorize morphogroups are very interesting, but that it is unfortunate you are unable to understand what features are most important to define and separate morphogroups and that there is substantial morphological variation in some morphogroups. The addition of Figure S3 is excellent and really demonstrates that some morphogroups have some shared features but in many of these groups it is very difficult to describe shared features of all the provided images. In many cases it therefore seems unpredictable why a particular image was assigned to one morphogroup instead of another. For me these issues lower my confidence in these results.

The variation within some of the morphogroups makes it difficult to have confidence that using a few representatives from a morphogroup is a good model for all the species with a morphogroup. This makes me more cautious in assuming your results on various denticle functions will be generalizable for a given morphogroup.

Author response 9

We understand this concern, and we agree with the view that our approach is comprehensive and less practical for other studies to replicate. The groupings made by Reif are often used as a framework for understanding denticle morphology. Reif's comprehensive image data were incorporated in our own grouping analysis, and we can thus describe the similarities with Reif's groupings to our own. We suggest a user-friendly guide to determine denticle

morphology, where we present Reif's groupings amongst ours so that others can navigate our groupings with more ease without further analysis. We have inserted a description of this in the Supplementary Information as Note 4, and we created a figure displaying this in the Supplementary, Fig. 9, and we have uploaded a novel Supplementary Table 15 to Figshare with the data used to create Supplementary Fig. 9 (see below).

We refer to Supplementary Note 4 in the Results section on lines 1096-1158:
"These findings suggest that the diverse denticle morphologies contribute to the adaptation of sharks to varying environments. **Our results follow the framework by Reif (1982) relating form and ecological function of shark dermal denticles (see Supplementary Note 4, Supplementary Fig. 9).** Consequently, we provide a detailed account of the relationships between the denticle morphology, function, and ecological niche of different shark clades in the following sections."

Lines 1096-1158: "***How to use the denticle morphogroups***
It is useful to have a frame of reference when studying denticle morphological. The most common reference was established by Reif (1982), distinguishing five categories of denticles named after their supposed ecological function: 1) generalized function, 2) abrasion strength (knobs), 3) defense (spines and mucus), 4) drag reduction (groove pattern), and 5) luminescence (scales and photophores)[20]. In order to make our suggested morphogroups applicable for others, we checked the shark taxa included by Reif against the F1-body location samples in our morphogroup dataset (see Supplementary Table 15 and Supplementary Fig. 2)[21]. We attached the morphogroups next to their associated taxon in a modified figure from Reif (1982) (Supplementary Fig. 9)[20].

The generalized function describes denticles with long ridges on the crown and well-developed lateral wings, represented by the *Mustelus* genus (smooth-hounds), and the families Scyliorhinidae (catsharks) and Hexanchidae (cow sharks). These taxa express morphogroups 3 (round ridge) 4 (robust ridge), 5 (thick ridge), 6 (catshark ridge) and 10 (highly structured). Protection against abrasions consists of two denticle types, one with large, thickened and smooth crowns, the other with large, thickened and ornamented crowns. The first includes most demersal sharks and fits the morphological description of our morphogroup 7 (smooth), whereas the other include taxa such as *Centrophorus* (gulper sharks), *Heterodontus* (bullhead sharks), and Orectolobidae (wobbegongs), expressing morphogroups 8 (irregular ridge), 9 (spiky) and 10 (highly structured). The function of defense against ectoparasites and epibionts display spine-like denticles to make attachment difficult. Reif named taxa such as *Somniosus* (sleeper sharks), *Deania* (deepwater dogfish), *Oxynotus* (rough sharks) and *Squalus* (dogfish), which can be found in our morphogroups 3 (round ridge), 5 (thick ridge), 8 (irregular ridge) and 10 (highly structured). The function of drag reduction always includes a groove pattern of parallel ridges on the denticle crown. Reif used two categories describing the distance between the ridges as either <80μ or >80μ. However, since we made no such distinction, we will treat it as one category. Reif included *Isurus* (mako sharks), *Sphyrna* (Hammerhead sharks), *Carcharhinus* (ground sharks), *Galeocerdo* (tiger shark), *Alopias* (thresher sharks), *Prionace* (blue shark), *Galeorhinus* (school shark), which we found in our morphogroups 1 (classic ridge 1), 2 (classic ridge 2), 4 (robust ridge), 5 (thick ridge), 8 (irregular ridge). Lastly, the function of luminescence describes denticle as either long, slender spikes (bristle or hook-like) with photophores dispersed in between, or large, thick denticles (concave or thorn-like) with photophores positioned on top. This include *Etmopterus* (lantern sharks), *Dalatias* (kitefin shark)

Chlamydoselachus (frill sharks), which are found among morphogroups 9 (spiky) and 10 (highly structured). However, species in this category must also carry photophores.

The morphogroups present under the Generalized Function mostly display ridged denticle crowns (grps 3-6), and a single without (grp 10), however, they vary in their functional properties (Fig. 5D). Besides all scoring high on drag reduction, associated with the ridged morphology, round ridge denticles (grp 3) has a more general functional profile scoring equally in hardness, stiffness, damage resistance, and mineral protection. Robust ridge denticles (grp 4) have a high mineral protection, while thick ridge denticles (grp 5) have a high stiffness. Highly structured denticles (grp 10) score either medium or high in all categories, also presenting a more general functional profile. From this, we can surmise that a general function for dermal denticles is the ability to reduce drag while additional functions depend on lifestyle. Abrasion Strength is described as an external feature by Reif, i.e., that denticles are thicker or more robust and can thus withstand abrasions, however, we measured the damage resistance of the internal denticle structure, which is a different measurement. This is reflected in the morphogroups. Spiky (grp 9) and highly structured (grp 10) have a relatively high damage resistance, however, smooth (grp 7), round ridge (grp 3), robust ridge (grp 4), and thick ridge denticles (grp 5) have a low damage resistance. For now, we can only assume that external morphology plays a role in abrasion strength, especially in smooth, robust ridge, and thick ridge denticles, all displaying a robust morphology. For the function of Defense with spine-like denticles, we do not observe obvious spine-like morphologies in the morphogroups (grps 3, 5, 8, 10). However, denticles with ridges also display cusps and cusplets, which can vary in morphology with some being spine-like. Here, we refer to the Supplementary Fig. 3 to show the morphological variation in each morphogroup. The function of Drag Reduction, on the other hand, displays six out of

seven (grps 1-5, 8) morphogroups with the ridged, or grooved, morphology. These all score high in drag reduction and belong to fast swimmers and highly migratory sharks. Lastly, the function of Luminescence is also in accordance with the expected morphology by Reif. These are either bristle or hook-like (grp 9), and concave or thorn-like (grp 10), each accommodating the presence of photophores."

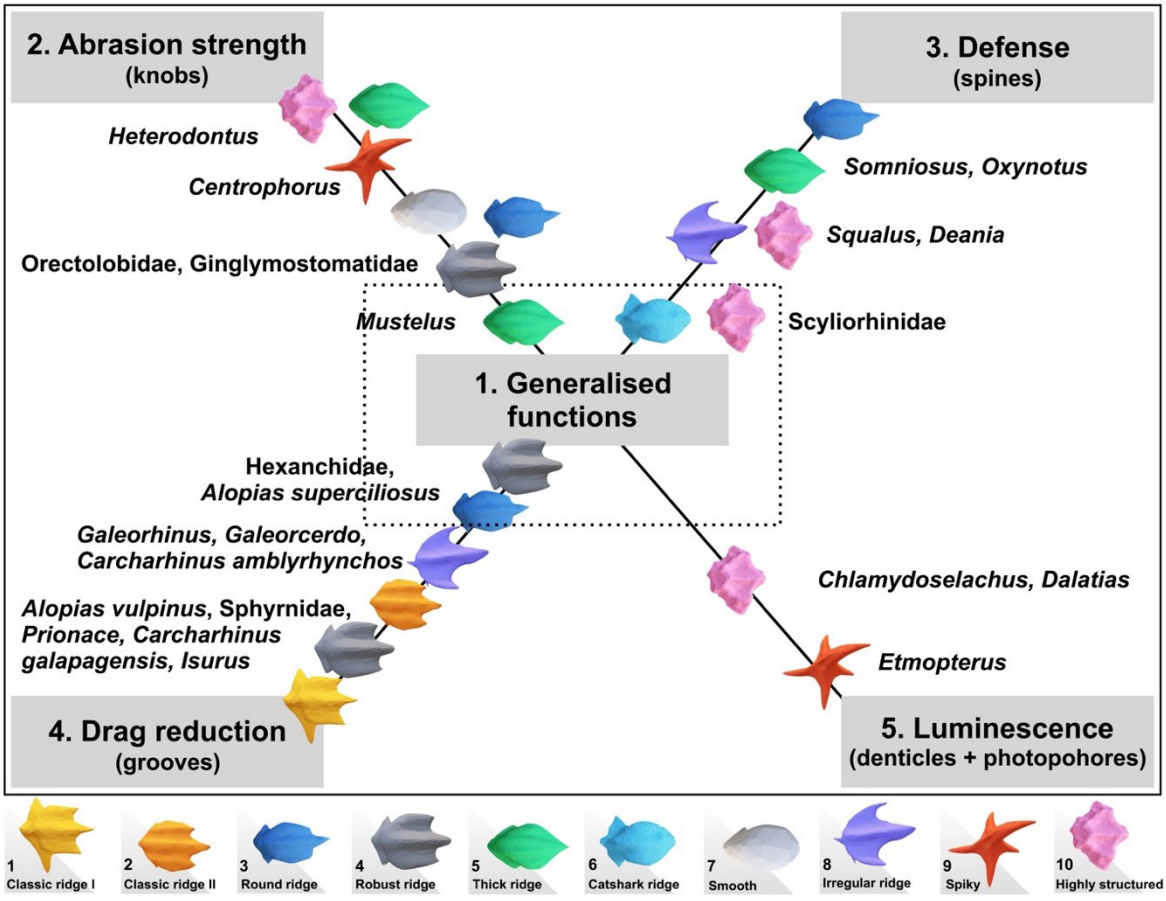


Fig. 9. Denticle morphogroups displayed across categories of function from Reif (1982). Modified version of Reif's five functional categories of shark dermal denticles based on shark lifestyles. Morphogroups are added for each taxon based on their F1-body location (flank region above pectoral fin) to demonstrate how they fit into this original denticle categorization. Stippled line includes morphogroups of the Generalized Function. See Supplementary Table 15 and Supplementary Fig. 2 for the data used to create this figure.

REVIEWER COMMENTS

Authors' note: Revised text can be found through track changes in the Manuscript and Supplementary Information file.

Reviewer #4 (Remarks to the Author):

I appreciate the lengthy responses to my previous comments and the efforts that the authors have made here. I still continue to think that this is by far the most ambitious treatment of shark denticles to date and it is very cool that the authors combine comparative data with multiple functional measurements.

Author response to introductory comment:

We are truly grateful for this comment from reviewer #4. It is evident that the reviewer has spent much time evaluating and improving our manuscript and understands the complicated nature of our project. We deeply appreciate the acknowledgement.

Comment 1

Despite a number of different comments and multiple rounds of revisions, you continue to describe your methods as 'objective' and you also describe experts' opinions on selecting features to quantify as 'subjective'…and in doing both these things you denote that subjectivity is a bad thing and objectivity is a good thing. I totally understand what you mean with these statements, but it is frustrating to see you still use this language despite multiple comments. I think you risk immediately offending anyone who studies and measures morphology. I also think that your responses in the 1st review also revealed that not all of the methods you use for creating 10 morphogroups of denticles are wholly 'objective'. I think you could deliver the same message you have here but without some of these terms. I also

think that by using these terms you are opening a semantic debate about the 'objectivity of science' (many folks would argue that science is never really objective), and a debate about if subjectivity is an inherently bad thing (again, I think many people would argue for or against it). These are interesting topics, but not really the focus of your paper.

Author response 1

Thank you for this comment. We agree. We are glad to be able to rectify our responses from the last round of revision and the reviewer has an excellent point in how we need to carefully report on our methodology. It is not our intention to offend other morphologists or start a debate on this topic and we have carefully rewritten these parts of the manuscript.

Main text, Introduction, Lines 63-64: "These important yet understudied questions are likely due to the difficulty in quantifying and classifying the vast morphological disparity of dermal denticles."

Main text, Introduction, Lines 76-82: "Specifically, representation learning is a machine learning approach that aims to automatically discover the most effective way to represent data prior to recognizing differences between pre-assigned data categories. Thus, through deep learning, we apply a novel approach to quantify denticle morphology, and combined with a recently published phylogenetic tree we can study the evolutionary tempo and mode of these different denticle morphogroups[11]."

