Response to Reviewer Comments on:
Evolutionary bursts drive morphological novelty
in the world's largest skinks

We would like to begin by thanking all three anonymous reviewers. All reviewers provided carefully considered comments and we have done our best to accommodate those suggestions here. On behalf of myself and my coauthors, we appreciate your time and contributions.

Below please find reviewer comments in green, and our responses in black. We have formatted a page break between each reviewer. Where necessary, we have interjected our responses among reviewer comments.

Thank you for your time and consideration,
Ian Brennan (on behalf of all coauthors)

Reviewer #1: A thoughtful and very interesting article that provides a clear "how-to" for all those interested in unravelling the particular patterns of evolutionary change across a lineage of animals. I had no issues with the methods used and feel only that the aim of the MS could steer away from any supposed disagreements among theoreticians, which just distracts from their method strengths. The MS starts in way that appears to take as a starting point that there is a dichotomy in the way evolution is interpreted with two seemingly competing schools of thought and "little consensus". But there really is no competition. It would be most unlikely to find any theorist who would not credit both types of change as occurring widely. What is generally at issue is the mode of change for a particular character or character set in a particular clade or organisms. What this MS does is provide a neat methodological approach to revealing the historical patterns of change that have brought about the diversity we see in a particular clade. The paper (rightly) makes no claims that this pattern applies to anything other than these lizards, and so the title should more restricted. I'd suggest adding that proviso to the title:

Evolutionary bursts drive morphological novelty in an iconic lizard clade.

Thank you for this suggestion and we agree a more appropriate title would include reference to the focal group. We have adjusted it to read:
"Evolutionary bursts drive morphological novelty in the world's largest skinks"

And also move the introduction somewhat away from any false dichotomies and towards the problems of a suitable methodological approach to reveal what has happened in a particular case. The method is where the paper has its greatest breadth of appeal.

We appreciate the positive comments regarding our manuscript, and hope that our methodology might be useful for others interested in quantifying the pace and pattern of trait evolution.

Other particular comments:

75. "heterogenous" should be heterogeneous

Thank you for catching this. We have corrected the spelling error.

139-146 - The claim that "trait variance accumulated more slowly through the first 40 million years of tiliquine evolution " is empty. The isolation of both Tribolonotus and Corucia mean that there is no pattern apparent regarding their differentiation. They represent two of the most highly modified lizard genera anywhere, both departing strongly from a generalized skink body plan and scalation. Given their specialization compared to some much less modified species nested within the Australian radiation, it must be the case that the period up to 25 Mya saw a near-total pruning of branches in this clade leaving only two highly modified offshoots. Given this, rates and modes of trait variation are not apparent until the more densely populated Australian radiation is reached.

We agree with the reviewer and we have adjusted the text to omit this comment in the results section. Instead we identify the inability to draw confident conclusions about evolution during this period:

> "Little information can be extracted from the early evolution of this group (75-30 mya), as estimates along bare branches leading to *Tribolonotus* and *Corucia* likely do not reflect the evolution of extinct unsampled lineages along these edges."

267. "functional" not the right contrast word here - physiological adaptations are functional too!. Replace with "mechanical".

We thank the reviewer for this comment and have corrected the text to reflect this.

332. Fig. 1.(d) is Tiliqua occipitalis

It sure is! Thank you for catching this species identification.

356. Rather than "typical" tiliquine - which arguably doesn't exist - better to use a descriptive term such as 'orthodox scincid body plan' to describe Liopholis, as the baseline body plan against which other tiliqines are assessed. Good to mention in text somewhere that among tiliquines, Liopholis and Lissolepis species are closest to this least modified scincid body plan.

Good point. We have changed the text to read "illustrated against a more 'typical' scincid body plan".

597. Caeruleulus is all but unpronounceable. Implore the authors to discover a more euphonious moniker. In this Appendix, the generic diagnoses are morphologically weak, especially for the distinction between Lissolepis and Paluarius.

We appreciate the feedback regarding our proposed new genus name. We agree that the previous suggestion was difficult to pronounce and so have changed the proposed name to *Caerulingua* (blue tongue). Hopefully this is a more pronounceable and euphonious option.

We have expanded our description to include a number of additional characters. As currently proposed, the diagnosis provides the best possible combination of characters that are likely to be readily available to someone seeking to distinguish *Paluarius* from *Lissolepis* or other skinks. In discussing these diagnoses we requested the input of one of Australia's leading herpetology morphologists and taxonomists Dr. Glenn Shea. We proposed an alternative option, which was to sink *Lissolepis coventryi* into *Liopholis*; however, his personal comment was that "to me, you can't define *Liopholis* inclusive of *coventryi*, as all the apomorphic defining characters of *Liopholis* (e.g., the single pair of nuchals) are lacking in *coventryi*". To address this, we instead diagnose the monotypic genus *Paluarius.*

Reviewer #2: I have now read the paper 'Evolutionary bursts drive morphological novelty' by Brennan et al. In this paper the authors generate a phylogenomic tree of Tiliquini (previously Egerniinae) and use this phylogeny and a set of 19 linear measurements to test if morphological evolution takes place gradually or in 'jumps'. Using different comparative methods, they find that most of the considered traits evolve in accordance with a Brownian Motion process, but that infrequent 'jumps' (sudden increases of the rate of morphological evolution) generate new divergent phenotypes, e.g., those of the genus Tiliqua. They use these results to make general statements about mode and tempo of evolution, namely that 'morphological traits diverge through evolutionarily discontinuous processes'.

My main concern is that the study is neither hypothesis driven nor does the group of Tiliquini skinks seem well suited to investigate mode and tempo of morphological evolution, due to a very sparse fossil record (the few existing fossils were not included in the analyses apart from calibrations) and their relatively low diversity. Instead of trying to answer obvious open questions about the evolution of Australian herpetofauna (e.g., if Tiliquini skinks encountered ecological opportunity upon their arrival in Australia), the authors chose to instead carry out what appears to be an exploratory analysis of their trait data. As currently framed, there is no way to defend the results against the most obvious criticism: the lack of fossil data. It is apparent from Figs. 3 and 4 that results are all biased towards the tips, suggesting bias from unsampled extinct lineages (the rate of morphological evolution stays constantly low for 40 million years, and disparity during this time is lower as expected under BM regarding all morphological modules). As such, the observed rate heterogeneity and calculated background rates of morphological evolution, respectively, are likely artefacts, however, this issue was not in any way considered or discussed. What is known about the deep-time evolutionary history of Tiliquini skinks, and how does this support the results?

While we appreciate the passionate views of the reviewer, we strongly disagree with their assessment that our investigation lacks value and rigor due to a lack of fossil information. Yes, the Tiliquini is a relatively small clade; however, our investigation focuses on the patterns of morphological evolution that have resulted in a remarkable group of lizards. We feel strongly that studies of tempo and mode should not be limited to clades with extensive fossil records or we would rarely have the opportunity to compare paleontological and neontological patterns of diversification. While it's true that there is a large gap in the deep history of this group (70-35 my), we acknowledge this in the text (Lines 144-147) and focus our attention on the densely sampled Australian clade.

