# Reviewer 2

The paper of Stiller et al. is focused on the elucidation of phylogenetic relationships of the 'spaghetti worms' (Terebelliformia) based on the analyses of the transcriptomic data supplemented by a much larger dataset of selected terebellid genes and morphological characters.

I consider the paper of a very good scientific quality with concise hypotheses and conclusions based on properly performed comprehensive phylogenetic analyses.

We that the reviewer for their time and effort in reviewing our paper and the constructive comments. We have implemented all suggestions in the manuscript and include in a detailed point-by-point response to all comments below (reviewers' comments in black, our replies in blue).

I only have minor comments/suggestions:

Introduction, Line 91: "Terebelliforms may have been ancestral "naked" with poorly developed chaetae and were free-living". The life history has not been discussed anywhere in the introduction as well as you do not investigate this later. Even though I believe the evolutionary history of the organism's lifestyles is very interesting, I suggest omitting the second part of the sentence unless you don't feel it is important for your paper.

Removed reference to lifestyle as suggested.

M&M, Line 134: How were the potentially present contaminant sequences, i.e. of environmental origin (e.g. from ingested food) or by lateral gene transfer) identified and removed? If not removed from your dataset, can these data influence the results of your phylogenetic analyses? Include the section on the identification/removal of such sequences in this M&M.

Thank you for this suggestion. We have now included two approaches to remove potentially contaminated sequences from the phylogenetic dataset. 1) We investigated the possibility of cross-contamination during lab work or sequencing. Spillover from one sample to the other can be expected to yield identical sequences. Inspection of the removed branches showed that the majority of the identical sequences were between Alvinellidae, which likely indicates evolutionary similarity rather than contamination. We nonetheless excluded the 824 terminals from their respective gene trees. 2) We use TreeShrink, which has been designed to identify suspiciously long branches of a species in certain gene trees, while considering the normal range of branch lengths for this species. This approach is expected to identify sequences that are more

dissimilar than any of the Terebelliformia+*Arenicola* sequences manifesting itself in an unusually long branch, which could be expected for a non-terebelliform contaminant. In total, these two approaches removed 5.9% of all branches present in the dataset, which indicates that the level of contamination, if present, was likely low. Summary of the cleaned gene trees with ASTRAL yielded the same topology as shown in Figure 1. We therefore conclude that if the potentially problematic sequences were of external origin, they did not significantly impact the phylogenetic analyses.

These analyses are now included as follows:

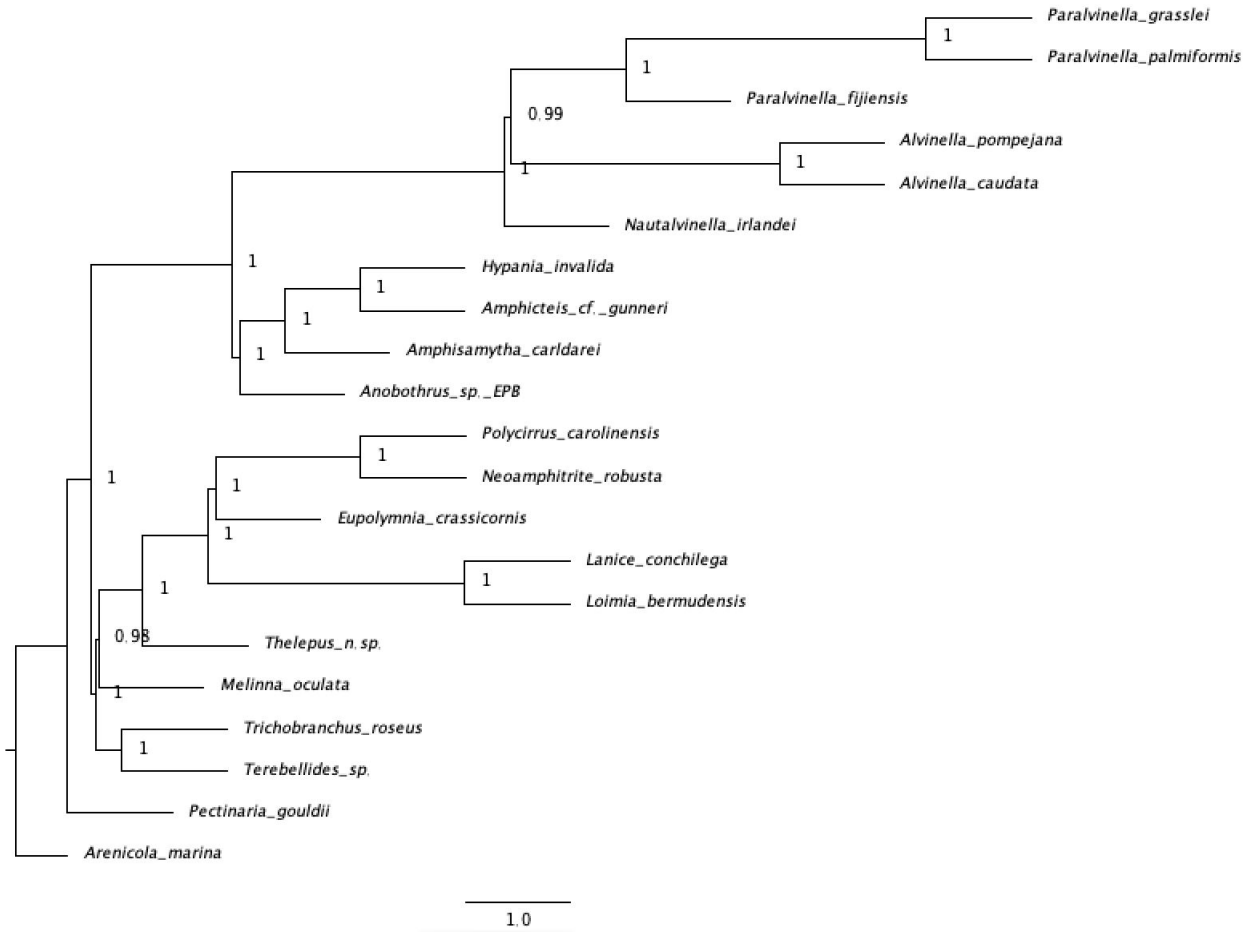
Material & methods

"Another experiment targeted sequences that were potentially introduced from extraneous sources such as parasites, food or symbionts of the studied species or from cross-contamination during lab work or sequencing. We therefore excluded potentially problematic branches from gene trees and investigated whether the exclusion impacted the phylogenetic hypothesis presented here. The first approach removed identical sequences, which may have originated through spillover between samples. We identified 456 loci that had two or more identical sequences (824 total) and we removed the corresponding branches from the gene tree using pxrmt [45]. The second cleaning step aimed at unusually long branches, which would be expected if some sequences originated from a non-terebelliform parasite or commensal that is evolutionarily more distant than the ingroup. We used TreeShrink v. 1.3.3 [46] to remove terminals that had uncharacteristically long branches considering the species-specific distribution of branch lengths across all gene trees (6272 branches total). Altogether, 5.9% of all branches were removed, indicating that if contamination existed in the samples, it was at a low level. The 'cleaned' gene trees were summarized with ASTRAL. "

Results

"Exclusion of potentially extraneous sequences also resulted in the same topology with improved support on one node (Supplementary figure S11)."

Supplementary file 1 includes commands to run the subsetting and the figure below showing the ASTRAL results.




**Supplementary Figure S7**: Coalescent-based species tree from ASTRAL based on gene trees after cleaning for i) identical sequences and ii) for unusually long branches. Node labels are local posterior probabilities. The topology is identical to the one in main text Figure 1.

Relevant files (table of identical sequences, TreeShrink output, cleaned tree files, resulting ASTRAL tree) have been added to the data repository.

M&M, Line 209: The sentence seems unfinished.

Sentence fixed to read "were amplified using a PCR mixture of ... or, when amplification failed, using PuReTaq Ready-To-Go PCR beads"

Results, Line 269: Bs=64 is not high and later you refer to bs=66 as "it lost support" (lines 288-289). Be consistent.