Main text, Results, Lines 101-112: "We wanted to categorize dermal denticles based on key morphological features, and for this we used the Variational Sparse Coding model (VSC), a branch of deep learning autoencoders to quantify the visual features of the shark's dermal

denticles[12,13]."

Comment 2

Somewhere – probably around lines 236-238 – please list the total samples (specimens) that were measured for each of the functional testing methods (nanoindentation, mineral protection, CFD). You mention "2-4 species from multiple body locations within each morphogroup" for nanoindentation (and I think also mineral protection?) and then "1-2 samples per morphogroup" for the CFD, but you should also just provide actual numbers. 1-2 samples per morphogroup could be 11 samples or it could be 19…. Similarly, I don't actually know what you mean by "2-4 species from multiple body locations within each morphogroup" means because the language isn't clear. Perhaps this means something between 21-39 samples or perhaps more if each species was sampled at multiple body locations. Both the total number of unique species+locations should be provided, and the range of individuals per species should be given – basically, it would be great to briefly give readers an understanding of what your sample size is here.

Author response 2

We agree and fully comply with this comment. See revised text below. Further, we have altered Figure 5 and revised the legend. We have removed the average sample size from the figure (e.g., "N=156") and, instead, explained the actual sample size in the legend.

Main text Lines 243-287: "While we acknowledge that species within a morphogroup may exhibit differences in material properties despite similarities in external morphology, we believe that our sample size provides valuable insights into the form function relationship (see Supplementary Table 2 for further sampling details). Our sample definition is described

as the species plus the body location, i.e., blackbelly lanternshark snout (S) or shortfin mako shark belly (B), and each sample provides numerous datapoints both for nanoindentation (NI) and for electron probe microanalysis (EPMA). The samples sizes for each morphogroup are as follows: classic ridge 1 (spadenose shark F1, F3, CP, B and shortfin mako shark B, totalling 156 datapoints for NI, 48 for EPMA, and one model for CFD), round ridge (whitespotted bamboo shark PEC1, mandarin dogfish PEC1, and shortfin mako shark F3, CP, totalling 114 datapoints for NI, 80 for EPMA, and one model for CFD), robust ridge (whitespotted bamboo shark F1, F3, CP, mandarin dogfish CP, and shortfin mako shark F1, totalling 231 datapoints for NI, 198 for EPMA, and one model for CFD), thick ridge (spadenose shark PEC1, shortfin mako shark PEC1, and tope shark C2, totalling 90 datapoints for NI, 52 for EPMA, and one model for CFD), catshark ridge (spadenose shark preANL and mandarin dogfish C1, totalling 50 datapoints for NI, 54 for EPMA, and one model for CFD), smooth (whitespotted bamboo shark S, B, spadenose shark S, mandarin dogfish S, B, and shortfin mako shark S, totalling 240 datapoints for NI, 149 for EPMA, and one model for CFD), irregular ridge (longfin catshark F1 and blue shark P3, totalling 53 datapoints for NI, 57 for EPMA, and one model for CFD), spiky (blackbelly lanternshark F1, F3, CP, B, PEC1 and mandarin dogfish F1, totalling 187 datapoints for NI, 87 for EPMA, and two models for CFD), highly structured (mandarin dogfish F3, frill shark B, blackbelly lanternshark S, and birdbeak dogfish B, totalling 113 datapoints for NI, 70 for EPMA, and two models for CFD). See Supplementary Figure 2 for body location abbreviations. For internal properties, we used NI to measure the enameloid hardness, which is the material's resistance to localized plastic deformation (H, 137 indents per morphogroup on average), reduced elastic modulus (E_r, 137 datapoints per morphogroup on average) […] Additionally, we use the F wt% as an indicator of mineral protection, not to be confused with mineral content (89 datapoints per morphogroup on average)[30,31]. […] for drag reduction with

computational fluid dynamics**.** We used one representative sample per morphogroup with the exception of spiky and highly structured where two samples were necessary to cover the morphogroups' morphology (Fig. 5C, Supplementary Fig. 6-7, Supplementary Note 3, see specific species used in Supplementary Table 2 and 9)."

Figure 5 with its revised Legend:

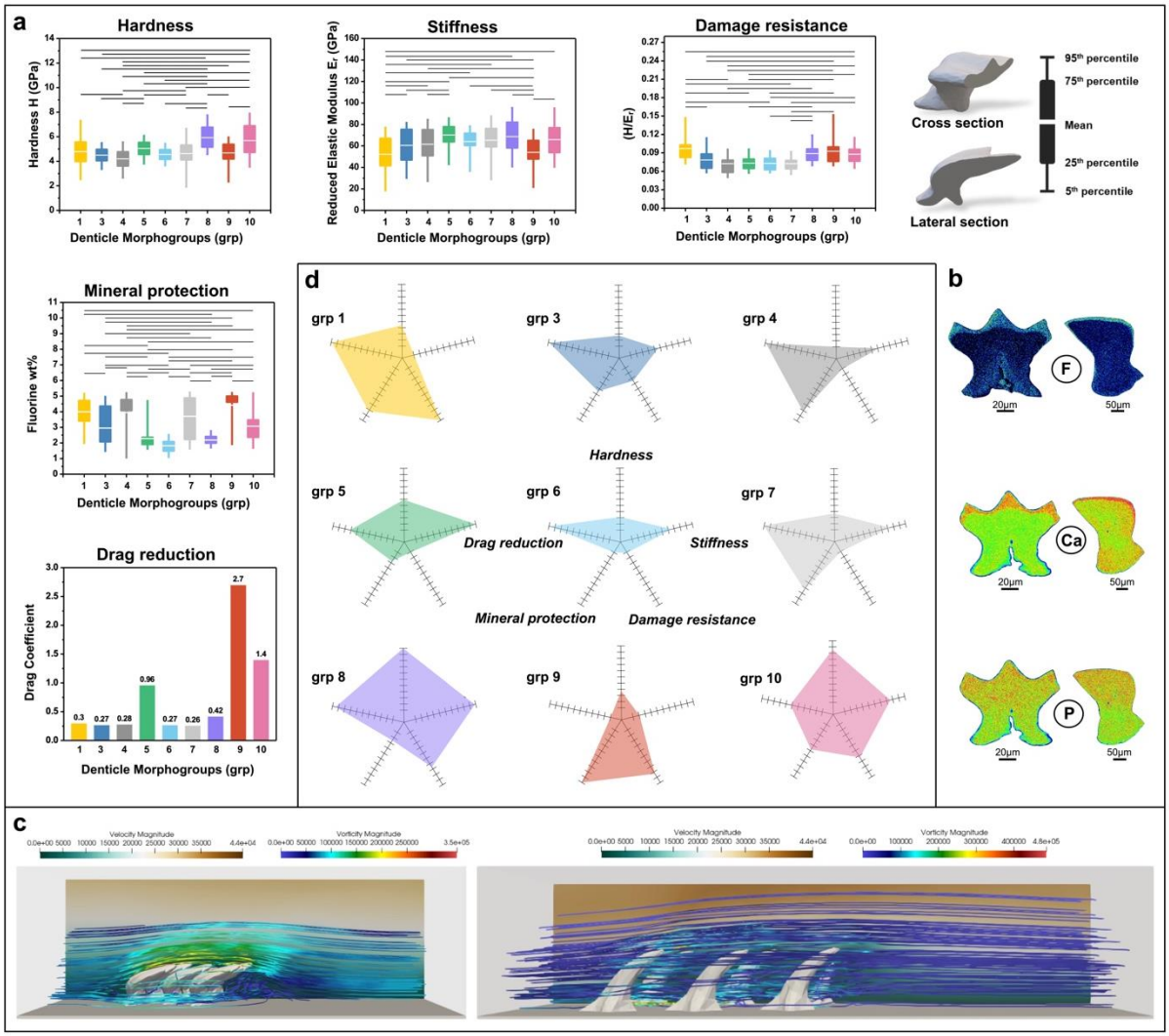


Fig. 5. Functional properties of denticle morphogroups. (**a**) The mechanical, elemental, and fluid dynamic properties of nine morphogroups (excluding group 2) including reduced elastic modulus (stiffness); fluorine wt% (mineral protection); drag coefficient (drag reduction) (Supplementary Tables 1, 4-9). (**b**) Examples of fluorine, calcium, and

phosphorous distribution in the enameloid of the shortfin mako shark *I. oxyrinchus* flank below first dorsal fin (cross section) and snout (lateral section), (Supplementary Table 8). (**c**) Water velocity magnitude of two extremes, the highly efficient *I. oxyrinchus* (left) and the poorly swimming blackbelly lanternshark *Etmopterus lucifer* (right). Each CFD model reflects the physical properties of the actual denticles used in prior NI and EPMA analyses. (**d**) All summarized as functional profiles for each morphogroup. Sample size varies for each morphogroup: for NI (H, E_r, H/E_r) N=156, 114, 231, 90, 50, 240, 53, 187 and 113 for morphogroups 1, 3-10, respectively. Similarly, for EPMA (F wt%) N=48, 80, 198, 52, 54, 149,57, 87 and 70. For the CFD results, one model was used for each of the morphogroups 1, 3-8 and two models were used for groups 9 and 10. Horizontal lines in (**a**) are error bars displaying significant difference between morphogroups with significance level $\boldsymbol{\alpha} = 0.05$ (Tukey's HSD test).

Comment 3

These ranges of numbers you're providing also don't line up with the samples sizes given in the boxplots in Figure 5. This must mean multiple measurements were made on each sample (I think you seem to indicate 10-30 per sample in the methods?). If so, you should not be treating each individual measurement as a statistically independent sample – your methods mention using an ANOVA, but your data seems to have a nested structure with multiple measurements per denticle and per individual. You should be accounting for this data structure in your tests of statistical significant differences in function (nanoindentation, mineral protection, CFD) among the different morphogroups. There is the additional complexity here of sampling multiple distinct regions of the same individual and the fact that species are related and thus no-independent, but without even worry about those you should probably be at least using a nested ANOVA or an ANOVA with some covariates.

Author response 3

Thank you for this important comment. We agree that the structure of our dataset—comprising multiple measurements per denticle, per body location, per species—raises the issue of non-independence, and we appreciate the opportunity to clarify how we approached this.

Ideally, a linear mixed-effects model (LMM) or a nested ANOVA would be used to account for the hierarchical nature of the data. However, due to the limitations in our sampling design—specifically, the relatively low number of biological replicates (i.e., unique species-body-location combinations) and the imbalance across morphogroups—these models were not statistically stable. In particular, a few of the morphogroups are represented by only two or three species-body-location samples, which precludes estimation of variance at higher taxonomic levels. Nested ANOVA is also highly sensitive to such unbalanced replication and can produce misleading or unstable variance partitioning.

We fully acknowledge that treating individual nanoindentation (NI) or elemental measurements (EPMA) as independent inflates the apparent sample size and can increase the risk of pseudo-replication. For this reason, we have deliberately avoided over-interpreting p-values, and instead focused our analysis and discussion on effect sizes, rank order patterns, and morphogroup-level trends that are robust to statistical noise. This interpretive strategy is now explicitly stated in both the **Methods** and **Figure 5 legend**.

Moreover, we wish to highlight the practical limitations that shaped our sampling strategy. Both **nanoindentation** and **electron probe microanalysis** are labor-intensive and resource-intensive techniques, requiring meticulous sample preparation and access to specialized equipment. Within the context of functional trait analysis in biological systems, our dataset

already represents a large number of measurements, both in scale and taxonomic breadth. That said, we absolutely agree with the reviewer that expanding biological replication—especially within morphogroups—is important for future work, and we hope to incorporate this as we secure additional resources.

In response to this comment, we have also revised parts of the main text and conclusions to more clearly acknowledge this limitation and to present our results in a more conservative tone that reflects the constraints of our dataset.

We have updated the figure legend and added the following statement to the Methods section to transparently describe this limitation:

Methods, Lines 716–724: “Statistical comparisons between morphogroups were performed using one-way ANOVA followed by Tukey’s HSD test ($\alpha = 0.05$), based on individual measurement values. We note that our dataset contains repeated measurements nested within denticles, body locations, and species. While a nested or mixed-effects model would be preferable to account for this structure, the low and unbalanced replication at higher levels precluded stable model fitting. We acknowledge this limitation and interpret the results accordingly, emphasizing overall effect patterns rather than statistical thresholds. This is especially important given the high cost and technical demands of NI and EPMA, which limited our ability to generate broader biological replication (Fig. 5).”