To address the reviewer's claims that a lack of fossil information biases our inferences—including that "observed rate heterogeneity and calculated background rates of morphological evolution … are artefacts"—we carried out additional analyses explicitly incorporating fossil data. While fossil knowledge of the Tiliquini is admittedly "sparse" when compared to macropod mammals or marine vertebrates, it is actually quite good for an extant lizard group, and one of the best represented Australian reptile groups. We estimated a combined evidence phylogenetic tree for the Tiliquini using morphological data from Thorn et al. (2023) in conjunction with our phylogenomic data (see Supplementary Material section *Phylogenetic Analyses and Timetree Estimation*). We included five fossil taxa known from within the crown of this group (*Proegernia mikebulli*, *Proegernia palankarinnensis, Tiliqua pusilla, Tiliqua frangens, Egernia gillespieae*). To investigate the influence of these fossil taxa on our inferences of the evolution of morphological traits we included three of these taxa (*P. mikebulli, T. frangens, E. gillespieae*) in new BayesTraits analyses of head length evolution (Fig.S12). The fossils of two other taxa (*P. palankarinnensis, T. pusilla*) are unfortunately too fragmentary to include. Results of fitting the Variable Rates model in BayesTraits to head length datasets including and excluding fossil taxa provide qualitatively identical results. The number and

position of shifts between the fossil and extant datasets are the same, and estimated evolutionary rates are highly correlated. These results lend strong support to our existing inferences of the morphological evolution of Tiliquini skinks, given our existing knowledge of their diversity. While it's likely that the discovery of additional extinct taxa may provide new insights, the contemporary morphological diversity of this group was rapidly partitioned among clades with a more traditional scincid morphology at the base of the Australian clade (emphasized by similarity between *Proegernia* taxa and *'Lissolepis'*). Because of this, additional fossil information may elucidate *additional* shifts in trait evolution, but are unlikely to erase inferred shifts. This is highlighted by the inclusion of *Tiliqua frangens* in the additional head length analysis. *Tiliqua frangens* has been estimated (Thorn et al. 2023) to have been more than twice the size of its close relative *Tiliqua rugosa* (SVL: 250 mm vs. 590 mm). In order to achieve this disparity in size, however, it's only necessary to increase the estimated evolutionary rate marginally. In comparison, a far higher evolutionary rate (5x the mean/background rate) is required to generate the comparatively huge heads of *Tiliqua*. This result is also relevant to the claim that our results are limited to inferences at the tips of our tree. While a number of shifts are estimated between closely related species, important shifts are also recovered deeper in the tree, particularly along the edge leading to *Cyclodomorphus* and *Tiliqua.* The rapid diversification of these clades appears to have been coupled with rapid morphological change as well.

The reviewer makes an additional comment that "observed rate heterogeneity and calculated background rates of morphological evolution, respectively, are likely artefacts" as a result of "bias from unsampled extinct lineages" along the branches leading to *Tribolonotus* and *Corucia*. While we agree that rates along those edges are likely underestimated, the estimation of downstream rates is not contingent upon those branches. To illustrate this point we reran the VR model in BayesTraits across all four modules limited to the Australian Tiliquini skinks (Fig.S13). Estimated rates and the position of inferred rate shifts is highly consistent when compared to our trees including *Tribolonotus* ssp. and *Corucia*. We also reran analyses that informed Figure 3, limiting them to the Australian clade. Here—as the reviewer correctly points out—mean rates are likely to be underestimated as a result of those two long branches. Reanalysis changes little about the timing of morphological expansion vs. packing, or timing of changes in evolutionary rates. We have included a note to this effect in the figure caption that reads:

> "Figures represent analyses of the Australian clade alone to avoid bias in estimated rates and diversity as a result of long bare branches leading to *Corucia* and *Tribolonotus*. For comparison with patterns estimated from data including these taxa, see Fig.31"

In my opinion, this paper would be better served by being reframed as a taxa-focused study, and by toning down/removing attempts to speak to general theories of evolution. Looking at the data it seems as if there could be a signal of early divergence of traits around the time this group is thought to have entered Australia, although the exclusion of most non-Australian Tiliquini makes this result difficult to interpret. Nevertheless, this would be a highly relevant question, because egerniines are potentially the first group of skinks in Australia (Thorn et al. 2021).

Ecological opportunity is a fascinating topic and one which we have previously studied in Australian reptiles (Esquerre et al. 2020 SysBio; Esquerre et al. 2022 BioLetters; Pavón-Vazquez et al. 2022 Evolution; Brennan et al. 2021 SysBio). Our focus here, however, is to understand the evolutionary patterns of extreme morphological diversity in this group of skinks. We agree that there are signs of early divergence of some traits that are coincident with the rapid crown divergence of this group. These traits and the modules they constitute, however, don't exclusively show early-burst

patterns that are consistent with traditional adaptive diversification as a result of ecological opportunity (see comparative model fitting). In fact, Early Bursts aren't favored to explain the evolution of any individual trait (Fig. S6). Instead, there are many additional nested shifts in trait evolution that suggest a more nuanced and complex history of morphological change. While we agree that there are likely artifacts of unsampled and extinct lineages, particularly along the stem leading to Australian Tiliquini, we believe our modeling of extant diversity still holds value.

As has been pointed out, the fossil record of Tiliquini skinks is limited but has pointed towards some interesting results which support our findings. Important to note is the fact that existing fossil material is largely fragmentary making phylogenetic placement difficult. This on top of already challenging phylogenetic estimation from morphological data (see Thorn et al. 2019/2021/2023) means incorporating this information may lead to unreliable conclusions. The recent redescription of *Tiliqua frangens* suggests that our inference of shifts near the tips of the tree is not exclusively a "bias" from sampling only extant taxa, as suggested by the Reviewer. *Tiliqua frangens* was a massive skink likely a close relative of *Tiliqua rugosa*. Despite the fact that *T. frangens* was potentially twice the size of *T. rugosa*, this change in size can be explained by the background evolutionary rate, suggesting that shifts estimated near the tips of tree are not recovered liberally and without reason.

Importantly, what is lost by lacking an extensive fossil record is the inference of *additional* morphological shifts and increased precision in shift magnitudes, not the inaccurate current assignment of shifts. We believe this is primarily due to what we know about extant tiliquine skinks and supported by fossil information. Thorn et al (2021) identify *Proegernia mikebulli* as "small and gracile" (~100 mm SVL) and that it "share [tooth crown] features with both *Lissolepis* and *Liopholis*". We place this taxon along the stem leading to the Australian clade, and this suggests that early branching Australian tiliquines share plesiomorphic characters. This supports the idea that early Australian group skinks probably exhibited a morphology more similar to traditional scincid design than to deeply nested and highly derived extant taxa. As a result, the inference of morphological rate shifts among extant taxa is primarily a reflection of significant amounts of morphological change occurring over comparatively short time frames, e.g. leading to *Cyclodomorphus/Tiliqua*. The addition of fossils into the crown group does not negate these inferred shifts. Admittedly however, without much more fossil information, little can be said about the early evolution of Tiliquini skinks given very few branching events between their origin and ~30 mya.

In addition, I found the paper very difficult to read due to the absence of any clear hypotheses, an understated methods section, and the use of superfluous and unscientific language. Not only is this unnecessary, and in places majorly oversimplified, but the use of colloquialisms/turn-of-phrases will make the meaning difficult to interpret for those who do not have English as a first language.

