

Thank you for the detailed comments and reviews on our manuscript. We address all of the points raised by the Referees below.

Referee 1 (A. Weigand)

Specific comments:

1) *In the discussion section, the authors state that "[t]he findings suggest that the non-pyrgulinid Hydrobiidae form a monophyletic, speciose and endemic clade and thus, by definition, represent a species flock". However, and since Bayesian posterior probability values have a strong tendency towards high values (see e.g. Cummings et al. 2003 - Comparing bootstrap and posterior probability values in the four-taxon case), a support value of 0.87 cannot be seen as a decisive support nor as an argument for the monophyly of this organism group. Thus, hypothesis 1 must be questioned. Furthermore, and because this hypothesis is the authors self-defined requirement to perform all subsequent diversification analyses, the monophyly of this group has to be unambiguously demonstrated first. I would suggest to additionally implement Maximum Likelihood analyses complementing the BI data and perhaps an even more dedicated partitioning scheme with different partitions based on the three CO1 codon positions and different partitions for the stem and loop regions of the 16S rDNA fragment. A nuclear marker would be great, but I see that this will be hard to achieve for all the specimens.*

Response: We are thankful for this comment and would like to clarify this point. The support value of 0.87 is comparatively low for a Bayesian posterior probability and we should be cautious in drawing the final conclusion that the flock is monophyletic.

As suggested by the referee, we performed additional maximum likelihood (RAxML) analyses using a default partitioned dataset (16S, COI; ML bootstrap: 30) and a partitioned dataset, in which the COI fragment was additionally partitioned by codon position (ML bootstrap: 61). Though in both analyses, the Prespa clade is monophyletic (BPP = 1.0), there is no significant support for a monophyletic Ohrid clade.

However, we think that a potential non-monophyly of the Ohrid clade will not have significant implications for the analyses and interpretations in our manuscript. First, our LTT-analyses are strictly based on Ohrid taxa, and a single colonization event into Lake Prespa would add only a single lineage to our analysis. Second, and more importantly, the only random tree (#2), suggesting a non-monophyly of the Ohrid clade, indicates that the split between Prespa and Ohrid taxa occurred c. 2.38 Ma ago and thus before Lake Ohrid came into existence. In other words, it does not affect intra-lacustrine diversification rates.

To make this clearer to the reader, we plan to rephrase the respected section in our manuscript: "Our findings indicate that this group is comparatively old and most likely monophyletic. However, due to the comparatively low support of 0.87, a non-monophyly of the Ohrid group cannot be excluded. Nonetheless, even under the unlikely assumption that the Prespa group renders the Ohrid group paraphyletic (as in random tree #2), the molecular-clock analysis suggests that the split between Prespa and Ohrid is older than 2 Ma and thus does not affect intra-lacustrine diversification rates within the Lake Ohrid clade."

2) *Again in the discussion, the authors argue that "[o]ur LTT-plot (Fig. 4a) shows that several lineages have already existed when the lake reached deep-water conditions or even before it came into existence (see Trajanovski et al., 2010 for a potentially similar pattern in the Dina leach flock). Therefore, the term "cradle" may not only be used for monophyletic species evolving within the lake (intra-acustrine speciation) but also for a group of monophyletic species that started to evolve within the (palaeo-)basin." Important here and addressed are the two oldest nodes of the non-pyrgulid Hydrobiidae (> 2mya) and their four corresponding lineages. However, both assumed to be monophyletic groupings within the non-pyrgulinid Hydrobiidae (i.e. the resulting 4 lineages) have no posterior probability support at all (= below 0.5 according to the authors), must be collapsed and hence cannot be regarded as being monophyletic. The argument of several lineages existing prior to lake Ohrid formation thus does not hold true. Same is true for the monophyly of the mentioned species groups.*

Response: Thanks for raising these important points. In our manuscript, we wanted to point out that the term 'cradle' may not only be used for a monophyletic group of extant species whose ancestors started to evolve within the lake (intra-lacustrine speciation), but also for lineages/species that started

to evolve within the (palaeo-)basin. However, we agree that the above sentence starting with "Therefore, the term 'cradle' [...]" might be misleading and thus we will remove it from the revised manuscript.

Regarding the second last sentence of the referee, we disagree with this statement. Both, our time-dated phylogenetic tree (Fig. 3) and the LTT plot (Fig. 4) indicate that at least some lineages may have existed before 2.0 Ma and thus prior to the lake formation. However, collapsing lineages with low support will result in a polytomy with exactly the same age. This assumption is also supported by Referee 2, see point 2).