Figure 5 legend addition:
“Statistical comparisons are based on individual measurements. Due to limited biological replication and nesting within species and body regions, p-values should be interpreted with caution. We emphasize the consistency and magnitude of trends across morphogroups over strict significance testing.”

We thank the reviewer again for this valuable feedback, which has helped us strengthen both the transparency and interpretive clarity of our work.

Comment 4

I appreciate your attention to looking at the effects of body size on your data. I think I understand some of what you have done and although it isn't ideal to have found differences in body size across your morphogroups, you seem to suggest that it has little effect. However, I'm not sure your results really support that idea. As I'm understanding it, your models have only body length and sex in them? And just with two parameters you gained some improvement over the null model (I don't quite understand what the null model is here)? A classification accuracy of over 50% seems impressive to me with only body size and sex as predictors, but I suppose it is all relative. Is there anything more in this section that you can say to give readers a sense that body length is a minor effect here? It isn't surprising that it might have a minor effect on such a high-dimensional dataset, but the fact that it does seem to separate groups is interesting but also could be an issue.

Author response 4

We appreciate the reviewer's thoughtful comments and the opportunity to clarify the role of body length and sex in our classification model, as well as to explain the meaning and implications of the null model. We have revised the methods as seen below with further explanations and clarifications to the model results.

Methods Lines 502-546: "To assess whether body length and sex influence morphogroup classification, we tested both variables using a series of statistical and machine learning approaches. These analyses included linear (logistic regression) and non-linear (random

forest) models, as well as ANOVA and importance score assessments. We began by defining a **null model** that included only an intercept term and no predictors, thereby assuming that morphogroup assignment is entirely random based on the observed class distribution. This serves as a baseline for evaluating whether body length and sex add explanatory power. The **ANOVA** revealed significant differences in body length across morphogroups ($F = 33.182$, $p < 0.001$; Supplementary Fig. 6A, Supplementary Table 12), and post hoc Tukey's HSD tests identified which morphogroups differ (Supplementary Fig. 6B). However, a random forest model assigned **97.7% importance to body length** and just **2.3% to sex**, suggesting that sex plays a minimal role and was therefore excluded from further analysis.

To quantify the predictive value of body length, we trained classification models using a 70:30 train-test split. The **logistic regression model** performed poorly, with mean accuracy of 28.5%, F1-score of 0.16, and a **McFadden's pseudo R^2 of 0.062**, indicating only a 6.2% improvement over the null model. Additionally, the absolute coefficients of body length and sex were similar in magnitude (ANOVA $F = 1.97$, $p = 0.099$), suggesting neither feature substantially improves prediction. The **random forest classifier** showed improved performance (accuracy = 54.5%, F1 = 0.55, pseudo $R^2 = 0.142$), with a 14.2% log-likelihood gain over the null model. This model is capable of capturing non-linear and high-dimensional interactions, so the limited performance is notable. **Although body length was assigned a high importance score of 97.7%, the overall model performance suggests that its contribution might not be as impactful as the importance score implies.** This indicates a modest overall gain, suggesting that these two predictors offer limited ability to explain morphogroup structure in this dataset. Similarly, the classification accuracy of 54.5%, while better than chance, is still relatively low—particularly considering that a random forest classifier, which is capable of modelling nonlinear and complex relationships, was used. If

body length and sex were strong determinants of morphogroup identity, we would expect substantially higher performance.

Further, we performed Pearson correlation and linear regression between body length and each of the first two UMAP dimensions (Supplementary Fig. 6D). UMAP1 had a statistically significant positive correlation with body length ($r = 0.088$, $p = 0.0315$; $R^2 = 0.008$), whereas UMAP2 showed a negative correlation with body length ($r = –0.194$, $p < 0.0001$; $R^2 = 0.038$). In both cases, the low R^2 values suggest that body length explains only a small proportion of the variation along the axes of UMAP1 and UMAP2. It should be noted that the low statistical significance is an artifact of large sample size and not meaningful in a functional context. Taken together, these results support the interpretation that body length and sex have only **minor effects** on morphogroup classification. While they may introduce weak structure, their explanatory contribution is limited in the context of our high-dimensional morphometric dataset.**”**

Comment 5

I also should have been a bit more prescriptive in my previous review. I would really like to see data on how body size relates (correlates) to the various UMAP axes. Significance values and correlations (r or R-squared).

Author response 5

We appreciate the reviewer’s request to quantitatively assess the relationship between body size and the low-dimensional embedding produced by UMAP.

In response, we conducted Pearson correlation and linear regression analyses between body length and each of the first two UMAP dimensions (Supplementary Fig. 6D). UMAP1

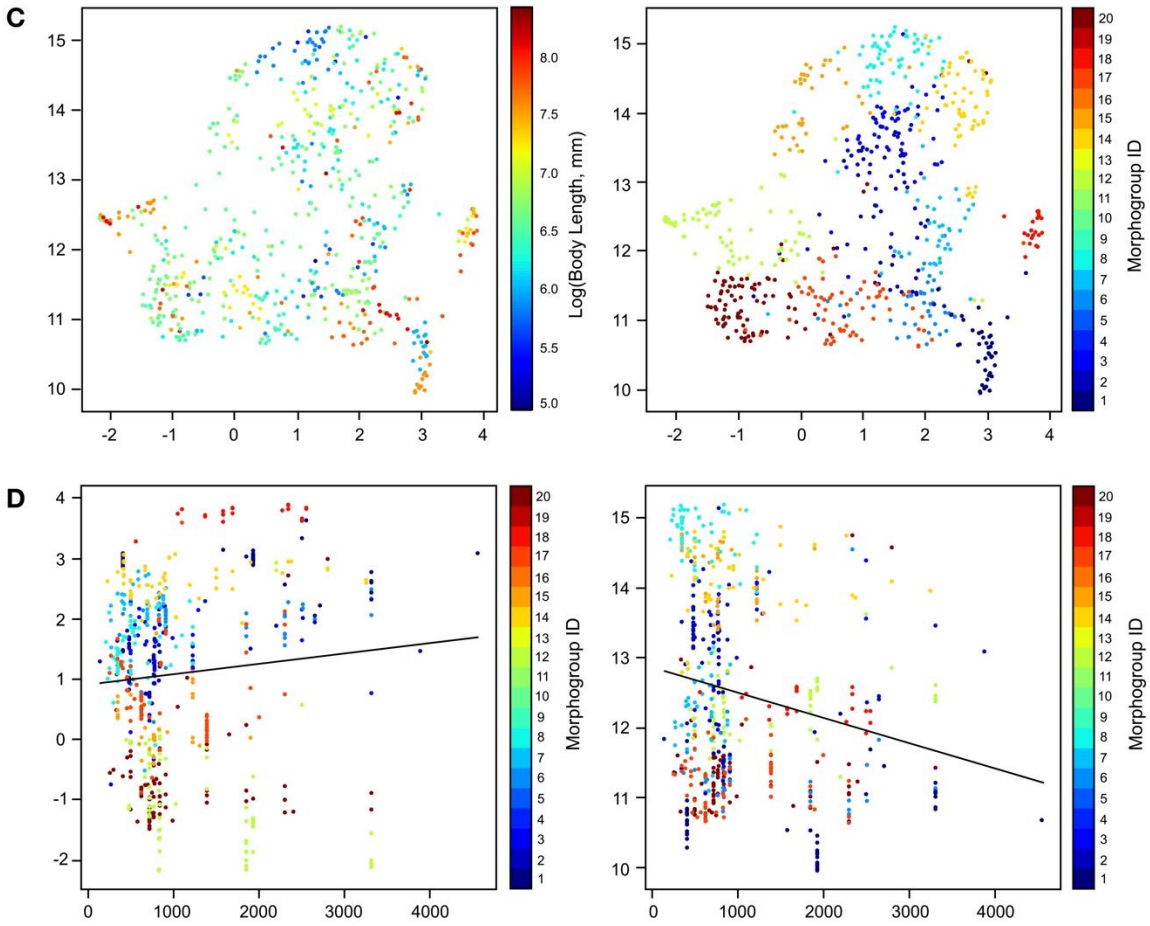
exhibited a statistically significant but weak positive correlation with body length ($r = 0.088$, $p = 0.0315$; $R^2 = 0.008$), while UMAP2 showed a slightly stronger negative correlation ($r = –0.194$, $p < 0.0001$; $R^2 = 0.038$). These results indicate that body size does explain a small portion of the variation in the embedding.

While these relationships are statistically significant, we emphasize that UMAP is a non-linear, non-parametric dimensionality reduction technique, and its axes are not explicitly defined in relation to interpretable variables such as body size. Thus, the observed correlations, though real, reflect weak structure and should be interpreted cautiously. The statistical significance is likely driven in part by the large sample size. Taken together with our classification results (random forest mean accuracy = 54.5%; pseudo $R^2 = 0.142$), these findings suggest that body length may exert a subtle influence on the latent structure, particularly along UMAP2, but it is not a primary determinant of morphogroup clustering.

We have added these analyses to the Methods and Supplementary Fig. 6, and clarified these interpretations accordingly.

Methods Lines 532-546: “[…] To further evaluate the role of body size in the morphometric structure, we performed Pearson correlation and linear regression between body length and the first two UMAP dimensions (Supplementary Fig. 6D). UMAP1 exhibited a statistically significant but weak positive correlation with body length ($r = 0.088$, $p = 0.0315$; $R^2 = 0.008$), while UMAP2 showed a stronger negative correlation ($r = –0.194$, $p < 0.0001$; $R^2 = 0.038$). Although these correlations are statistically significant, the low R^2 values indicate that body length explains only a small proportion of the variance along each UMAP axis. These trends suggest a limited influence of body size on the embedded morphometric structure.

It is important to note that UMAP is a non-linear, non-parametric dimensionality reduction technique. Its axes do not preserve linear relationships from the original high-dimensional data, and correlations with biological variables should be interpreted cautiously. Combined with our classification results—where body length had a high importance score (97.7%) but yielded modest model performance (random forest accuracy = 54.5%, McFadden's pseudo R^2 = 0.142)—these findings suggest that body size may exert some weak structure, but it is unlikely to be a dominant factor shaping the morphogroup clustering."



[Addition to] Supplementary Fig. 6 (C-D). Effect of body length on denticle morphogroups. […] (C) Body length distribution across denticle morphogroups. Left panel is colour-coded by log-transformed body length; right panel shows the original morphogroup

assignments (1–20). (D) Correlation between body length and UMAP axes. UMAP1 shows a weak positive correlation ($r = 0.088$, $R^2 = 0.008$), and UMAP2 a slightly stronger negative correlation ($r = –0.194$, $R^2 = 0.038$). While statistically significant, these relationships explain little variance and indicate limited influence of body size on the embedded morphometric structure. "

Reviewer #5 (Remarks to the Author):

The paper employed the VSC model and K-means clustering algorithm to quantify and group the geometry features of shark denticles by analyzing SEM images. Combing with nanoindentation, electron probe micro analysis and CFD simulations, the authors tried to establish the form-function relationship of the dermal denticles. Also, the tempo and model of shark denticle evolution has been studied through phylogenetic tree analysis. The topic is interesting, the analysis is comprehensive, and the paper is well written. However, I have a few questions for the authored to be addressed.

General concerns:

Comment 6

The authors already got the shark denticle 3D models using SEM methods. But in the CFD part, the authors used micro 3D-scan to obtain the 3D models. Could the authors explain why they didn’t directly use the 3D models from SEM and the differences between these two methods?

Author response 6

We appreciate the reviewer’s observation and the opportunity to clarify this point. The SEM-derived 3D model was created only for morphogroup 2, which lacked a physical specimen

suitable for micro-CT scanning. However, this model was used solely for illustrative purposes and was not included in any computational fluid dynamics (CFD) simulations. All CFD analyses in our study were based exclusively on micro-CT–derived 3D models, which provide full volumetric data essential for accurate fluid simulations. Unlike SEM, which only captures surface morphology in two dimensions, micro-CT scanning allows non-invasive reconstruction of complete 3D geometry, including internal concavities and thickness variations—features that are critical for CFD modeling, especially in complex denticle shapes. We have clarified this point in the Methods section (lines 694–703), and we include the relevant passage below for convenience.

Methods Lines 802-808: "The 3D model of morphogroup 2 was created using SEM images because no physical specimen was available for micro-CT scanning. Using ImageJ, we extracted denticle outlines, which were imported into SolidWorks as SVG files and reconstructed into 3D geometry via Extrude and Revolve tools. Parameters such as thickness and surface features were adjusted to approximate the morphology seen in the SEM images. This model was not used in fluid dynamic simulations or any drag analyses, and served only as a visual representation."