We understand that some phrases used in our manuscript may have been difficult to understand for readers who are not native English speakers. We have addressed these as recommended (more clarification noted below in reference to specific comments).

I also have several other issues with the paper listed below:

Just from a quick look there seem to be many misaligned sections in your alignment. I suggest to thoroughly check your bioinformatics pipeline and redo your alignment if need be.

To make sure that alignment errors have not had downstream impacts we undertook an extensive realignment exercise. We started by returning to the pseudo-reference genome files for each sample (fasta file of best assembled contig per each AHE target). We downloaded the complete CDS of

Anolis carolinensis from NCBI and used a blast search to identify exonic targets in our AHE dataset (n=358). We then downloaded additional squamate genomes as well as Sphenodon and Gallus, extracted these targets, and aligned them in the correct reading frame using MACSE. We added our Tiliquini sampling to these alignments using MAFFT (see Supplementary Material section *Alignment Specifications*). We again estimated individual locus trees with IQTREE and a species tree with hybrid-ASTRAL. The topology of this tree is nearly identical to the one originally estimated for the group, with the exception of the position of *Lissolepis coventryi* outside the remaining Australian Tiliquini. Low support for the placement of *L.coventryi*, however, means the topology is fundamentally consistent with that originally presented.

Multiple sequence alignment is a notoriously difficult task and given the size of our dataset, some sections of misalignment are an unfortunate side-effect. At the start of this research project we experimented with various alignment strategies (MAFFT, MACSE, MAFFT+MACSE refinement, etc). Resulting alignments produced qualitatively identical species tree topologies with some variation in estimated branch lengths of gene trees (as expected). As a result, we proceeded with the *refineAlignment* strategy specified in the Supplementary Materials. Because divergence dating realistically has a subjective element (see Bromham et al. 2017 *Bayesian molecular dating: opening up the black box*) we accept a reasonable error around our divergence dating estimates as a result of many compounding assumptions and choices we must make. We feel confident in our inferences of trait evolution however, as shift placement and magnitude are similar between the first presentation of this project (initial submission) and this submission, despite using trees with differences in absolute ages and sampling.

In your Astral phylogeny (and in your alignment) only one Tribolonotus species is present, whereas in the 'trimmed down' version of the tree that was used for the morphological analyses there are four. I didn´t find any explanation for this or how these taxa were added, after (?) generating the phylogeny.

We appreciate the reviewer bringing this to our attention and regret this omission in our methods. Molecular alignments include *Tribolonotus gracilis* and *T. pseudoponceleti*. Two additional taxa *T. blanchardi* and *T. ponceleti* were added to the timetree based on topology and ages from a timetree of Squamata recently published by Title & Singhal et al. (2024). This information is now included in the Supplementary Materials section *Phylogenetic Analyses and Timetree Estimation*.

The figures are not cited in order in the text which makes it very difficult to follow.

Figures 1-4 are first referenced in order in the text. Unfortunately the first reference to Figure 6 does precede the first reference to Figure 5. We have now fixed this in the text.

I was surprised to see no citation of E. Vrba's work which is clearly significant to the topic.

We appreciate the suggestion of E. Vrba’s work, as she has certainly been instrumental in many elements of our understanding of evolution, particularly palaeoecology, selection, and the idea of “exaptation”. However, we’re uncertain of how to link her work to our own. We are happy to receive recommendations of Dr. Vrba’s work which may be particularly relevant to this manuscript.

Have you tested how the outgroup calibrations impacted the node ages of your tree?

In our new dating strategy outgroup node priors are based on posterior estimates from our combined evidence analysis in BEAST. This BEAST analysis incorporated only tip dates of sampled fossils and no node calibrations, and as such should not be biased by prior placement or shape. Estimated ages from BEAST analyses were used to generate node priors for MCMCtree analyses.

Figure 1: What is the meaning of the background colours in the ternary plot?

Background colors were intended to display a "heatmap" like effect to emphasize the position of most inferred points in the ternary plot. We recognize the color scheme made this confusing and have removed this element of the figure.

In Figure 2 you show results of reconstructed slopes of PC1 and PC2 values between (ancestral) nodes. With PC1 being the independent variable that explains most of the variation (80%) in the morphological data taken exclusively from extant specimens, it is expected that reconstructed changes between ancestors of this group occur mainly along this axis. I think this is a circular result and I don´t really see a connection to the gradual vs. pulsed question.

The presentation of the majority of change occurring along PC1–the major axis—is expected and not intended as a surprise. Ancestral trait values were estimated for individual traits, then run through PCA, so circularity should not be an issue.

Nonetheless, we have clarified our methodology to show how we believe it is reasonable. We start from a dataset of all traits, including estimated trait values at ancestral nodes which were inferred using our rate-transformed BayesTraits trees (see Materials and Methods section *Phenotypic Analyses*). We then ran PCA on all traits together to reduce the dimensionality of the data. In this way, identifying PC1 is a result of analyzing our data, and not part of the data itself (as circularity would imply). We then identify change between parent and child node pairs as primarily elaborative or innovative by estimating the angle/slope of change of the line connecting them in Euclidean space and indicated by color (Fig.S15). We identify the strength of this change as the distance between the points and indicated by color saturation. While there is an expectation that nodes will have a greater spread along the x axis as a result of PCA, this is informed by what we know about morphological variation of extant tiliquine skinks. This is reflected in the greater saturation of colors on elaborative branches indicating greater distance traveled between nodes along this axis. In short, it is the direction and not the distance between nodes which dictates the categorization. As a result, many branches (including internals) show change primarily along the innovative axis, which indicates our ability to recover these trends. Additionally, while the major trend (variation in PC1) is interesting, more importantly are the trends of subgroups which highlights groups that adhere strongly to the group-wide trend (*Liopholis*) and groups that depart strongly (*Cyclodomorphus*).

It is not mentioned which individuals were kept in the trimmed down tree for T. scincoides where the subspecies T. s. scincoides is paraphyletic.

We kept only a single *T. scincoides* and trait values were averaged across all subspecies, as an effort to represent the species.

How does the choice of outgroup taxa influence the background rate of evolution? I suggest re-running the analyses without outgroups.

All analyses were run without outgroups, so analyses focused on the Tiliquini alone.

I`m sceptical that using isolated fossil jaws resembling certain modern species for fossil calibrations of the respective species is a good idea, when these have not been included in a cladistic analysis. What about T. cf. scincoides of the Curramulka Local Fauna (Pledge 1992)?

Skepticism around the phylogenetic affinities of fragmentary fossils that have not been included in rigorous cladistic studies is valid and appropriate. We have now excluded these fossil taxa from our assessment, and in their place we have run combined evidence analyses of morphological and molecular data to resolve the position of included fossil taxa and estimate more accurate divergence dates.

What is the scale of colors in Fig. 2? What is the threshold for a branch being colored in either blue or red when a branch shows changes along both PC axes? This needs to be explained.

