It is a massive undertaking to try and review a paper of this magnitude, so I will do my best to structure my review in the most helpful way possible. To wit, I have three *a priori* concerns with a handful of basic tenets of the manuscript and the work that went into it:

1)As always, I have concerns with publications based wholly or in part on privately-held specimens, such as SMA 0011. I'm aware of the Sauriermuseum's current policies, but it remains a private museum and a private specimen and my concern for setting a precedent is high.

2)I disagree wholeheartedly with the notion that one can split genera based on the results of a phylogenetic analysis that maintains monophyly of the original genus. (see the case of *Brontosaurus* vs. *Apatosaurus* for a primary case example). Unless a phylogenetic analysis selects a cladogram (or cladograms) with a poly or paraphyletic assemblage of what was a previously monophyletic group, that cladogram contains zero evidence whatsoever for carving a group up. From what I can determine here, none of the groups previously held to be monophyletic (by modern taxonomy) are recovered otherwise, which means that any splitting is purely arbitrary and thus contrary to the very point of both numerical phylogenetics and modern nomenclatural acts. This is a major point of contention for me. Furthermore, I don't think using "distance" as a measurement in character-based phylogenetic analysis is a valid tool for diagnosing subclades (i.e., species). I discuss this below, but characters are not all equal, lots of characters go unnoticed (or dismissed as unimportant), and the methodology by which we assign states to those characters is, at best, arbitrary.

3)Finally, I don't know that a phylogenetic analysis including juvenile specimens alongside adult specimens is going to give you a particularly trustworthy result. This is particularly important for sauropods (and this group of sauropods) for two reasons. First, what little we understand about juvenile diplodocoids suggests that they are very different from the adults morphologically, particularly in the skull, which is the source of the majority of informative characters in the group. Second, the juvenile material we have is *extraordinarily fragmentary*, even by sauropod standards. A specimen exhibiting either one of these conditions is a potential bugbear; when they coexist as they do here I would recommend extreme caution at best. I'm aware of the weighting methodology you employed here, but I'm not sure it is sufficient and it may introduce more questionable results than it clears up (see *Australodocus*).

Of those three concerns, I would rank them in this order (from most troubling to least): 2, 3, 1.

With those three points listed, allow me to continue by enumerating my other general comments:

Introduction:

This first paragraph of course needs to be revised now in light of Pablo Gallina's newest paper.

Description:

The generic diagnosis does not entirely do it for me. The first autapomorphy is extremely arbitrary (and therefore essentially an unreproducible observation), and in any case I disagree completely based on the photographs presented. The third autapomorphy is potentially the result of the general lack of atlantal neurapophyses preserved in the record.

Similarly, I have issues with a few of the diagnoses of the species. “Robust humerus” again means very little, particularly given that taxa like *Apatosaurus* which sit just outside Diplodocinae have also been noted for their “robust” humeri. The absence of the tubercle on the humerus is potentially a size-related character. The proportion character of the radius is interesting but I’d like to see some reasoning behind why that measurement and cutoff point was chosen, and what the distribution is in the group.

Line 298 - “see next chapter” - Is this a leftover from your dissertation?

The description of the material was thorough and well-described, as is typical for both authors. As an aside, as someone who has written (and is still writing) an alarming number of descriptions in this group, I feel like I can’t be alone in saying sometimes I wish I could just write “the lacrimal looks like every single other lacrimal” sometimes and leave it at that.

I will say that the supraoccipital crest looks pathologic or (less likely) taphonomic to me (without, of course, having put the bc in my hands yet). The bone looks exceptionally porous here and just generally not healthy, and I wouldn’t put too much stock in the lack of a sharp ridge there, especially given where the material falls out phylogenetically.

The discussion of ontogenetic stage is also slightly problematic for me, in that I think a couple of logical leaps have been made that might be worth revisiting. The first of these is the whole “ontogeny recapitulates ontogeny” argument. I’m reasonably certain the phrasing is tongue-in-cheek, but I don’t think that you can make the argument that juvenile characteristics are plesiomorphic in the group based on two closely related juvenile skulls, one of which does not have an adult counterpart. Also, given that smaller adult sauropods, both within diplodocidae (i.e., *Tornieria*) and in other diplodocoid groups (i.e. *Suuwassea*) also typically have rounder snouts than *Diplodocus* and *Apatosaurus*, I think this simply not a justifiable argument to make. Secondly, I would urge you STRONGLY to take caution in applying the ontogeny of the distantly related *Europasaurus* to that of this taxon. The ontogenetic stages from the Carballido paper were based on that material, which had numerous representatives of multiple age classes to draw on, and I don’t know that you can apply the lessons of when one taxon gained all the characteristics of adulthood to another just yet.