Comment 7

Although the external morphology features determine the functions of dermal denticles, the hydrodynamic function also depends on the fluid environment, the locomotion gait of sharks and the arrangement of denticles. With the same flow conditions and same setup, the conclusions about the hydrodynamics of different groups denticles may not be very solid.

Author response 7

Thank you for this important observation. We agree that the hydrodynamic function of dermal denticles is influenced not only by their external morphology but also by the surrounding fluid environment, the shark's locomotion gait, and denticle arrangement.

In this study, our aim was to investigate the general functional implications of denticle morphology rather than species-specific adaptations. Each morphogroup represents a collection of denticles grouped by morphological similarity across multiple shark species. Although the representative denticle for each morphogroup comes from a specific species, we avoided extrapolating environmental conditions of that species to the entire morphogroup. Instead, we subjected all morphogroup representatives to the same controlled flow conditions to enable direct comparison of their geometries. The denticle arrangements were preserved, as each CFD model was generated from micro-CT scans of real shark skin.

To clarify this approach, see revised methods below.

Methods Lines 864-872: "We emphasize that these flow conditions are selected based on the examination of the denticle geometry rather than the whole body of the sharks. Adding to this, species-specific environmental factors that may influence the drag conditions, such as water temperature, were not considered. However, we did examine a range of water viscosities affecting the drag of each denticle design (Supplementary Note 3.4, Supplementary Fig. 15B). We found that drag coefficients decreased as viscosity decreased (i.e., as Reynolds number increased) and stabilized at low viscosity conditions where inertial forces dominated the flow behavior, confirming that the chosen flow conditions were suitable for evaluating denticle hydrodynamics."

We acknowledge the reviewer's concern and believe that controlling the variables across morphogroups is critical to avoid misleading conclusions in CFD-based comparisons.

Concerns in Supplementary Information:

Comment 8

The dimensions of shark denticles in simulations? It seems the authors enlarged the denticles, which is okay if the authors can keep the dimensional parameters, such as the Reynolds number, are the same. However, the local Reynolds number (calculated based on the model length) the authors used is around 2*10^4, which is much larger than that for shark swimming. The local Re of shark denticles changed around 1,000[1].

Author response 8

Thank you for this thoughtful comment. We would like to clarify that our primary CFD simulations and mesh independence tests were conducted at a Reynolds number (Re) of **10,000**, not 20,000 as stated in the comment. The Re = 20,000 value may stem from our **sensitivity analysis** (Supplementary Fig. 15B), in which we explored a range of flow conditions up to Re = 10^7 to test the robustness of morphological effects on drag. These higher Reynolds numbers were used for **trend verification only**, not for our main simulations.

We selected **Re = 10,000** deliberately because **lower Reynolds numbers**, such as **Re ≈ 1,000**—which are typical of swimming sharks—produce fully **laminar flow**, dominated by viscous forces. In this regime, **flow separation and vortex formation are minimal**, and **the drag becomes largely insensitive to surface geometry**. Therefore, if we had used Re = 1,000, it would have **precluded any meaningful comparison** of denticle shape on hydrodynamic performance. By contrast, using Re = 10,000 provides enough flow

complexity—specifically, the onset of **separation and vortex shedding**—to detect differences in drag caused by **morphological features** such as ridges, concavities, and spacing. It also avoids the need for fully turbulent models, balancing **physical realism and computational feasibility**.

Finally, our sensitivity analysis shows that the **relative differences** in drag performance across denticle morphogroups remain consistent across a wide Re range (from creeping flow to 10^7), as shown in Supplementary Fig. 15B. This confirms the **robustness of our comparative results**, even if the absolute drag values vary.

Please see the revised text below.

Methods Lines 875-878: "[…] the size of the denticles is approximate at $D_x = 2.08$ mm, $D_y = 0.45$ mm, and $D_z = 2.21$ mm (Note: these dimensions may vary for denticles of different shark species and their overall geometry is illustrated in the 3D plots of Supplementary Fig. 8)."

Methods Lines 881-890: "For all simulations comparing denticle geometries, we used a constant Reynolds number of **Re = 10,000**, as described in Supplementary Note 3.4. At lower Reynolds numbers (e.g., Re ≈ 1,000), the flow is fully laminar and surface geometry has minimal impact on drag. This limits the ability to detect morphological effects. By contrast, Re = 10,000 produces flow with moderate complexity, allowing surface morphology to influence drag characteristics while remaining computationally tractable. These conditions thus enable a meaningful comparative analysis of denticle forms."

Comment 9

Could the authors explain why you set the outflow boundary condition as zero pressure? Have you tried other outflow boundary conditions? I am not sure whether there is not enough space between the denticles and the outflow boundary. Also, for the bottom where the denticles are placed, should it be set up as no-slip boundary condition to mimic the shark body?

Author response 9

Thank you for raising these important technical points. We are happy to clarify the rationale behind our boundary condition settings.

For the **outflow boundary**, we applied a **zero-pressure condition**, which is standard practice for steady-state incompressible flow simulations. This boundary condition minimizes pressure buildup at the outlet and avoids the introduction of non-physical flow reversals. It ensures a numerically stable exit of fluid from the simulation domain and aligns with conventional CFD treatment for drag force estimation in open-channel or external flow configurations.

To address the reviewer's concern about proximity to the outflow boundary, we conducted additional tests with an extended downstream domain. These tests revealed **no appreciable difference in drag coefficients**, confirming that the original outlet location did not induce artificial backpressure or flow distortion. As such, we retained the more compact domain to optimize computational efficiency without sacrificing accuracy.

Regarding the **bottom boundary condition**, we confirm that a **no-slip condition** was applied to mimic the interaction between denticles and the shark's body. This ensures the fluid velocity at the denticle base is zero, which accurately reflects the physical constraint of skin-

attached denticles under swimming conditions. As described in Supplementary Information (Lines 95–96): *"The denticles are assumed to be rigid bodies in the simulation domain, and the water/denticle interface is assumed to be no-slip."*

We appreciate the opportunity to clarify these modeling decisions and hope this addresses the reviewer's concerns.

Comment 10

The temperature where sharks live varied widely, which greatly affects the ocean water viscosity and is not negligible.

Author response 10

Thank you for pointing this out. We fully agree that water temperature, and thus viscosity, varies considerably among shark habitats and that this variation may influence drag dynamics. However, our study was designed to isolate the effect of denticle morphology, not to replicate species-specific swimming conditions.

To achieve a controlled comparison across morphogroups, we intentionally used a fixed set of fluid parameters, including constant temperature and viscosity. This approach is widely used in drag reduction studies to isolate geometric effects under standardized conditions. Introducing temperature-specific viscosity for each species would introduce multiple confounding variables, undermining our ability to directly compare hydrodynamic performance attributable to denticle shape alone.

Furthermore, at Reynolds numbers typical of shark swimming (Re ~1,000), flow remains laminar and morphology has minimal effect on drag. As discussed in our response to

comment 8, we chose Re = 10,000 to ensure sufficient flow complexity (including separation and vortex formation) necessary to reveal the functional consequences of morphology. Within this framework, absolute viscosity becomes less critical than the relative trends between morphogroups.

To account in principle for broader viscosity effects, we performed a parametric sensitivity analysis by varying the Reynolds number over several orders of magnitude (up to 10^7). As shown in Supplementary Figure 15B, the relative differences in drag between denticle shapes remained consistent over this range, indicating the robustness of our results.

Finally, as described in the Supplementary Information (lines 90-91), our model assumes a constant fluid density ($\rho = 1$ mg/mm^3) and assigns dynamic viscosity as $\nu = 1/Re$. This modeling approach provides a simplified yet effective way to incorporate viscosity effects via the Reynolds number while maintaining a clean comparison between morphologies.

Comment 11

The surface drag Fd should both include viscous drag and pressure drag. At least at Re=10^4, the pressure drag is not negligible.

Author response 11

Thank you for this important comment. We appreciate the reviewer's attention to detail, and we agree that at Re = 10^4, both viscous and pressure drag are significant and must be properly accounted for in the calculation of total surface drag.

The confusion likely stems from a misleading sentence in the Supplementary Information:

"We only measure the viscous stress in the direction of the flow velocity to calculate the drag force, assuming that the pressure is always perpendicular to the denticle surfaces."

This wording unintentionally implied that we excluded pressure drag from our simulations, which is not the case. We apologize for the lack of clarity in this sentence.

After careful review of our simulation pipeline and implementation of the drag force calculation, we confirm that our code uses the full surface stress tensor, which includes both viscous and pressure components, to compute the total drag force in the flow direction. Thus, the reported drag coefficients in our figures and analyses do indeed accurately reflect the total surface drag, as expected in CFD modeling of incompressible flows.

To avoid further confusion, we have revised the Supplementary Information as follows:

Revised Supplementary Lines 206-210: “We calculate the total drag force using the full surface stress tensor, which includes both viscous shear stresses and pressure forces acting on the denticle surfaces in the direction of flow. This standard approach ensures that both viscous and pressure drag contributions are captured in the resulting drag coefficient (C_D)".

Again, we thank the reviewer for bringing up this important clarification. It allowed us to correct an inaccurate description and reinforce the rigor of our computational setup.

Comment 12

Typo: Line 171 in Supplementary Information, “element”.

Author response 12

This has been corrected.

Comment 13

Line 216: “L = 3”. What is the unit?

Author response 13

$L = 3$ **mm** is the characteristic length scale used for cuboid fluid domain and we apologize for the typo made here. We have corrected it.

Comment 14

Supplementary Fig. 15: the discrepancy between the simulation and experiment of flow past sphere is more than 30% at Re=10^4, which is not a slight difference for a validation case. If the authors included the pressure force, the discrepancy might disappear. The pressure force is more sensitive to the morphology changes.

Author response 14

Thank you for this insightful comment. We fully agree that a >30% discrepancy in drag coefficient at Re = 10^4, as shown in Supplementary fig. 15, is not trivial for a validation case. This feedback is valuable, and we welcome the opportunity to clarify both the scope and limitations of our validation.

First, we emphasize - as clarified in Author Response 11 - that our simulations compute the drag force using the full surface stress tensor, which includes both viscous and pressure forces. The previous wording in the Supplementary Information may have been misleading, and we have corrected it to reflect the correct treatment of pressure drag.

After re-examining our setup, we confirm that the reported drag coefficient (C_D) includes the total surface drag. The discrepancy observed in the sphere benchmark is not due to the omission of compressive forces, but rather to small differences in geometry and mesh fidelity between our model and the physical sphere used in Achenbach's experiments [1]. At moderate to high Reynolds numbers (e.g., Re = 10^4), such small discrepancies in shape or surface roughness can lead to amplified differences in pressure distribution, making pressure drag particularly sensitive to even subtle morphological details.

While the absolute value of drag is different, our simulation reproduces the expected trend in drag behavior with respect to Reynolds number, as shown over a wide range of Re in Supplementary Fig. 15B. In addition, our CFD solver (based on FEniCS and the VarMINT formulation) has been independently validated in published work [2,3], further supporting the reliability of the numerical framework.

[1] Achenbach, E. (1972). Experiments on flow past spheres at very high Reynolds numbers. Journal of Fluid Mechanics, 54(3), 565–575.

[2] https://github.com/david-kamensky/VarMINT

[3] Zhu, Q., & Yan, J. (2021). A moving-domain CFD solver in FEniCS with applications to tidal turbine simulations in turbulent flows. Computers & Mathematics with Applications, 81, 532–546.

We have now added reference [1] to Supplementary Note 3.4 on line 221 to support this clarification.

Finally, we would like to reiterate that the goal of our study is not to validate CFD accuracy in absolute terms, but rather to compare the relative drag performance of different denticle

morphologies under identical simulation conditions. Thus, although the absolute error is not negligible, the internal consistency of our simulations is preserved, which is critical for comparing shape effects.

Comment 15

For the convergence study, I don't think the size differences between the coarse, medium and fine meshes are large enough, which might affect the choice of mesh size. Also, please provide the quantitative comparison of three meshes instead of flow plots which can't tell whether the simulation is converged or not.

[3] Graybill, M. T., & Xu, N. W. (2024). Experimental Studies of Bioinspired Shark Denticles for Drag Reduction. Integrative and Comparative Biology, 64(3), 742-752.

Author response 15

Thank you for this valuable comment. We acknowledge the reviewer's concern about the refinement ratio between coarse, medium, and fine meshes. The mesh resolutions used in this study — (A) Mesh 1: 275,442 elements, (B) Mesh 2: 550,884 elements, and (C) Mesh 3: 1,101,768 elements — are indeed modest in refinement ratio. However, these were selected after benchmarking our in-house FEniCS-based solver against available GPU/CPU resources. Given the computational complexity of simulating 3D flow over biologically accurate denticle geometries, these mesh sizes represented the maximum feasible resolution for systematic comparison across morphogroups.