We regret that we did not include a complete explanation of the interpretation of this figure in the corresponding legend. The color scale indicates the Euclidean distance between parent and child nodes travelled along a given branch, as an indication of how much variation has occurred. We now indicate this information in the legend for Figure 2. The threshold for determining the primary direction of change (elaborative vs. innovative) are now figured in Supplementary Figure S15 and explained in the Supplementary Materials section *Modelling Trait Evolution and Disparity*. For a quick reference we provide that text explanation here:

> "We then identify change between parent and child node pairs as primarily elaborative or innovative by estimating the angle/slope of change of the line connecting them in Euclidean space and indicated by color. For branches which exhibit both elaborative and innovative change (likely most branches exhibit some combination of both), the primary trend is a discrete binary state described by the angle of change (innovation: >45°<135°, >225°<315°; elaboration: >135°<225°, >315°<360°)."

Further, we show in the supplementary figure that a more conservative coding of this threshold does not substantially impact the conclusions from this figure.

In general, I find the methods section too short, the methods not well explained and therefore it is very hard to follow what you did.

We can appreciate that the methods section in the main text is unfortunately brief, however this is a common casualty of abbreviated journal formats (Current Biology articles have a word limit of 5,000 words). To address this, we include a vastly expanded version of the methods appears in the supplement, see sections:

- *Alignment Specifications*
- *Investigating Data Completeness and Informativeness*
- *Morphological Measurements*
- *Phylogenetic Analyses and Timetree Estimation*
- *Modelling Trait Evolution and Disparity*
- *Concatenation and the Anomaly Zone*

- *Topology Tests*

The node ages of your phylogeny differ significantly from those of Thorn et al. (2019), and e.g., the age of Egernia in your tree equals the minimum age for E. gillespieae, which is nested within the genus in Thorn et al. (2019). This needs to be at least discussed.

To address this comment we have undertaken an extensive combined evidence dating exercise, outlined above and in the supplementary section titled *Phylogenetic Analyses and Timetree Estimation*. In reference to the minimum age of *E. gillespieae*, we do not recover this taxon inside the crown of *Egernia* and instead consistently recover it along the stem leading to *Bellatorias+Egernia*. While we appreciate that there is some conflict between the dates in our combined evidence analyses and molecular data alone, we anticipate that much of this is driven by rate heterogeneity in the morphological dataset that may be inflating divergence times in the combined analysis (see Fig.S7).

There is only very little information about the bioinformatic processing of your sequence data, and almost no sequencing stats are reported. You should report e.g., the total number of reads per sample, and of that, the % of on-target reads, the % of duplicates, and the total raw data retained after processing, the average sequencing depth per sample, and some statistics about the completeness of the final alignment (e.g., % of missing sites).

We appreciate the Reviewer's interest in our raw data and processing practices. Many of the requested statistics regarding alignment completeness and occupancy are available in the *Alignments/Tiliquini_Summary_PerLocus.csv* file, which is available in the GitHub directory. In addition, we have summarized some basic statistics of interest in the supplementary materials (alignment length, sequencing coverage, reads per sample, missing data, sample occupancy, locus informativeness) including several plots of this information. To avoid generating an overwhelming amount of supplementary material we have chosen not to include sample specific statistics (read depth, duplicates, % on target) as these statistics are not reported commonly and are of interest to very few readers. All of our data are readily available and we are happy to share any additional information including raw read files.

Ln 12-14; These sentences are not helpful or scientific. What is the purpose of referring to some animal traits as mundane?

We include these sentences as a general interest introduction to our paper. We use the adjectives "mundane" and "bizarre" to highlight that traits exist on a broad spectrum we as scientists and the general public observe every day. To address the reviewer's comments we have restructured the first few sentences of our abstract to be more concise.

Ln 15; How does this connect to previous sentences?

We have adjusted our abstract to better link the introduction with our primary question.

Ln 15-17; There have been many studies of evolutionary mode of many traits across many taxa. Here the authors pose a large-scale question of whether 'most traits' evolve by gradualism or punctuated evolution; which can't be answered by looking at one clade of Australian skinks.

While it's true that looking across a single group does not necessarily produce a pattern that can be applied to all skinks, or even all animals, we provide observations that can be compared across groups and a methodology that can be applied elsewhere. Nowhere in the text do we suggest that large-scale investigations of evolutionary tempo/mode have not been undertaken, in fact, we cite several in the text (see Cooney et al. 2017; Ronco et al. 2021). These are fantastic studies against which we compare our results, just as we mention above. We believe our study builds on those ideas and experiments presented elsewhere by expanding the trait dataset to look across different body regions, which allows us to compare temporal patterns among clades and traits.

Ln 22; only 19 traits are used in the study.

The reviewer is correct that we analyzed 19 traits from 21 traits that we collected. We have amended the text to reflect this.

Ln 25; how is this inconsistent with punctuated equilibria?

Following Gould & Eldredge's (1972) definition of punctuated equilibrium, species traits do not "change in any appreciable way" during their geological history (i.e. along the branches). Under this idea trait change is confined to speciation events (i.e. at nodes). Our modeling framework does not tie trait change to cladogenesis strictly, allowing change to occur along branches, with varying evolutionary rates punctuating an otherwise gradual process.

Ln 26; define punctuated gradualism?

Following Malmgren et al. (1983), punctuated gradualism describes a pattern where periods of evolutionary stasis are interrupted by rapid phyletic change without requirement of cladogenesis. The term is uncommonly used but we apply it here to explain the general pattern of gradual trait change interrupted by periods of rapid trait change that are in contrast to the background process. In the manuscript (Discussion: *Evolving Novel Phenotypes by Bursts*) we describe our use of this term as such:

> "Instead, we find broad support for a model in which the background process of evolution by a random walk (Brownian Motion) is punctuated by bursts in evolution that result in jumps to new adaptive zones. We consider this process more akin to the "punctuated gradualism" model of Malmgren et al. (1983). While we recognize that our inferred process is not entirely consistent with Malmgren et al. (1983) who proposed punctuations between evolutionary stasis and gradualism, we suggest that these distinctions may instead be indicative of similar processes occurring at different scales along the micro-to-macroevolutionary continuum."

Lns 48-50; unnecessary colloquialisms/turn-of-phrases, e.g., 'clumpiness', 'line of least resistance' make meaning difficult to interpret for those who do not have English as a first language.

We appreciate the reviewer's comments, however we consider these phrases useful and familiar as they are often used in evolutionary biology literature. "Clumpiness" is how this pattern is described directly by the authors cited (Deline et al. 2018). The phrase "line of least resistance" is also a well established term used in genetics and evolution to describe phenotypic evolution along the axis of greatest genetic variation (Schluter, 1996).

Ln 87; modest species richness, of which is probably not enough to test the question of gradual vs. pulsed evolution with any real level of rigour. Especially given absence of fossils.

We agree that the use of "modest" is subjective and fluctuates depending on the group of interest. The focal clade (Tiliquini skinks) is not terrifically large, though the Reviewer seems to insinuate that studies of tempo and mode are only valuable if investigated at great phylogenetic scale. This is untrue, and varied patterns at shallower scales are often missed in broad-scale investigations. Fossils are indeed valuable to our understanding of patterns of evolution, but some groups just do not have an extensive fossil record yet are still worthy of study (for example birds have a poor fossil record but hundreds of studies on their evolution have been published). To further our point we show that the inclusion of known tiliquine fossils does not qualitatively change our inferences (see Fig.S12).