Material included

A large list, but I find myself wondering about some of the exclusions - CM 11255? USNM 2673?

Character list

I know this isn’t final formatting, but it would be nice to have some way to break this up between characters - perhaps bolding C1, C2, C3 … in addition to/instead of bolding the word “comments”, or adding a paragraph break, or an indent, or something. As it is, it’s hard to read straight through, and harder to scan for a particular character.

On occasion there are semantic issues with the coding, such as this example for C4: "*Because the plesiomorphic state is state one, the character was left unordered*" and this example in C114: "*Since the reduction of the number of dentary teeth was accomplished and reversed several times, it is not entirely clear how evolution worked in this case. The character was therefore left unordered.*" In both examples, it is more correct to say that because the derived state is ambiguous, the character is most parsimoniously left unordered - it might seem like a small detail, but it's deceptively important. Particularly for the second example, it is important to note that your interpretation of ordering is *a posteriori* based on a chosen topology and that interpretation could vary strongly depending on character/OTU selection.

C60: this character is problematic for me (see above), but I think it could be helped if it the method of measurement was described a little bit more - as written, I can't quite parse what you mean here.

C69: this character seems like it could be problematic re: taphonomy

Characters with ontogenetic variation (e.g., C111, C113): As above, I think these sorts of characters become highly problematic when they characterize variation in your OTUs that is both inter- and intra-specific/generic. Especially when your criterion for distinguishing valid taxa is so relatively small (5 steps), I think these sorts of characters (or including the juvenile OTUs in question) are candidates for removal.

C135: What sort of serial variation do you see in this character? It seems like it would be tied with elongation.

C142: How did you code taxa without bifurcated neural spines?

C163: The derived state seems like the conflation of two different states to me.

C191: break, not brake

C192: It would be preferable if you used aEI (scale centrum length to the average of centrum height and width) rather than EI to avoid confusing changes in centrum elongation with those in cross-sectional shape as well as to account for deformation, as in Chure et al., (2010) and others.

Methods

I really, really like the overlap method and would like to see it expanded upon a little bit below. Anything to start breaking data out of that matrix and into the eye of the reader is a plus in my book, and this seems like it could be really useful. I'd actually like to see it in the actual paper instead of the supplemental info.

I'm not sure i'm buying into the implied weighting scheme. I get what Goloboff was getting at with it, and I think in certain instances it can be a useful tool. However, I think in this particular sample, with as much convergence and homoplasy as exists (and with as many subadult specimens as are added) the result is that you're downweighting what are some locally very important characters and getting some funky results (*Australodocus*, yes, but also higher in the tree as well). I think it was a clever idea that ultimately didn't prove useful for the sample. As a result, I wouldn't place a ton of weight on it and probably would not include it in the paper itself, although I'd probably provide the results in an appendix so people can chase it down if they want to.

What would seem to be a more useful exercise is to look at the individual CIs of synapomorphies uniting interesting clades found in your chosen cladograms. This allows a user to identify those homoplastic characters (which is more immediately transparent than the implied weighting) and evaluate the relationships they imply individually. Character CI is easy to pull out of Mesquite, so it would be a relatively easy exercise to do, and I honestly believe would be more informative. If you find that you have a large number of autapomorphies, you might also consider looking at the Retention Index or the RC (or all three). In any case, you're going to have to make a judgement call about what is an unacceptably low value for your chosen index, but my suspicion is that the outliers will be readily apparent. Additionally, you might revise your outgroup to include less distant/convergent taxa, thus lowering apparent homoplasy.

Similarly, you get some funky results in the equally weighted scheme as well (thinking primarily of *Cetiosauriscus* and *Haplocanthosaurus*, but also the paraphyletic Macronaria). Different never means wrong, but I'd like to see a little more of a breakout discussion of what's causing those variations (OTU selection? particular characters?) and what happens when you play with those variables (taxon/character pruning, etc.). It's possible you might find some decent support for a non-intuitive relationship, and that in turn could set off a whole new project.