To address the reviewer's request for a more rigorous and quantitative convergence analysis beyond flow plots, we have now included a comparison of drag coefficient evolution over time, as well as velocity magnitude profiles across all mesh levels. These are shown in

Supplementary Fig. 14I–J and discussed in the revised Supplementary Note 3.4 (lines 266–283).

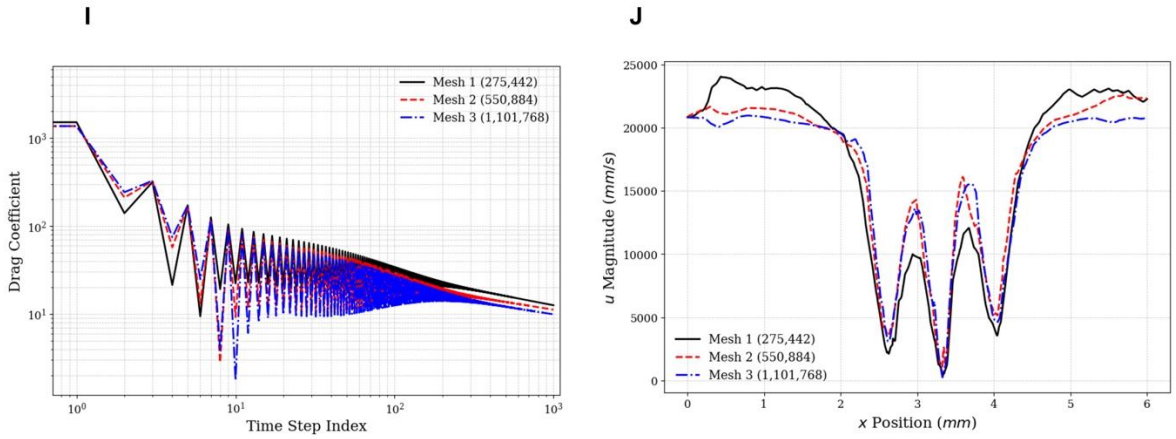
Specifically:

Supplementary Information, Note 3.4, Lines 266-283: ” Finally, we performed a mesh convergence analysis to assess the numerical sensitivity of the drag computation to mesh resolution. The three mesh levels used in our test — Mesh 1 (275,442 elements), Mesh 2 (550,884 elements), and Mesh 3 (1,101,768 elements) — were selected based on practical computational limits of our in-house solver architecture. Although the refinement ratio between successive meshes is moderate, it reflects the highest feasible resolution for full-domain, time-dependent CFD simulations across multiple morphogroups.

To evaluate convergence, we analyzed two metrics: (1) the time evolution of the drag coefficient across the three meshes, and (2) the velocity magnitude distribution in the near-wall region where sensitivity to resolution is typically highest.

As shown in Supplementary Fig. 14I, the drag coefficient converges to a stable value in all three meshes after an initial transient period. The difference between Mesh 2 and Mesh 3 is minimal, indicating that Mesh 2 is already sufficient for capturing the integrated hydrodynamic response.

Additionally, Supplementary Fig. 14J displays velocity profiles sampled along a horizontal line located above the center of the denticle matrix (from (0, 0.45, 1.5) to (6, 0.45, 1.5)). Profiles from Mesh 2 and Mesh 3 are nearly indistinguishable, even in regions with steep gradients or recirculation. Together, these results demonstrate that the solution exhibits effective mesh convergence for both global and local flow features under our simulation framework.”



Supplementary Fig. 14. Mesh independency test for the lantern shark *E. lucifer* F1 denticles (grp 9): […] **(I)** Time-step convergence of drag coefficient across three mesh resolutions. (**J**) Velocity magnitude profiles along a horizontal line above the denticle array for three mesh resolutions.

We thank the reviewer for highlighting the recent work by Graybill and Xu [3]. While their study focuses on simplified denticle analogs at lower Reynolds numbers, we agree it provides valuable complementary context and have now cited it in the revised Supplementary Information Line 186.

REVIEWER COMMENTS

Authors' note:

Revised text can be found through track changes in the Manuscript and Supplementary Information file. Additionally, changes can be found in **bold** in the Response Letter. Figures and tables included in this Response Letter for explanatory purposes, and not contained in the manuscript files, are labeled "RL" (e.g., Figure RL1 and Table RL1). Line numbers in the Response Letter refer to manuscript files without track changes.

All updated figures, tables, and CFD re-runs have been incorporated into the revised manuscript and Supplementary Information.

Reviewer #4 (Remarks to the Author):

I think the authors have done a nice job of responding-to and addressing my comments. And of course I continue to think that this is the most detailed treatment of shark denticle morphology and potential functions to date. I think this paper will be very popular for students and other scientists interested in this and similar topics.

I have only two last comments and they are again about clarifying sample sizes. Supplementary table 2 is really helpful, as are the additions you have made in your text and figure captions. Thanks for those additions. I suspect that your samples for NI and EPMA all come from one individual shark per species. I am also unsure about how many repeated measurements are made on a single denticle. I have two requests:

Author response to introductory comment:

We thank the reviewer for revisiting our manuscript and for the valuable time invested in helping us improve it. We greatly appreciate the positive feedback and hope that the final version of our work will be both useful and engaging for fellow enthusiasts.

Comment 1

Can you confirm that all your sharks are represented by one individual per species and state that clearly somewhere in your methods (perhaps when you introduce the sampling of denticles for NI and EPMA). This should be clearly stated.

Author response 1

We agree and have complied fully.

Methods Lines 652-654 (when we introduce sampling for NI and EPMA): "**For each species, we sampled one individual (one specimen per species)**. NI and EPMA measurements from all samples were assigned to AI-defined morphogroups after data collection."

Comment 2

When you make NI and EPMA measurements on sections, how many different denticles do your measurements come from? I'm guessing it might be one denticle for each method but I'm not sure. This should be clarified multiple places (Supplemental Table 2, Lines 245-270, Material and methods lines 691+). For example, there are 32 NI data points for *Scoliodon laticaudus* F1. How many of these are from one denticle? Or how many data points are collected for each denticle if more than one denticle is sampled? This should be clearly stated in some way…again likely in multiple parts of your paper.

Thanks for all your patience here and with my reviews. I think your paper has continued to improve in quality and clarity, and I think you have done an excellent job here.

Author response 2

We understand the need for this clarification and we have provided the number of denticles per sample used for EPMA and NI, respectively, in the Supplementary Table 2, in the Results, and in the Methods section.

Revised Supplementary Table 2. Overview of sample size. A list of the shark species and body locations **(samples)** used to represent each morphogroup in the mechanical **(NI)**, elemental **(EPMA)**, and fluid dynamic analyses **(CFD)**. Values show the exact number of data points collected from each morphogroup. Coloured values in parentheses reveal the exact number of NI and EPMA data points achieved, **followed by brackets with the number of denticles used for each sample**, in blue and red respectively. See Supplementary Figure 2 for body location abbreviations.

Morphogroup	*Species* (bodylocation (NI data points [number of denticles], EPMA data points [number of denticles])	NI data points [number of denticles]	EPMA data points [number of denticles]	CFD models
1 Classic ridge I	*Scoliodon laticaudus* (F1 (32 [5], 6 [2]), F3 (31 [2], 6 [2]), CP (30 [3], 11 [2]), B (33 [2], 14 [2])) *Isurus oxyrinchus* (B (30 [2], 11 [2]))	156 [13]	48 [10]	*Scoliodon laticaudus* (F1)
2 Classic ridge II	-	-	-	-
3 Round ridge	*Chiloscyllium plagiosum* (PEC1 (34 [2], 25 [2])) *Cirrhigaleus barbifer* (PEC1 (30 [2], 25 [2])) *Isurus oxyrinchus* (F3 (29 [2], 16 [2]), CP (30 [3], 14 [2]))	114 [9]	80 [8]	*Scoliodon laticaudus* (C1)
4 Robust ridge	*Chiloscyllium plagiosum* (F1 (51 [2], 38 [2]), F3 (60 [2], 60 [2]), CP (59 [2], 67 [2])) *Cirrhigaleus barbifer* (CP (31 [2], 12 [2])) *Isurus oxyrinchus* (F1 (30 [2], 21 [2]))	231 [10]	198 [10]	*Isurus oxyrinchus* (F1)
5 Thick ridge	*Scoliodon laticaudus* (PEC1 (30 [2], 3 [2])) *Isurus oxyrinchus* (PEC1 (31 [2], 20 [2])) *Hemitriakis japanica* (C2 (29 [2], 29 [2]))	90 [6]	52 [6]	*Hemitriakis japanica* (C2)
6 Catshark ridge	*Scoliodon laticaudus* (preANL (24 [2], 21 [2])) *Cirrhigaleus barbifer* (C1 (26 [2], 33 [2]))	50 [4]	54 [4]	*Scoliodon laticaudus* (preANL)
7 Smooth	*Chiloscyllium plagiosum* (S (60 [2], 67 [2]), B (45 [2], 25 [2])) *Scoliodon laticaudus* (S (41 [2], 2 [2])) *Cirrhigaleus barbifer* (S (30 [2], 26 [2]), B (29 [2], 15 [2])) *Isurus oxyrinchus* (S (35 [3], 14 [2]))	240 [13]	149 [12]	*Scoliodon laticaudus* (S)
8 Irregular ridge	*Apristurus herklotsi* (F1 (25 [4], 10 [2])) *Prionace glauca* (P3 (28 [2], 47 [2]))	53 [6]	57 [4]	*Apristurus herklotsi* (F1)
9 Spiky	*Etmopterus lucifer* (F1 (31 [3], 17 [2]), F3 (30 [2], 19 [2]), CP (30 [3], 12 [2]), B (35 [4], 15 [2]), PEC1 (31 [2], 10 [2])) *Cirrhigaleus barbifer* (F1 (30 [2], 14 [2]))	187 [16]	87 [12]	*Etmopterus lucifer* (F1) *Cirrhigaleus barbifer* (F1)
10 Highly structured	*Cirrhigaleus barbifer* (F3 (28 [3], 16 [2])) *Chlamydoselachus anguineus* (B (29 [2], 29 [2])) *Etmopterus lucifer* (S (30 [3], 14 [2])) *Deania calcea* (B (26 [2], 11 [2]))	113 [10]	70 [8]	*Deania calcea* (B) *Chlamydoselachus anguineus* (B)

Results lines 235-267:

"Our sample definition is described as the species plus the body location, i.e., blackbelly lanternshark snout (S) or shortfin mako shark belly (B), and each sample provides numerous data points both for nanoindentation (NI) and for electron probe microanalysis (EPMA). **For NI and EPMA, we extracted data from 2–5 different denticles per sample, which is detailed in Supplementary Table 2 and in the following description.** The sample sizes for each morphogroup are as follows: classic ridge 1 (spadenose shark F1, F3, CP, B and shortfin mako shark B, totalling 156 data points for NI **across 13 denticles averaging 12 indents per**

denticle, 48 for EPMA **across 10 denticles averaging 4.8 data points per denticle**, and one model for CFD), round ridge (whitespotted bamboo shark PEC1, mandarin dogfish PEC1, and shortfin mako shark F3, CP, totalling 114 data points for NI **across 9 denticles averaging 12.7 indents per denticle**, 80 for EPMA **across 8 denticles averaging 10 data points per denticle**, and one model for CFD), robust ridge (whitespotted bamboo shark F1, F3, CP, mandarin dogfish CP, and shortfin mako shark F1, totalling 231 data points for NI **across 10 denticles averaging 23 indents per denticle**, 198 for EPMA **across 10 denticles averaging 19.8 data points per denticle**, and one model for CFD), thick ridge (spadenose shark PEC1, shortfin mako shark PEC1, and tope shark C2, totalling 90 data points for NI **across 6 denticles averaging 15 indents per denticle**, 52 for EPMA **across 6 denticles averaging 8.7 data points per denticle**, and one model for CFD), catshark ridge (spadenose shark preANL and mandarin dogfish C1, totalling 50 data points for NI **across 4 denticles averaging 12.5 indents per denticle**, 54 for EPMA **across 4 denticles averaging 13.5 data points per denticle**, and one model for CFD), smooth (whitespotted bamboo shark S, B, spadenose shark S, mandarin dogfish S, B, and shortfin mako shark S, totalling 240 data points for NI **across 13 denticles averaging 18.5 indents per denticle**, 149 for EPMA **across 12 denticles averaging 12.4 data points per denticle**, and one model for CFD), irregular ridge (longfin catshark F1 and blue shark P3, totalling 53 data points for NI **across 6 denticles averaging 8.8 indents per denticle**, 57 for EPMA **across 4 denticles averaging 14.3 data points per denticle**, and one model for CFD), spiky (blackbelly lanternshark F1, F3, CP, B, PEC1 and mandarin dogfish F1, totalling 187 data points for NI **across 16 denticles averaging 11.7 indents per denticle**, 87 for EPMA **across 12 denticles averaging 7.3 data points per denticle**, and two models for CFD), highly structured (mandarin dogfish F3, frill shark B, blackbelly lanternshark S, and birdbeak dogfish B, totalling 113 data points for NI **across 10 denticles averaging 11.3 indents per denticle**, 70 for EPMA **across 8 denticles averaging 8.8 data points per denticle**, and two models for CFD). See Supplementary Figure 2 for body location abbreviations."