Ln 90; What is 'pleisomorphic grade'?

This phrase was intended to draw attention to the nested nature of highly derived phenotypes within a generally conserved group. We note that this phrase was commented on by another Reviewer and to address this we have rephrased the term. The sentence now reads "Additionally, tiliquines are ideally suited for studies of the tempo and pace of morphological evolution because they have been suggested to show a morphological gradient with highly derived phenotypes nested deeply within a generally conserved body plan."

Ln 92; how can you determine if their morphological deviations are 'remarkable'? There is no benchmark here. Why ask this question if you can't address it?

Here we use the phrase "remarkable" as shorthand for identifying morphologies that exceed what we expect under a null model. This is specified elsewhere in the text (e.g. "To determine if our observed patterns follow a null expectation of the accumulation of disparity through time we simulated univariate and uncorrelated and correlated multivariate datasets for each trait and module applying parameter estimates from observed data for theta, sigma, and covariance."), but phrased differently here. We can in fact directly address this by modeling the evolution of our traits, comparing it to evolution under our null model (constant rate Brownian Motion) and identify outliers. We even quantify the amount those outliers deviate from the modeled null process allowing us to remark on just how remarkable they are.

Ln 100; what are your hypotheses, why are you testing this in this group? What patterns would you expect to observe in the present analyses under different scenarios (gradualism, jumps) and do these differ from your results?

We explain our reasoning for choosing this group in the text:

> "tiliquines are ideally suited for studies of the tempo and pace of morphological evolution because they have been suggested to show a morphological gradient with highly derived phenotypes nested deeply within a generally conserved body plan."

This structuring of morphological diversity among derived clades with more 'traditional' morphologies clustering at the base of the tree provides an interesting basis for understanding how trait evolution progresses to reach such extremes.

We are not entirely sure what the Reviewer means when they ask about expected patterns under different scenarios. We can easily visualize and anticipate what evolutionary rates through time would look like under a constant rate BM (flat line) or Early Burst (linearly or exponentially declining line) process, but through our modelling exercise we identify that these processes are not favored and unlikely to have given rise to the diversity we see today. Similarly, we could visualize how traits evolve under some of these varied models (Fig.1 below), with values extracted from empirical model fits. In fact, Figures 3 and 6 show exactly this comparison between what trait evolution might look like under an alternative model (BM/gradualism). In Fig.3 the left and central panels show differences in how disparity has accumulated through time under the VR model versus under a constant rate BM process. In Fig.6, we show how individual traits change along branches under these two competing modelling hypotheses. What's apparent is that variance accumulates more consistently under a single regime BM model, whereas the accumulation of variance is largely partitioned into a small number of branches in the VR model.

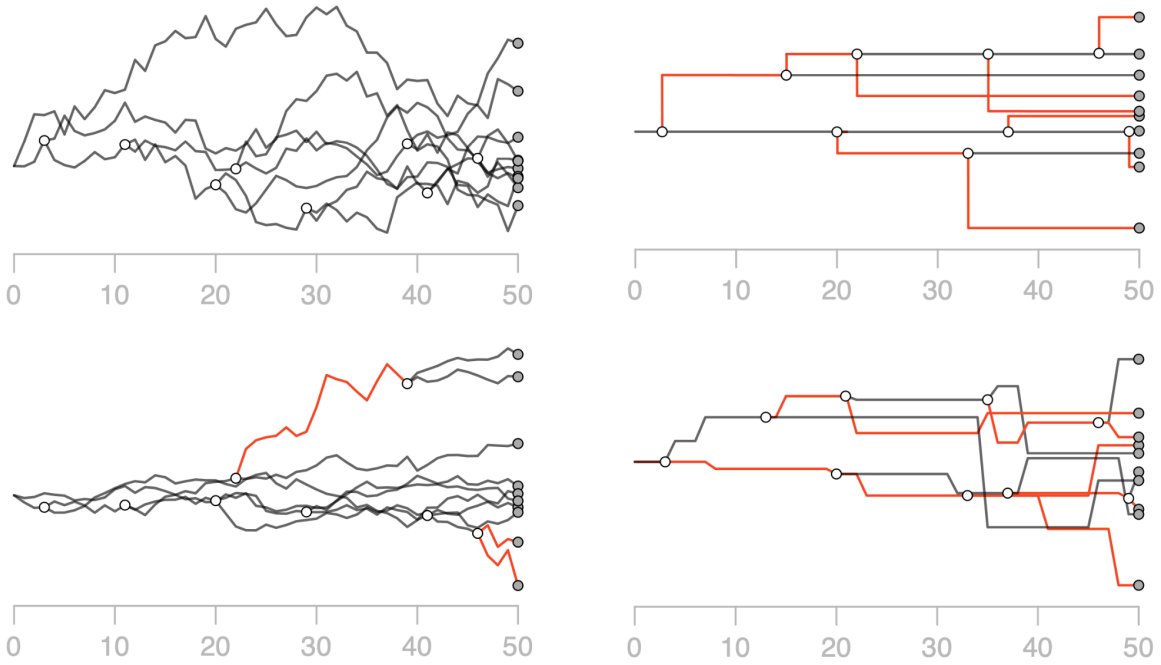

Figure 1. Trait evolution under four alternative modes. Clockwise from top-left: (1) unbiased random walk following Brownian Motion, (2) punctuated equilibrium in which trait stasis is interrupted by jumps in trait values at speciation events (orange branches), (3) pulsed evolution following a Levy model in which trait stasis can be interrupted anytime at or between speciation events by trait jumps, and (4) a relaxed BM model (Variable Rates) in which jumps in trait values are the result of increased evolutionary rates. White circles represent internal nodes and grey circles are extant species.

Ln 102-103; how can this result be meaningful in the absence of fossils?

It's true that fossils have the ability to add information and precision that can not be gleaned from extant-only investigations, and they are key to fully understanding patterns in biodiversity. This includes fitting some categories of trait models (e.g. BM with trend). That said, there remains a wealth of information we can collect in the absence of fossils. Our estimates of ancestral trait values and patterns/rates of morphological evolution provide valuable hypotheses about trends in tiliquine skink evolution that we hope will form a basis for when new fossil material becomes available. For many organismal groups there is limited fossil material, however, we believe that this should not prevent

researchers from investigating evolutionary patterns in these groups. For example, modern birds have generally poor fossil record but there remain hundreds of studies on their evolution. We agree caveats exist and acknowledge these in the text.

Ln 109; are you actually measuring the anomaly zone or just guessing?

We applied the framework of Linkem et al. (2016) to identify branches which fall within the anomaly zone based on demographic signatures and coalescent histories. We visualize these results in the supplementary section *Concatenation and the Anomaly Zone*. To clarify this point we have added a reference to how we determined anomalous branches in the manuscript.

Ln 130; what happened to 21 traits?

We regret the confusion we caused in our explanation of morphological data collection. We collected data for 21 traits directly, but after parsing some (dropping some, splitting some) we finalized a set of 19 traits for analysis. To eliminate this issue we have rephrased the data collection section of our methods to focus on just the 19 focal traits.

Ln 159; these are not moments.

We did not intend the mathematical definition of "moment", to clarify this, we have changed the language to "instances" and "periods".