Discussion

You might reword the ending of the first paragraph here - on first reading it appears that you're saying that almost all studies were based on Wilson (2002) (lines 4464-4465), which is only true pre-Whitlock (2011) as you note later. Following up on that, you note that *Suuwassea* was the biggest mover, without mentioning that it ended up in Dicraeosauridae in most (all?) analyses post-W11. Also, *Tornieria* was included in at least W11 and Mannion et al. '12*, contra* lines 4470-4472 - its position was one of the large differences between the two analyses.

As I hinted at a little bit in the beginning, I have problems with some uses of phylogenetics as a tool for taxonomy. One use I don't put much stock in is using "distance" in a character-based phylogenetic analysis to say two OTUs are the same species (e.g., YPM 429 + AMNH 6341, Seismosaurus + MDT, etc.). Yes, only one character separates YPM 429 and AMNH 6341 in your analysis. Is that the only difference between them? Is it worth more or less than two smaller differences you might have missed/thought insignificant? And then, there are four (!) state changes between *Seismosaurus* and two *Diplodocus* specimens, but that's still not enough to make them different species. Where do we draw the line? I'm not casting aspersions on your anatomical expertise here at all - these are issues we all wrestle with and struggle with when we do these sorts of analyses.

I know that you've attempted to apply a criterion to determining "importance" of a number of character state changes, and I applaud the thought, but it seems to me to be a completely circular argument - you're basing your step count criterion on a relationship you've recovered from the tree you're trying to use to determine species and genera. Furthermore to that point, it still does not account for the issues raised above - in point of fact, you're not even using the same five/ten character states for all of your assessments, so there's no possible way they can be equivalent. Ultimately I just don't think anybody can say "this number of state changes is a species-level divergence and this one is a generic-level divergence".

I also don't think wiggling taxa around between species is a particularly useful exercise. Do we need to have a modifier to tell us we're talking about only one specimen, or two specimens? Those already exist - the specimen number(s). Do we need to talk about a largish group of specimens? Well, that already exists too - the genus. The rest seems like rearranging the furniture to me, and I think it just makes life confusing unnecessarily. It might seem like a personal preference type issue, but I think it really impacts communication to swap specimens back and forth into and out of species all the time - imagine a scenario where two taxa are four changes apart in your analysis but six in another! Who wins?

I very much appreciate the list of synapomorphies at each recovered node - nobody ever does this and it is such a useful tool to have.

CM3452 + K +B: Wouldn't a more parsimonious interpretation, given the lack of a skull in *Barosaurus* and the ontogenetic stage of CM3452, CM11255, and *Kaate* be that a spur is a juvenile characteristic?

AMNH 969 + *Galeamopus*: As you note, this is potentially fragile based on the character states uniting the group. You've got to dance with the girl you brought, but I appreciate the caution here.

Curious to see *Australodocus* nested in Diplodocidae in the implied weight trees - I would have thought spongy internal bone texture would have been unambiguous enough (and therefore highly weighted enough) to push it back towards where it (probably) belongs out in TSF-land.

I appreciate the huge amount of information here, from clades to holotypes to non holotypes and beyond, but it's a bit daunting. I know it's an online pub, but you might consider editing some of this stuff out a bit (Gigantosaurus africanus, for example) and into a supplemental info file just as a sop for the reader who's made it 170+ pages in already.

Line 5668 - Why is it still *Diplodocus hayi* here and not *Galeamopus*?

Assignation of CM 3452 to *Barosaurus*: this is problematic for me. It falls in there, as a subadult, based on one character that is also known from a well-established specimen of *Diplodocus*. This, to me, is not evidence that CM 3452 belongs anywhere except Diplodocinae indet. at the moment. I don't think this one is going to hold up over the long run.

Biogeography: as above, the new Cretaceous Argentine specimen is going to cause some re-writes here.

Conclusions:

I've discussed my concerns with your methodology for splitting genera/species above, but it should appear above in your methods section.

Line 7689 - Willi Hennig, not Willy.

In sum, I think there is a lot of good work here. The description of the material and the phylogenetic analysis (minus what I consider a failed experiment in weighting to eliminate onotogenetic effects) are both well accomplished, and I think that the result of the specimen-level analysis is interesting if not ground-shaking. I’m less convinced of the validity of a few of the autapomorphies of the new taxon, and I have some fundamental issues with some of the choices made with regard to splitting/grouping taxa based on character state distances, but if those are adequately resolved I don’t see any obstacle to publishing. I might strongly encourage some heavy-handed editing (or at least shifting to supplemental info), particularly with the dozens of pages of taxonomic info-dumps, but I would like to point out that it is all actually useful information to have and will be of particular benefit to new researchers.