Methods lines 645-665:

"Nanoindentation (NI) and electron probe microanalysis (EPMA)

Ten shark species were found to represent the morphological variation in nine morphogroups, excluding group 2 due to lack of physical material (see species in Supplementary Table 2). **The species were chosen based on taxonomic breadth, prior morphogroup formation**: spadenose shark *Scoliodon laticaudus*, whitespotted bamboo shark *Chiloscyllium plagiosum,*

blue shark *Prionace glauca*, shortfin mako shark *Isurus oxyrinchus,* Japanese topeshark *Hemitriakis japonica*, longfin catshark *Apristurus herklotsi*, blackbelly lanternshark *Etmopterus lucifer*, mandarin dogfish *Cirrhigaleus barbifer*, frill shark *Chlamydoselachus anguineus,* birdbeak dogfish *Deania calcea.* For each species, **we sampled one individual (one specimen per species).** NI and EPMA measurements from all samples were assigned to AI-defined morphogroups after data collection. Supplementary Table 2 details how NI and EPMA data has been distributed between the morphogroups. From these ten species we sampled a total of ten different body locations comprising 37 samples, **where one sample is defined as species+body location. Body locations include** tip of snout (S), abdomen between pectoral fins (B), anterior edge of the pectoral fin (PEC1), flank region above pectoral fin (F1), below second dorsal fin (F3), laterally on the caudal peduncle (CP), lateral flank of dorsal caudal lobe (C2), pre-anal fin region (preANL), anterior edge of dorsal caudal lobe (C1), and posterior region of the pectoral fin (P3) (Supplementary Fig. 2). **For each sample (species+body location), we used between 2–5 denticles to extract data for NI and two denticles to extract data for EPMA (see details in Supplementary Table 2). It was difficult to work with smaller sized denticles, such as those of the spadenose shark, and we had to extract the data from several denticles to reach a comparable number of data points.**"

Reviewer #5 (Remarks to the Author):

The authors have addressed some of my previous questions, and I appreciate their thorough study on shark skin diversity and their efforts in relating denticle morphologies to functions. However, concerning the CFD simulations, there remain several unresolved issues that undermine the solidity of the authors' conclusions regarding drag reduction.

Author response to introductory comment:

We thank the reviewer for the careful reassessment of our work and for the constructive feedback. Although conducting the additional analyses required considerable time and effort, they have significantly strengthened the manuscript. We appreciate the reviewer's recognition of our study on shark skin diversity and the relationship between denticle morphology and function, and we believe the revised version presents a more robust and comprehensive treatment of these aspects.

Comment 3

In the authors' response to Comment 8, they stated:

"We selected Re = 10,000 deliberately because lower Reynolds numbers, such as Re ≈ 1,000— which are typical of swimming sharks—produce fully laminar flow, dominated by viscous forces. In this regime, flow separation and vortex formation are minimal, and the drag becomes largely insensitive to surface geometry. By contrast, using Re = 10,000 provides enough flow complexity—specifically, the onset of separation and vortex shedding—to detect differences in drag caused by morphological features such as ridges, concavities, and spacing."

From this statement, it appears the authors could not find evidence supporting drag reduction at realistic local Reynolds numbers (Re = 1,000), prompting their selection of a higher Reynolds number (Re = 10,000) to better illustrate their findings. However, this selection detracts from the realism of actual shark swimming conditions. The investigation into Reynolds number effects shown in Supplementary Fig. 15B is valuable but does not adequately justify the Reynolds number chosen for the study.

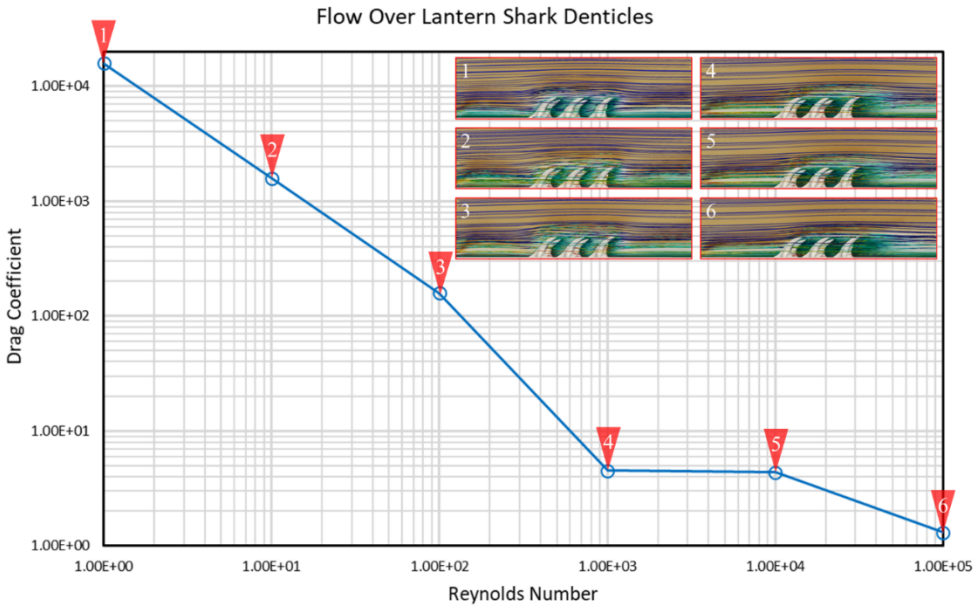
Author response 3

We thank the reviewer for highlighting the importance of using a realistic Reynolds number representative of actual shark swimming conditions. We agree that Re ≈ 1,000 more accurately reflects the typical local flow regime experienced by swimming sharks [1]. To address this, we have rerun all simulations at Re = 1,000, while maintaining the same computational setup as for Re = 10,000. The updated results are summarized in Table RL1, where we also compare the values with those at Re = 10,000 to quantify the differences. In addition, the Supplementary Fig. 15B in the original manuscript has been revised to include the new Re = 1,000 results, showing the variation in drag coefficient for the blackbelly lanternshark at different Reynolds numbers. For clarity, these updated results are also presented in the Revised Supplementary Figure 15B.

Table RL1. Comparison of Terminal Drag Coefficient for Different Shark Denticle Types at Re = 10,000 and Re = 1,000.

Denticles Type	Morphogroup	Re = 10,000	Re = 1,000	Change (%)
Spadenose Shark F1	Classic Ridge 1	0.3030	0.3255	+7.43%
Spadenose Shark C1	Round Ridge	0.2744	0.3338	+21.63%
Shortfin Mako Shark F1	Robust Ridge	0.2800	0.2931	+4.68%
Tope Shark C2	Thick Ridge	1.1955	1.2027	+0.60%

Spadenose Shark preANL	Catshark Ridge	0.2697	0.2942	+9.10%
Spadenose Shark S	Smooth	0.2608	0.2817	+8.03%
Longfin Catshark F1	Irregular Ridge	0.4200	0.4435	+5.60%
Blackbelly Lanternshark F1	Spiky	4.3777	4.4273	+1.13%
Mandarin Dogfish F1	Spiky	1.0888	1.1093	+1.88%
Frill Shark B	Highly Structured	0.3389	0.3531	+4.19%
Birdbeak Dogfish B	Highly Structured	2.5049	2.5751	+2.80%



Revised Supplementary Figure 15B. Variation of Drag Coefficient for the blackbelly lanternshark at Different Reynolds Numbers.

Lowering the Reynolds number from Re = 10,000 to Re = 1,000 increases the terminal drag coefficient by **6.10% on average** across morphogroups (Table RL1), with the largest change for **Spadenose C1 / Round ridge morphogroup** (**+21.63%**). The direction and magnitude are consistent with canonical external flows, where drag coefficient increases modestly as Re decreases within (10^3–10^4). The qualitative ranking of morphogroups is preserved. Consistently, the mesh-resolution and domain-length sensitivity exhibit the same qualitative behavior across both Reynolds numbers, indicating stable numerical convergence. We therefore adopt Re = 1,000 as the primary condition and update **Main Figure 5a, 5c, 5d** and **Main Figure 6a–c** accordingly.

As shown in the Revised Supplementary Figure 15B, the blackbelly lanternshark exhibits only a slight increase in drag coefficient when Re decreases from Re = 10,000 to Re = 1,000. The visualizations of the flow field confirm that wake and separation structures remain similar between the cases of Re = 1,000 and Re = 10,000, as shown in Figure RL1.

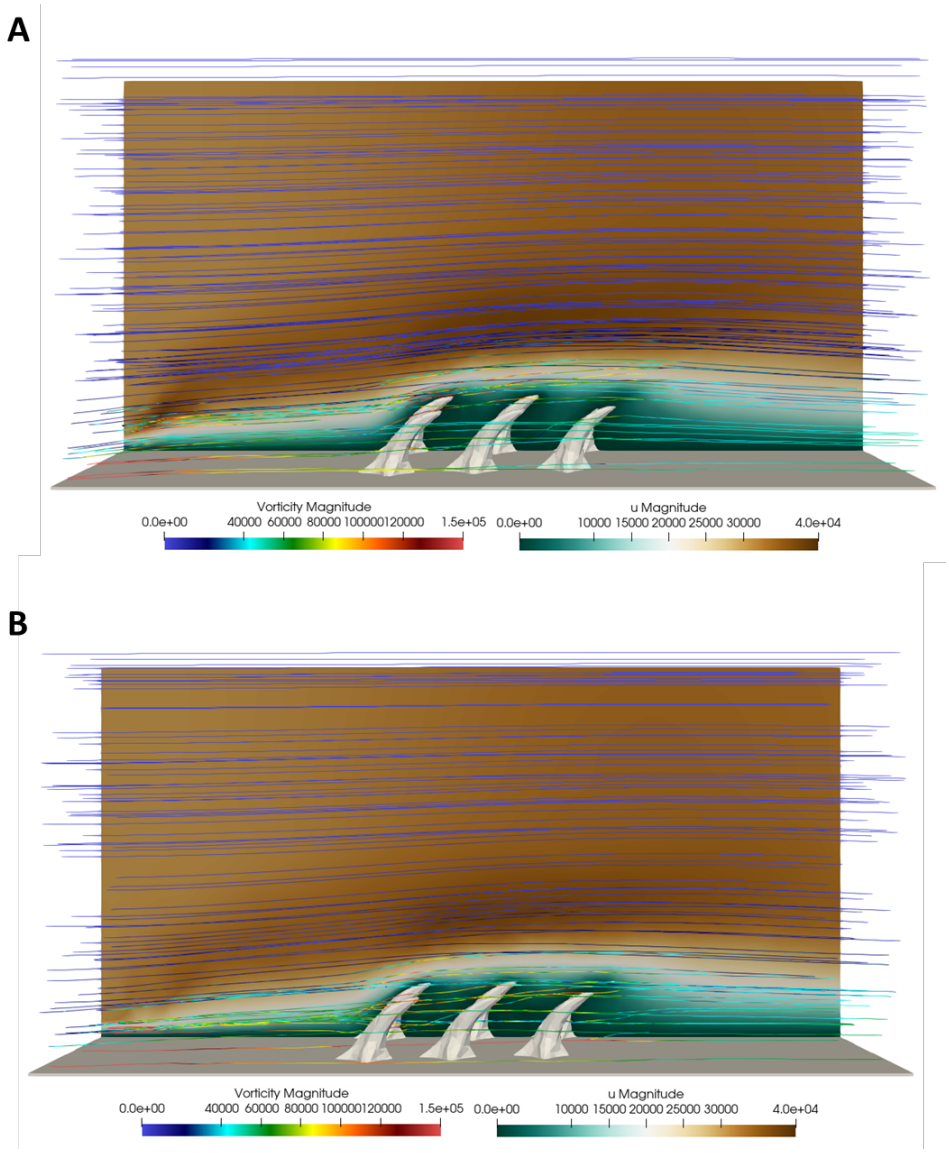


Figure RL1. Perspective view of streamlines colored by vorticity magnitude for (A) Re = 1,000 and (B) Re = 10,000.