Ln 167-169; This sentence takes space but adds nothing.

This sentence acts as a segue.

Ln 179-180; 'tremendously' and 'impressive' seem overblown.

These are subjective terms but skink richness is extremely high relative to most other squamate groups (save snakes, geckos, iguanians). In fact, Scincidae is the second largest squamate family (second only to Colubridae), as well as the second largest family of terrestrial vertebrates (Chapple et al. 2023). Regardless, we have changed these phrases to make the comparison of richness more objective.

Ln 206-207; in the intro you state that most trait evolution conforms to BM…

In the introduction we note that observed measures of trait diversity typically *fall short* of BM expectations, suggesting that diversity likely *does not* accumulate following a constant gradual Brownian process. In the discussion we highlight that instead, there is growing evidence that phenotypic diversity instead accumulates via rapid bursts and not constant gradual change.

Ln 218 - 223; over generalisation of results.

Without more specific recommendations we are unable to address this comment. In the section referenced by the reviewer we describe in simpler terms the frequency of preference for jump/burst models (13 of 19 traits) and how these models may be extrapolated to Simpson's view of the macroevolutionary landscape. We do not suggest that these patterns are applicable at any larger

phylogenetic scale, but do describe how they can be interpreted from our data (visualized as large trait changes along individual branches).

Ln 244; again, unnecessary colloquialism.

We have amended the phrase.

Paragraph 'Evolving Novel Phenotypes By Bursts' lacks a clear argument in my opinion.

Our argument stands that patterns of morphological evolution are heterogeneous and constitute a mix of gradual and pulsed processes, with pulses contributing significantly towards the formation of novel phenotypes. To emphasize this, we have made it more explicit in the leading paragraph of this section.

MCMCtree: there are multiple representatives of the same species in the dated phylogeny, which is usually not recommended for this approach due to the tendency to overinflate ages at the tips, which has consequences for the relative ages in the rest of the tree. For what reason were within-species splits included in the molecular dating?

Within species splits were included to assess divergence among known divergent populations and described subspecies. For most of these subspecific taxa (*Egernia stokesii* ssp.*, Cyclodomorphus melanops* ssp.*, Tiliqua gigas* ssp.) subspecies are geographically disjunct and unlikely to show contemporary gene-flow. In addition, distinguishing between species-level and subspecies-level splits is often a subjective exercise for taxa lacking focused population genetic or ecology studies. Molecular clock violations resulting from intraspecific and interspecific sampling should largely be absorbed by the relaxed molecular clock applied (clock = 2; ILN--independent log-normal) which allows individual branch rates to be drawn from a common (but diffuse) distribution. Investigating intraspecific sampling using Bayesian divergence dating methods like MCMCTree is a common practice (see https://doi.org/10.1093/sysbio/syp082; https://doi.org/10.1093/molbev/msab311).

In our updated dataset we exclude most intraspecific samples but retain samples which are of direct interest (*Egernia stokesii* ssp., *Cyclodomorphus melanops* ssp.). To assess the effect of intraspecific sampling on estimated divergence dates we ran additional MCMCTree analyses with no intraspecific samples. Analyses including subspecies do not inflate ages in any meaningful way, see figure below. The orange tree includes subspecies while the blue tree does not. 95% confidence intervals for estimated node ages overlap for all comparable nodes between the two trees.

Ln 379; how many exons were from the same gene? Were they treated as one genealogy or separate?

Of 388 AHE loci, we identified 358 as exonic in origin and of these, 25 targets appear to be separate exons of shared genes (APOB, DSP, DST, FAT4, HIVEP2, IRS1, KAT6B, PLXNA1, RBM15, SHANK2, USF3, ZNF536). Due to the limited number of linked targets, we treated each target AHE locus independently. To investigate further, we removed linked targets from our dataset and reran hybrid-ASTRAL. This produced a near-identical topology, confirming our expectations that this limited number of linked targets had little effect on our species tree topology. We have included a supplemental table that indicates the mapping of AHE targets to squamate CDS (Alignments/Tiliquini_AHE_CDS_RBH.csv). To identify exonic matches in our AHE dataset we downloaded the full RefSeq CDS of *Anolis carolinensis* from NCBI and then used the reciprocal blast feature in the R package *metablastr* (Benoit & Drost, 2021) to compare our AHE targets to CDS loci.

Ln 380-381; should be substitution model.

Substitution models are models of molecular sequence evolution. We have clarified this in the text.

Ln 382; All ML programs will arbitrarily resolve polytomies in gene trees, did you collapse these before running ASTRAL? This needs to be stated in the methods.

Instead of collapsing splits with low support via a subjective threshold (e.g. <80 BS) we used the threshold-free weighting scheme implemented in ASTRAL (Zhang & Mirarab, 2022). We applied the 'hybrid-ASTRAL' design which down-weights quartets with low support or long terminal branches. We believe this is a preferable and reproducible design.

Ln 383; what locus partitions? If all exons are assigned their own model this will be over-parameterised. Partitioning into rate categories is preferrable.

Each locus was assigned to a separate partition and ModelFinder estimated and applied the best fit model for each partition. While it's possible that this may result in overparameterization, we estimate identical topologies estimated under maximum likelihood regardless if data are partitioned into a single partition, four rate-based partitions, or individual locus partitions.

Ln 390; there is no information about the setup of the MCMCtree run. What parameters, how many iterations was it run, with what sampling frequency? Did you run your MCMCtree run with priors only, and compare the prior to the posterior sampling? These comparisons should be shown in the supplement.

We have expanded our supplementary materials and methods to incorporate the requested information. We include a plot of the prior and posterior age estimates for each node.

Reviewer #3: The study presented here provides an interesting concept of how to interpret elaboration and innovation in phenotypic change, and detecting which of these concepts are predominantly observed in a case study of skinks. Their case study also provides high taxonomic sampling and a novel genomic time calibrated evolutionary tree for the group. They conclude by showing how phenotypic disparity and rates of evolution have changed across time, clades, and inferred evolutionary modules in this group. The comprehensiveness of the analyses employed and thoroughness dedicated to investigating the data available are to be commended, and I believe this will be an extremely important and exciting contribution. This study will be a major source of citations in the field of macroevolution and phenotypic evolution.

We appreciate the very positive review and recognition of our work. We hope that this manuscript lives up to the expectations provided by this reviewer.

Having said that, there are aspects of the methodology that could be clarified by the authors, especially regarding their approach to infer evolutionary rates. There are also important corrections or adjustments they need to address in the introduction to make some statements accurate. But overall, I considered those relatively easy changes to be performed. Please find my detailed comments below.

MAJOR COMMENTS:

Introduction:
Lines 35-39: it's not entirely accurate to say that trait evolution through time "almost always fail to exceed expectations of a neutral model". Results on the predominant mode of evolution are highly dependent on the taxonomic and chronological scale of sampling. Whereas studies focusing on shallow divergence times and low taxonomic scales (similar to the case study presented in this study) do tend to show a predominant Brownian motion model and rare instances of early bursts (Harmon et al. 2010), studies focusing on much longer evolutionary timescales or broad taxonomic sampling (e.g., across families or orders) frequently find more variable modes, including early bursts, punctuated change, among others (Halliday et al. 2019; Simões et al. 2020). This is natural considering that the former represent a spyglass look on the evolution of trait trajectories, but when observed under a bird's eye perspective may reveal more complex patterns of evolution. Finally, there are methodological issues that prevent the detection of alternative evolutionary modes and tend to favor Brownian motion as the preferred mode using maximum likelihood approaches (Slater & Pennell 2013), or when there is no consideration of fossil data (Slater et al. 2012).