The first referenced statement was inconsistent and we adjusted it to read:

"On those small matrices, support for a clade of *Melinna*, Trichobranchidae and Terebellidae was moderate to high (bs=64-92), but support for Terebellidae plus Trichobranchidae was low to moderate (bs=32-56, Supplementary figure 3-4)."

The second statement of "lost support" was relative to the bootstrap support of another tree. We clarified the statement by removing "lost support".
"Relationships were maximally supported in the coalescent-based species tree but the node supporting Ampharetinae as the sister to Amaginae and Amphicteinae was lower in the 80% concatenated matrix (bs=95) and in the 90% concatenated matrix (bs=66)."

Results, multiple lines: I am confused by the use of Melinnidae/Melinninae throughout the text of the MS. You said that "Ampharetidae is traditionally categorized as Ampharetinae and Melinninae" and that Ampharetidae was paraphyletic (Ampharetinae and Melinninae did not cluster together). Because Ampharetidae sensu lato was paraphyletic, it is restricted here to the former Ampharetinae, which hence becomes Ampharetidae." Does the same situation apply for Melinninae/Melinnidae? At line 264, you use the term Melinninae but at line 271 and below as well as in Fig. 1. you use Melinnidae. This is also confusing in the abstract. Clarify, please.

Thank you for pointing out that a clear statement of the name change for Melinninae to Melinnidae was missing from the results. The reviewer was correct that a nomenclatural change was also necessary in this case. We made multiple changes to the manuscript:
- In the main text, we use Melinninae (the traditional name) until the main phylogenetic hypothesis is presented in the Results section that demonstrates the separation of Melinninae and Ampharetinae and the required nomenclatural changes. The paragraph describing the placement of *Melinna* now ends with the following sentence to clearly signal the change of terminology from now on:
"Because Melinna did not group with Ampharetinae, Melinninae now becomes Melinnidae."
- Abstract: "... and and Melinninae. The latter, traditionally a subfamily of Ampharetidae, was unexpectedly the sister to Terebellidae, and hence becomes Melinnidae, and Ampharetinae becomes Ampharetidae."
- Figure 5: The manuscript now includes a figure that summarizes the names used by Holthe 1986, Nogueira et al. 2013, and the current study to give an immediate overview of nomenclatural changes and how they relate between the three studies.

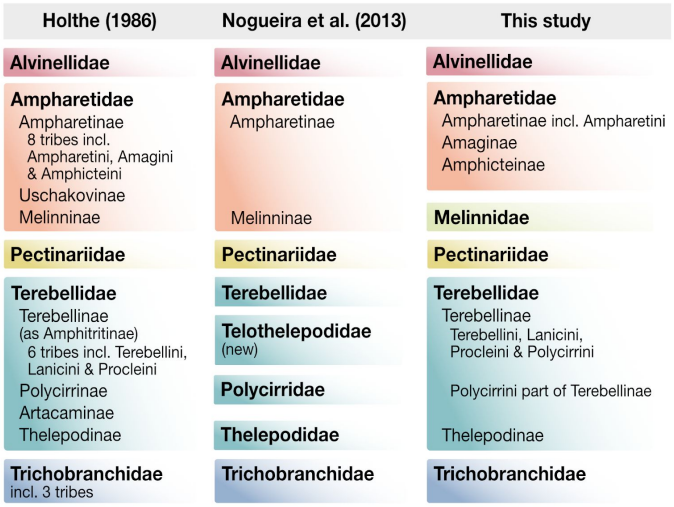



**Figure 5**. Summary of nomenclatural changes of this study in relation to two previous works.

Results, Line 262: Include also the reference for the results of the phylogenetic analysis.

Thanks. We added a reference: "The monophyly of Trichobranchidae, Alvinellidae, and Terebellidae was confirmed (Figure 1).

Figure 1: The picture is beautiful but I suggest including the legend for the names of the pictured organisms in the same way as done in Fig.2.

Terminals with accompanying photographs have been highlighted in bold.

Figure 2: Maximum nodal support is probably shown by the asterisk in the tree but this should be explained in the figure legend.

Corrected as suggested.

Figure 2: "tcp denotes terminals which are shared with the transcriptome phylogeny." It would be better visible if you label these taxa in bold in the tree.

Changed as suggested.