We also provide classical flow over spheres analysis, as demonstrated in the Revised Supplementary Fig. 15A (**see Author response 6**), where the drag coefficient varies minimally within this Reynolds number range (from Re = 10,000 to Re = 1,000). This is also consistently observed in the denticle analysis.

Given that Re = 1,000 better represents realistic shark swimming conditions and that the differences in drag coefficients remain small, we fully agree with the reviewer's recommendation. We have therefore revised the study to adopt Re = 1,000 as the primary flow condition, and the Revised Supplementary Fig. 15B has been updated to include the new Re = 1,000 results. This revision ensures that the simulations more accurately capture the realistic hydrodynamic behavior of shark denticles while maintaining consistency across all geometries.

See revised methods in the Supplementary Information and Main Methods with track changes for updated descriptions and equations.

[1] Graybill, M. T., & Xu, N. W. (2024). Experimental Studies of Bioinspired Shark Denticles for Drag Reduction. Integrative and Comparative Biology, 64(3), 742-752.

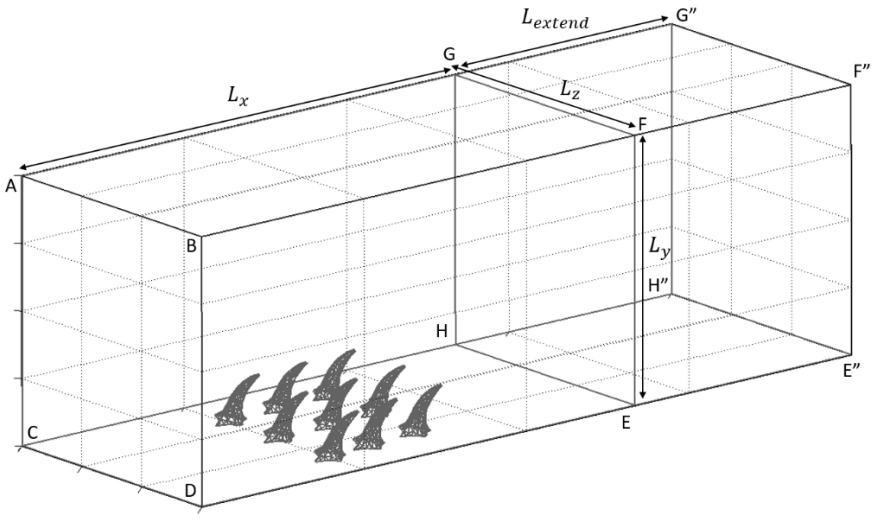
Comment 4

Regarding Response 9: While the zero-pressure boundary condition is generally acceptable when the object is sufficiently distant from boundaries, it may introduce non-physical errors otherwise. The authors noted, "These tests revealed no appreciable difference in drag coefficients." Could the authors specify the exact values of these differences between different domain sizes? Furthermore, it would be beneficial if the authors provided explicit quantitative values whenever claiming, "no appreciable difference" or "appreciable difference."

Author response 4

We thank the reviewer for the valuable suggestion to provide quantitative evidence supporting our statement regarding the outlet boundary condition. To evaluate the sensitivity of the drag coefficient (C_D) to the downstream domain length, we extended the computational domain in the flow direction while keeping its height and width constant. The cuboid computational domain was originally defined with dimensions $L_x = 6$ mm, $L_y = 3$ mm, and $L_z = 3$ mm. To assess the influence of the outlet length, an additional downstream extension L_{extend} was added to L_x in 1 mm increments, resulting in total domain lengths of $L_x + L_{\text{extend}} = 7$ mm, 8 mm, and 9 mm. The newly added Supplementary Figure 12D illustrates the extended domain geometry,

where L_{extend} denotes the added outlet section, and the newly added Supplementary Table 15 summarizes the corresponding drag coefficients, mesh sizes, and computational runtimes.



Newly added Supplementary Figure 13D. Extended flow domain used to assess outlet-distance sensitivity.

Newly added Supplementary Table 15. Effect of downstream domain extension on drag coefficient (C_D).

Domain size (mm³)	C_D	ΔC_D vs 6mm (%)	Mesh (elements)	CPU Time (hours)
6 × 3 × 3	4.377	N/A	275,442	16.2
7 × 3 × 3	4.295	-1.9%	322,215	19.8
8 × 3 × 3	4.249	-2.9%	367,453	23.6
9 × 3 × 3	4.216	-3.7%	413,052	27.5

As shown in newly added Supplementary Table 15, extending the domain results in only a −3.7% total variation in C_D compared with the original case. The incremental changes in the drag coefficient between successive extensions are −1.9% (7 mm), −1.1% (8 mm), and −0.8% (9 mm), indicating convergence of the drag coefficient with increasing outlet length. Meanwhile, the computational cost increases substantially: the mesh size grows from 275,442 to 413,052 elements as shown in the table, and the simulation time rises from 16.2 hours to 27.5 hours, corresponding to a 69% increase in simulation time for less than 4% change in drag coefficient. This confirms that the chosen domain provides a balanced compromise between computational efficiency and physical accuracy. Therefore, to optimize computational cost

without compromising result fidelity, we retained the 6 × 3 × 3 mm³ domain for all reported simulations.

The results of this comparison study are included in the Supplementary Information on Lines 214-229, together with the corresponding geometry, as shown in the newly added Supplementary Fig. 13D, and the newly added Supplementary Table 15.

Comment 5

Regarding Response 11: Aside from the specific mistake highlighted previously, I recommend carefully reviewing the entire manuscript and supplementary materials. The terms "total drag," "viscous drag," and "pressure drag" appear frequently mixed or incorrectly interchanged throughout the fluid dynamic analyses. For example, in the supplementary document (lines 127- 128), it states: "F_D is the denticle surface drag force due to the fluid viscosity," which requires clarification or correction.

Author response 5

We thank the reviewer for this valuable follow-up comment and fully agree that the terminology for drag components must be used consistently. We have carefully reviewed the entire manuscript and supplementary materials to ensure that "total drag," "viscous drag," and "pressure drag" are clearly defined and correctly applied throughout.

In all CFD analyses, the total drag force (F_D) is calculated from the surface integral of the complete stress tensor acting on the denticle surfaces, which includes both viscous shear and pressure (normal) contributions. Accordingly, we have corrected the previous inaccurate description in the Supplementary Information (Lines 106–108) to read:

"**F_D represents the total drag force obtained from the surface integral of the complete stress tensor acting on the denticle surfaces, including both viscous shear and pressure contributions,[…]"**

All related figures, captions, and text have been checked and revised to maintain consistent terminology. We appreciate the reviewer's careful attention to this issue.

Comment 6

Regarding Response 14: I fully acknowledge inherent discrepancies between CFD simulations and experimental results due to differing setups and measurement techniques.

However, at canonical geometries and a Reynolds number of Re = 10,000, results should align closely. Thus, a discrepancy exceeding 30% raises significant concerns about the accuracy of the presented simulations. I also strongly disagree with the authors' statements regarding their validation and numerical simulations. Despite computational limitations making high Reynolds number.

Therefore, if we had used Re = 1,000, it would have precluded any meaningful comparison of denticle shape on hydrodynamic performance predictions challenging, current CFD methods typically provide accurate hydrodynamic predictions at Re = 10,000. Additionally, the authors used a validation case involving forces on a sphere from simulations and experiments, concluding that "although the absolute error is not negligible, the internal consistency of our simulation is preserved." From a scientific computing perspective, such a statement is questionable, especially given the results provided. Absolute errors matter for the current study at Re = 10,000.

Author response 6

We appreciate the reviewer for highlighting the importance of quantitative validation, particularly within the Reynolds number range relevant to our shark-denticle simulations (Re ≈ 1,000 to 10,000). We acknowledge that the earlier validation results showed noticeable discrepancies at Reynolds numbers greater than 100, which directly correspond to the flow regime of interest in this study. To ensure that our computational setup provides reliable predictions within this range, we re-examined and refined the validation for the flow over a sphere using a more accurate numerical configuration and mesh resolution. The refined example will replace the original analysis in the revised Supplementary Information.

In the revised configuration, the computational domain was defined as 26 × 10 × 10 mm³, and the sphere had a diameter of 1.0 mm positioned at (5.5, 2.5, 2.5 mm). To enhance numerical accuracy near the sphere surface and within the wake region, we refined the mesh density in the critical flow zones. The near-sphere region extending from −0.6 to 0.6 mm in all directions was assigned a fine resolution with $D/\Delta x = 20$, while the intermediate region ($-1.5 \leq x \leq 5.5$ mm and $-1.5 \leq y, z \leq 1.5$ mm) used $D/\Delta x = 10$. The outer domain was discretized more coarsely with Δx = 0.5 mm, resulting in approximately 436,906 elements in the new setup. This refinement significantly improved the resolution of the boundary layer and wake

structures, which had been identified as the primary sources of error in the previous configuration.

Figure RL4 presents the updated computational domain and mesh configuration. Figure RL4A shows the original mesh, while Figure RL4B illustrates the refined adaptive mesh near the sphere and in the wake region. Figure RL5 compares the drag coefficients obtained from the original simulations, the revised simulations, and the canonical experimental data across different Reynolds numbers. These figures clearly demonstrate the improved mesh representation and its positive effect on the accuracy of the predicted drag coefficients. We acknowledge that our previous numerical setting may not be suitable, although the overall trend was similar.

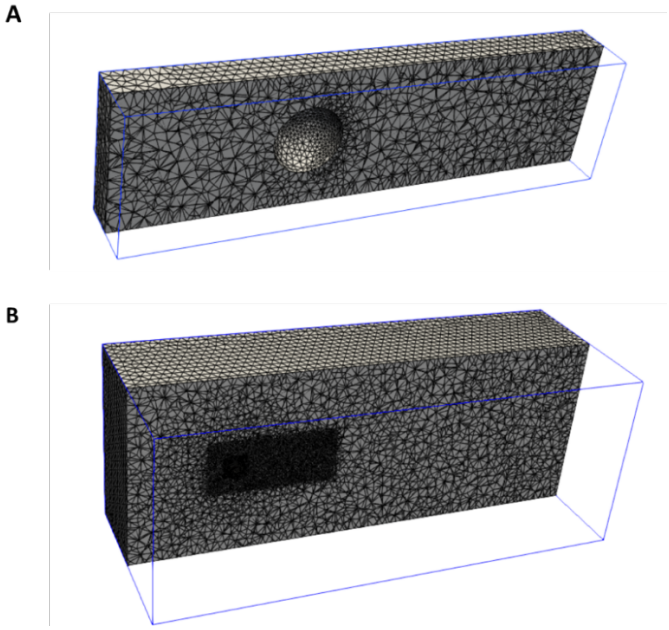


Figure RL4. Computational domain and mesh refinement for the sphere validation case. (A) Original coarse mesh configuration. (B) Refined adaptive mesh.

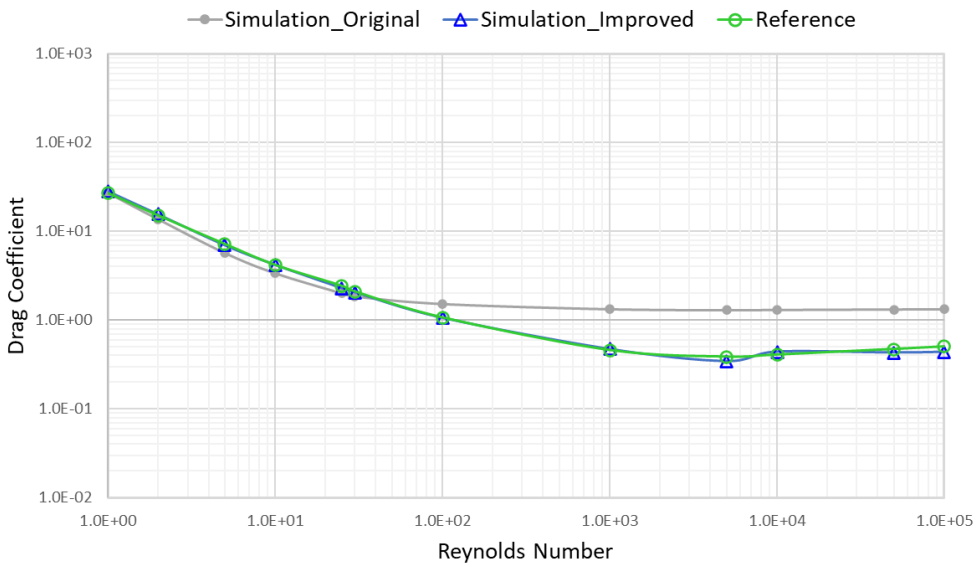


Figure RL5. Drag coefficient comparison between original and improved simulations across Reynolds numbers.