These are really valuable observations, and we certainly agree that the predominant inferred mode of morphological evolution is sensitive to phylogenetic scale. In our original phrasing we were perhaps too general. We have amended this first introductory paragraph to incorporate mention of phylogenetic scale and also address the comment regarding Line 42 below:

> “Despite the dramatic variety we see in nature, on deep timescales, measures of these differences (variance, disparity, et al.) often fail to exceed the expectation of a neutral model where the accumulation of variation is the result of a random evolutionary walk through time (Brownian Motion). In other words, Earth's flora and fauna realize only a fraction of all possible shapes and sizes. This diversity undershoot suggests bounds on phenotypes as a result of genetic, functional, or developmental constraints. If constant and gradual processes are unlikely explanations for the accumulation of trait diversity at deep timesecales, then how important are heterogeneous processes at shallower ones?”

Line 42: " If this is the case, then how do novel morphologies arise?" This question implies that no morphological novelties may arise from Brownian motion, which is not really the case. Brownian motion does not mean absence of evolution, as frequently implied in the literature. Fluctuations around a stable main trade value are different from a stable trait value through time (Harmon 2018).

We agree that new morphologies may arise via Brownian Motion, and did not intend to suggest that BM implies stasis, as it's apparent from our results that plenty of trait evolution occurs under BM processes. Our question instead was meant to ask how ***novel*** morphologies appear in light of general disparity undershoots. Not small changes to size, but dramatic changes to shape. We embrace a less restrictive definition of novelty than most outlined in Erwin (2015), instead encapsulating the idea of major transitions among adaptive peaks following Hallgrimsson et al. 2012 and Peterson & Muller 2013. Basically, we are asking if some variation to tempo and mode are driving the accumulation of disparity. We have changed the introductory paragraph to clarify this.

Lines 53-55: authors should make mention of different (and previously established) definitions of evolutionary innovation¬—e.g., (Erwin 2015; Simões et al. 2020). Guillerme et al. are hardly the only study investigating this topic in recent years and their definition conflicts with some previous utilization of this term in evolutionary biology.

These are very valuable references and we thank the reviewer for bringing them to our attention. We did not mean to suggest that Guillerme et al. and Endler were the only researchers who have worked in this space, but instead we were highlighting that we find their terminology useful and interesting. Nor do we mean to confound the use of "innovation" (sensu Endler/Guillerme) with "key innovation" (sensu Erwin). To address this we have incorporated the suggested references, as well as others, and better placed our results in context with paleontological results in the Discussion. In an effort to highlight that this is just one of several definitions of "innovation" we have also changed the text in the introduction to read:

> "These competing processes <u>are sometimes called</u> "elaboration" (along major axis) and "innovation" (away from major axis)"

Lines 88-89: "the modest species richness of the group allows high-dimensional macroevolutionary study to be computationally tractable" this is not entirely true. High dimensional morphometric studies for extremely diverse clades have been conducted in recent years, including studies sampling substantially more taxa and trait variables with the same methods employed in the present study—e.g., (Felice & Goswami 2018; Goswami et al. 2022; Watanabe et al. 2019)—certainly indicating that large-scale studies are computationally tractable.

We agree with the reviewer that a number of recent studies of high-dimensional morphospace have really pushed the envelope of what is computationally feasible. That said, we don't argue against it, just to say that the reasonable size of this clade means that analyses can be run and interpreted at reasonable timescales. We acknowledge that we are neither the first, nor the largest multivariate dataset.

Lines 90-91: " a plesiomorphic grade with highly derived morphotypes nested deeply within the phylogeny". I do not understand what the authors mean by this sentence. Plesiomorphy is a characteristic of characters or traits, not taxa or clades. If they mean that the group in question has

many plesiomorphic traits, then they are plesiomorphic relative to what? And how does this make it more interesting to study morphological evolution? One can easily assume that most clades have a combination of both plesiomorphic and derived traits relative to their sister groups.

This phrase was noted by another Reviewer, and so we apologize for the confusion. We intended to draw attention to the nested nature of highly derived phenotypes within a generally conserved group. To correct this, the sentence now reads "Additionally, tiliquines are ideally suited for studies of the tempo and pace of morphological evolution because they have been suggested to show a morphological gradient with highly derived phenotypes nested deeply within a generally conserved body plan." We believe this better explains why it's interesting to study the morphological evolution of this group.

Methods section:

Lines 408-410: "… Variable Rates model in which rates of change vary across individual branches of the tree—similar to the relaxed clock model of timetree estimation." As it currently reads, it gives the impression that the variable-rate model is a qualitatively different model of evolution relative to the other models tested (BM, OU, etc). However, BayesTraits (and most clock models, following the analogy provided by the authors) actually assume a Brownian motion to model rate variation among all branches—see also a review of clock models in the introduction of Lartillot et al. (2016). VR and clock models do allow for much greater flexibility of rate inference across branches, but it does not represent a qualitatively different model to BM. So perhaps the authors should make this explicit, and perhaps rename VR to BM-VR.

This is a very good point, and one we certainly agree with. The Variable Rates model applied here only differs from a 'traditional' single-regime BM in allowing each branch to have its own independent rate. We appreciate the recommendation to rename the model to BM-VR but to maintain consistency and with the nomenclature of BayesTraits and avoid confusion we have opted to retain the name Variable Rates (VR). We have rephrased the sentence in question to now read:

> "and a Variable Rates model in which rates of change vary across individual branches of the tree---in principle a multirate BM model similar to the relaxed clock model of timetree estimation"

Lines 368-371: "Rough alignments were compiled using MAFFT50 and refined using MACSE51, alignment parameters are specified in the supplementary materials" this is very good practice! Could you also report here the average length of each AHE locus? This is supposed to be one of the key advantages of using AHEs rather than UCEs (longer reads per locus). It would also be beneficial to explain to non-specialist readers the advantages of using AHEs rather than UCEs, especially as the latter have much broader usage in phylogenomics.

Thank you for the kind comment. While we agree that there are advantages to using AHEs over UCEs (length, locus informativeness, model-ability), we're unfortunately tight on word limits. We have included some general summary statistics that might be of interest in the supplement *Investigating Data Completeness and Informativeness*. For quick reference average AHE locus length is ~1884 bp (min=332, max=2708). For comparison, in another squamate dataset of similar phylogenetic depth that includes both AHE and UCE loci (SqCL kit) our average alignment length is ~2kbp for AHE and ~580bp for UCE targets.

Lines 420-436: The authors first inferred ancestral states along the tree assuming Brownian motion in phytools and a constant evolutionary rate. They did this from root to tip in 0.1myr window intervals. Then used reconstructed ancestral states to calculate disparity through time across each time window. This is all fine and standard from what is currently available for inferring ancestral states and calculating disparity (although this section would benefit from making it explicit whether you used your own custom scripts are one of the available R packages for calculating disparity through time. If the latter, please cite it).

My main question comes to their inference of evolution rates, however. Since the authors assumed constant evolutionary rates to simulate character evolution in the tree to begin with, what is the point of inferring rates of change after this first step? Instead, why didn't the authors use the rate value estimates produced by BayesTraits? On the other hand, the results in the main text talk about rate scalar (r) values > 2 for inferring significantly accelerating rates, which I assume is inferred with BayesTraits, which gets a bit confusing. Please clarify.

Thank you for these detailed comments and we regret causing confusion. In this portion of the methods we describe that we estimated our ancestral states by providing our "VR rate-transformed trees" which are the trees with branch lengths scaled by BayesTraits estimated rate scalar. Because the branch lengths are now relative to the amount of rate variation * time, we can estimate ancestral states using a constant process. For example, take two branches with length 0.5 million years. Branch A has an estimated evolutionary rate of 0.2, Branch B 0.4, and the mean evolutionary rate across the tree is also 0.2. The rate scalar for Branch A is 1, the rate scalar for Branch B is 2. To accommodate this we multiply the branch lengths by the rate scalar, so Branch A remains 0.5 million years long and Branch B is now 1 million years long. A constant Brownian process will allow twice as much variation to occur along Branch B because *variance* = rate^2 x *t*. This is in principle the same process as suggested by the Reviewer, however we're not familiar with an implementation that would allow a user to directly apply the evolutionary rates to estimate ancestral trait values (but would be open to trying a different approach if one is available). Our design is highly efficient as it takes advantage of existing comparative methods in R. We implement this in the function *process_PPP* in the *Scripts/plotting_BayesTraits.R*. Using this design (scaling the tree) also allows us to estimate the likelihood of the Variable Rates model for comparing model fit via AIC.

Lines 688-696: it would be highly beneficial to have a supplementary table listing the five models of modular evolution, the anatomical regions represented by each module for each model, and what are the specific traits that would go into each candidate module. It would become much easier to understand the different model hypotheses being tested.

This is a great idea and we appreciate the recommendation. Previously we had included the suggested table (the input for *EMMLi*) as a supplementary data file (Data/Morph_Module_Models.csv). We now include that as a table directly in the supplement and we include a new supplementary figure which outlines these modules across the skink body (Fig.S10).

Discussion:
Lines 172, 218, 234: the way these sentences currently read, it sounds like the authors are addressing this question for all lizards (i.e., a broad level squamate analysis), which is far from what is attempted here. Please be specific to the group in question (Tiliquini), or it may sound like they are overreaching.

We appreciate this comment and have adjusted the language to avoid making overly broad claims based on the data and results we present.
Line 172 now reads “we investigate the morphological evolution of the tiliquine skink body”
Line 218 now reads “Our analysis of the tiliquine lizard body plan”
Line 234, in our opinion is appropriate as is, referring to the focal group of skinks as “these lizards”

MINOR COMMENTS:

Title: the title reads a bit too generic, not telling readers what is the study system investigated. The title currently reads more like a review on the topic. I suggest something like "Evolutionary bursts drive morphological novelty in Tiliquini skinks".

We agree with this assessment and appreciate the suggestion provided. The new title reads “Evolutionary bursts drive morphological novelty in the world’s largest skinks”.

Line183: what do you mean by emerging fossil system? I have never heard of this term before.

This was an unfortunately vague term, which we intended to mean that the group has been a recent focus of extensive paleontological work. We have replaced it in the text with:

> “Despite their importance as models of reptilian sociality and recent fossil discoveries, phylogenetic hypotheses for the Tiliquini have relied on a handful of molecular markers and limited species sampling.”

Line 229: Malmgren et al. (1983) ref 43?

Yes, thank you for noting this. We have corrected it in the text.

Lines 437-438: Lastly, to summarize the major avenues of morphological change we ran PCA on (1) all traits jointly and across… Across what? Modules or clades?

We have changed this to read “we ran PCA on (1) all traits jointly across all species, and across …”

Tables for morphological measurements do not have table numbers associated.

Thank you, we have now listed these as Tables S3 and S4.

Fig. S5, trait correlations: please specify the correlation index used for measuring correlation (I assume standard Pearson correlation, but you must be explicit).

We have corrected this to now read: “Plot generated using *corrplot* (Wei & Simko 2021) using the Pearson (parametric) correlation.”

Line 680: correct "EMMLi [CITE Goswami & Finarelli]."

Thank you, we have added the appropriate citation.

References cited above:

Erwin, D.H. 2015. Novelty and innovation in the history of life. Current Biology, 25(19): R930-R940.
Felice, R.N. & Goswami, A. 2018. Developmental origins of mosaic evolution in the avian cranium. Proceedings of the National Academy of Sciences, 115(3): 555-560.
Goswami, A., Noirault, E., Coombs, E.J., Clavel, J., Fabre, A.-C., Halliday, T.J.D., Churchill, M., Curtis, A., Watanabe, A., Simmons, N.B., Beatty, B.L., Geisler, J.H., Fox, D.L. & Felice, R.N. 2022. Attenuated evolution of mammals through the Cenozoic. Science, 378(6618): 377-383.
Halliday, T.J.D., dos Reis, M., Tamuri, A.U., Ferguson-Gow, H., Yang, Z. & Goswami, A. 2019. Rapid morphological evolution in placental mammals post-dates the origin of the crown group. Proceedings of the Royal Society B: Biological Sciences, 286(1898): 20182418.
Harmon, L. 2018. Phylogenetic comparative methods: learning from trees. Open Access Publication.
Harmon, L.J., Losos, J.B., Jonathan Davies, T., Gillespie, R.G., Gittleman, J.L., Bryan Jennings, W., Kozak, K.H., McPeek, M.A., Moreno-Roark, F., Near, T.J., Purvis, A., Ricklefs, R.E., Schluter, D., Schulte Ii, J.A., Seehausen, O., Sidlauskas, B.L., Torres-Carvajal, O., Weir, J.T. & Mooers, A.Ø. 2010. Early bursts of body size and shape evolution are rare in comparative data. Evolution, 64(8): 2385-2396.
Lartillot, N., Phillips, M.J. & Ronquist, F. 2016. A mixed relaxed clock model. Philosophical Transactions of the Royal Society B: Biological Sciences, 371(1699): 20150132.
Simões, T.R., Vernygora, O.V., Caldwell, M.W. & Pierce, S.E. 2020. Megaevolutionary dynamics and the timing of evolutionary innovation in reptiles. Nature Communications, 11: 3322.
Slater, G.J., Harmon, L.J. & Alfaro, M.E. 2012. Integrating fossils with molecular phylogenies improves inference of trait evolution. Evolution: International Journal of Organic Evolution, 66(12): 3931-3944.
Slater, G.J. & Pennell, M.W. 2013. Robust Regression and Posterior Predictive Simulation Increase Power to Detect Early Bursts of Trait Evolution. Systematic Biology.
Watanabe, A., Fabre, A.-C., Felice, R.N., Maisano, J.A., Müller, J., Herrel, A. & Goswami, A. 2019. Ecomorphological diversification in squamates from conserved pattern of cranial integration. Proceedings of the National Academy of Sciences, 116(29): 14688-14697.