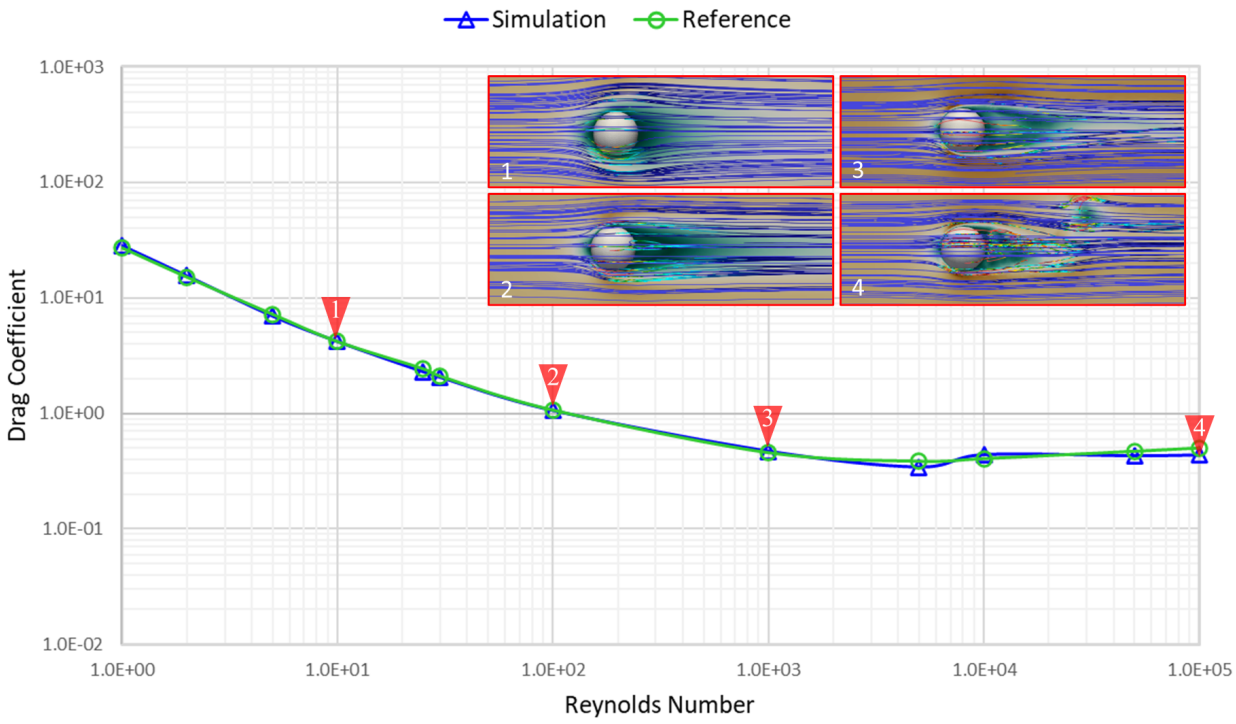
Table RL2 summarizes the quantitative comparison of drag coefficients between the original simulations, the experimental reference data, and the improved simulations. The refined results show a consistent improvement in matching experimental trends, especially at higher Reynolds numbers. In the original setup, drag coefficients were systematically overpredicted for Re > 100, indicating that wake dissipation and surface pressure recovery were under-resolved. The refined configuration corrected these issues by improving the near-wall resolution and wake capturing, resulting in a closer alignment with the canonical experimental values.

Table RL2. Comparison of drag coefficients between original simulation, revised simulation, and experimental reference across Reynolds numbers.

Re	Original	Improved	Reference
1	26.98	28.41	26.98
2	13.62	15.57	15.07
5	5.69	6.98	7.18
10	3.37	4.19	4.18
25	2.01	2.30	2.43
30	1.85	2.05	2.10
100	1.52	1.06	1.07
1,000	1.33	0.47	0.46

5,000	1.30	0.34	0.38
10,000	1.30	0.44	0.41
50,000	1.32	0.43	0.47
100,000	1.33	0.44	0.50

The comparison clearly shows that the improved simulations reduce the overall deviation from experimental data across all Reynolds numbers. The most significant improvement is observed at higher Reynolds numbers, where the refined drag coefficients now follow the expected experimental decay behavior. Overall, the refined validation demonstrates that the earlier discrepancies were primarily caused by insufficient near-surface and wake resolution rather than by the numerical formulation itself. The updated validation confirms that the present CFD setup accurately reproduces canonical drag behavior across a wide range of Reynolds numbers. This provides a robust foundation for shark denticle simulations and supports the reliability of the numerical results presented in this study. The improved simulation data and computational setup have been incorporated into Supplementary Information. **The results have been presented in a revised plot comparing the improved simulations with the experimental data across all Reynolds numbers, as shown in the Revised Supplementary Fig. 15A. Corrections have been made in Supplementary Information on Lines 189-193.**



Revised Supplementary Figure 15A. Comparison of drag coefficients for flow over a sphere at different Reynolds numbers.

Comment 7

Regarding Response 15: Concerning the statement, "Profiles from Mesh 2 and Mesh 3 are nearly indistinguishable," could the authors provide exact quantitative values of these differences? Drawing conclusions without quantitative comparisons undermines the professionalism and rigor of the analysis, a recurring issue in the manuscript.

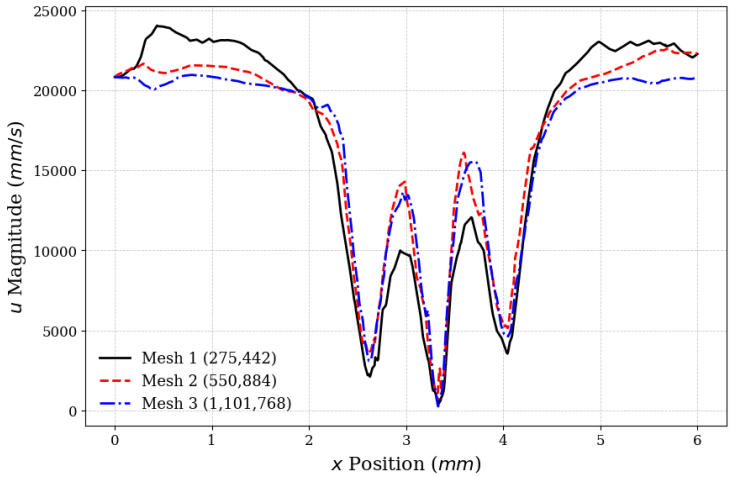
Author response 7

We thank the reviewer for the suggestion to include a quantitative mesh convergence analysis. To evaluate numerical sensitivity, we computed the relative L1 and L2 error norms of the velocity magnitude between coarser meshes (Mesh 1 and Mesh 2) and the finest mesh (Mesh 3, treated as the reference solution) along a representative sampling line above the denticle array. The sampling line, shown in Figure RL5, was positioned slightly above the center of the denticle matrix, extending from (0, 0.45, 1.5) mm to (6, 0.45, 1.5) mm in the global coordinate system. The relative L1 and L2 errors were evaluated using trapezoidal integration along the line as follows:

$$\text{Relative L1 Error } = \frac{\|\boldsymbol{u}_{Mi} - \boldsymbol{u}_{M3}\|_1}{\|\boldsymbol{u}_{M3}\|_1} \times 100\%, \forall i = 1,2 \quad (1)$$

$$\text{Relative L2 Error } = \frac{\|\boldsymbol{u}_{Mi} - \boldsymbol{u}_{M3}\|_2}{\|\boldsymbol{u}_{M3}\|_2} \times 100\%, \forall i = 1,2 \quad (2)$$

where $\boldsymbol{u}_{M_i}$ and $\boldsymbol{u}_{M3}$ denote the velocity magnitude of mesh $i = 1,2$ and the reference solution from Mesh 3, respectively. $\|\cdot\|_1$ and $\|\cdot\|_2$ indicates standard L1 and L2 norm operator, respectively.



Current Supplementary Figure 14J. Velocity magnitude profiles along a horizontal line above the denticle array for three mesh resolutions.

As summarized in the **newly added Supplementary Table 16**, the relative L1 and L2 errors for Mesh 1 were 12.07% and 12.96%, while those for Mesh 2 were 5.26% and 6.17%. These results indicate a consistent reduction in error with mesh refinement, confirming numerical convergence. The velocity magnitude profiles plotted in current Supplementary Figure 14J also show that Mesh 2 and Mesh 3 exhibit nearly identical distributions, demonstrating that the intermediate mesh resolution provides an accurate and computationally efficient representation of the flow field. **The detailed quantitative data and comparison plots have been included in the Supplementary Information on lines 262-290 for completeness.**

Newly added Supplementary Table 16. Mesh information. Mesh elements and terminal drag coefficients (C_D) for Mesh 1, Mesh 2, and Mesh 3. Relative L_1 and L_2 norm errors of velocity magnitude along the sampling line, and norm errors of the C_D for differences relative to Mesh 3.

Mesh	Elements	C_D	Relative L1/L2 Error (%)	L1/L2 Error (%) of C_D relative to Mesh 3
Mesh 1	275,442	12.58	12.07/12.96	26.77/26.77
Mesh 2	596,950	11.20	5.26/6.17	12.86/12.86
Mesh 3	1,063,695	9.93	N/A	N/A

The authors have addressed some of my previous questions, and I appreciate their thorough study on shark skin diversity and their efforts in relating denticle morphologies to functions. However, concerning the CFD simulations, there remain several unresolved issues that undermine the solidity of the authors' conclusions regarding drag reduction.

In the authors' response to Comment 8, they stated:

"We selected Re = 10,000 deliberately because lower Reynolds numbers, such as Re ≈ 1,000—which are typical of swimming sharks—produce fully laminar flow, dominated by viscous forces. In this regime, flow separation and vortex formation are minimal, and the drag becomes largely insensitive to surface geometry. Therefore, if we had used Re = 1,000, it would have precluded any meaningful comparison of denticle shape on hydrodynamic performance. By contrast, using Re = 10,000 provides enough flow complexity—specifically, the onset of separation and vortex shedding—to detect differences in drag caused by morphological features such as ridges, concavities, and spacing."

From this statement, it appears the authors could not find evidence supporting drag reduction at realistic local Reynolds numbers (Re = 1,000), prompting their selection of a higher Reynolds number (Re = 10,000) to better illustrate their findings. However, this selection detracts from the realism of actual shark swimming conditions. The investigation into Reynolds number effects shown in Supplementary Fig. 15B is valuable but does not adequately justify the Reynolds number chosen for the study.

Regarding Response 9: While the zero-pressure boundary condition is generally acceptable when the object is sufficiently distant from boundaries, it may introduce non-physical errors otherwise. The authors noted, "These tests revealed no appreciable difference in drag coefficients." Could the authors specify the exact values of these differences between different domain sizes? Furthermore, it would be beneficial if the authors provided explicit quantitative values whenever claiming, "no appreciable difference" or "appreciable difference."

Regarding Response 11: Aside from the specific mistake highlighted previously, I recommend carefully reviewing the entire manuscript and supplementary materials. The terms "total drag," "viscous drag," and "pressure drag" appear frequently mixed or incorrectly interchanged throughout the fluid dynamic analyses. For example, in the supplementary document (lines 127-128), it states: "F_D is the denticle surface drag force due to the fluid viscosity," which requires clarification or correction.

Regarding Response 14: I fully acknowledge inherent discrepancies between CFD simulations and experimental results due to differing setups and measurement techniques. However, at canonical geometries and a Reynolds number of Re = 10,000, results should align closely. Thus, a discrepancy exceeding 30% raises significant concerns about the accuracy of the presented simulations. I also strongly disagree with the authors' statements regarding their validation and numerical simulations. Despite computational limitations making high Reynolds number

predictions challenging, current CFD methods typically provide accurate hydrodynamic predictions at Re = 10,000. Additionally, the authors used a validation case involving forces on a sphere from simulations and experiments, concluding that “although the absolute error is not negligible, the internal consistency of our simulation is preserved.” From a scientific computing perspective, such a statement is questionable, especially given the results provided. Absolute errors matter for the current study at Re = 10,000.

Regarding Response 15: Concerning the statement, “Profiles from Mesh 2 and Mesh 3 are nearly indistinguishable,” could the authors provide exact quantitative values of these differences? Drawing conclusions without quantitative comparisons undermines the professionalism and rigor of the analysis, a recurring issue in the manuscript.