3) *Based on the inferred constant rate of diversification the authors conclude that "[the] initial working hypothesis – rate homogeneity – cannot be rejected. [...] If we assume that the rate of diversification in the non-pyrgulinid Hydrobiidae from Lake Ohrid is constant, linking environmental/climatic fluctuations to changes in tempo of diversification becomes impossible (see specific goal iv)." The diversification rate is defined as the net sum of speciation rate - extinction rate. Rate homogeneity thereby refers to the following (see Ricklefs 2007 - Estimating diversification rates from phylogenetic information): "The simplest diversification process presupposes that rates of speciation and extinction are the same for all lineages and do not vary over time. This is the assumption of rate homogeneity." Thus, rate homogeneity is mostly used as a model and cannot be inferred from a constant rate of diversification as it refers to constant rates of speciation and extinction. Ricklefs further states that "different combinations of speciation and extinction rates can produce the same expected clade size". This means, that different combinations in speciation and extinction rates may lead to similar inferred net diversification rates. As an example: potential environmental/climatic fluctuations (as proposed by some studies for Lake Ohrid) may have lead to extinction events and subsequent adaptive radiations within relative short periods of time. This is known for other taxa including hydrobiids from other regions of the world. The outcome would be a temporally increased rate of extinction and subsequent increased rate of speciation. However, and for the observed 0.1 my intervals, the "phylogenetic window" may be too broad to see those changes in extinction and speciation rates leading to a similar net diversification rate (i.e. birth and death of lineages) as expected under rate homogeneity. Thus, rate homogeneity as defined as constant speciation and extinction rates cannot be inferred nor supported by the data presented as only a constant rate of diversification is observed. Finally Ricklefs (2007) states that "it is unlikely that rate homogeneity can be unambiguously supported for any clade." However, the inference drawn by the authors that "Lake Ohrid never experienced catastrophic environmental events that resulted in the extinction of all or most of its endemic taxa and thus caused a "reset" of diversification processes" can be partially drawn as old lineages can be observed. However, lineages may have even survived those events. If this point is still dealt with in the discussion, it should be re-formulated according to what is supported by the data. I would be really careful in drawing the final conclusion that a high ecosystem reliance can be supported by the "rate homogeneity" of the investigated gastropod taxon. The opposite may be true: If a significant rate shift is observed and can be temporarily linked to a given environmental event, this may be seen as support for the influence of this event on the diversification process. Related to this discussion, see also point 4.*

Response: We thank you for pointing us to the difference between rate homogeneity as defined by Ricklefs (2007) and constant diversification rates, which is the parameter that we actually measured. We will replace the term "rate homogeneity" by "constant diversification rate" in the revised manuscript. The definition of rate homogeneity is a very strict theoretical concept and, as correctly identified by the referee, Ricklefs acknowledges that "*it is unlikely that rate homogeneity can be unambiguously supported for any clade*". The test of rate homogeneity is moreover challenging because there is still an ongoing debate whether extinction rates can be reliably estimated based on molecular phylogenies (Beaulieu and O'Meara, 2015; Morlon et al., 2011; Rabosky, 2010, 2014). Our used TreePar approach, for instance, is known to overestimate the extinction/speciation ratio but identifies shifts in net diversification rate correctly (Laurent et al., 2015; Stadler, 2011).

We are still confident that the window frame (grid) of 0.1 Ma is sufficient for the TreePar analysis, because 1) this is a reasonable timeframe as changes in diversification rates will probably not occur within smaller time slots in both genetic markers, and 2) we had enough power to detect shifts in two out of ten random trees at the time of 0.1 Ma. We furthermore think that the grid used is less important than the magnitude of changes in the diversification rate from one time slot to the next. Unfortunately, only two studies analysed the power of the TreePar approach under various scenarios of tree size and extinction fractions (Laurent et al., 2015; Stadler, 2011) but none of them tested explicitly the influence of grid size. We however re-analyzed random tree #1 using a grid of 0.01 which resulted in a

significantly better log-likelihood by including a single shift at 0.07 Ma (const: 45.882 vs. 1 shift: 35.959 → $p = 1.0$). Note that with a window of 0.1 Ma, a single rate shift in this tree almost reached the critical significance level ($p = 0.947$; see Table 2). This very recent single shift, again, represents a timeframe in which incomplete lineage sorting may play a crucial role.

As for the third point, that a high ecosystem resilience may not be supported by a lack of diversification rate shifts, we only partly agree. Operational criteria for resilience are lack of regime shift and lack of extinction events (for details see the discussion paper of Jovanovska et al. in the same special issue) and we agree that the latter is very difficult to apply to molecular phylogenies (see also above). In molecular phylogenetic studies, the impact of catastrophic events has been inferred from molecular-clock analyses and/or particular topological patterns (e.g., founder flush, recent re-colonizations; see e.g., Schultheiß et al. 2009). If such a pattern is not observed in a molecular phylogeny, more complex methods are needed such as diversification-rate analyses. We think that this methodological approach is powerful even for such recent and comparatively small datasets (we could detect rate shifts in two out of ten random trees at 0.1 Ma). Our analyses did not detect shifts in diversification rates in the pre-lake phase or early phase of the lake. We therefore argue that the lack in diversification rate shifts is related to the ecosystem resilience of Lake Ohrid against particular environmental perturbations such as the observed glacial/interglacial cycles.

*4) The authors refer to the species coverage, an important factor when calculating diversification rates based on phylogenetic tree hypotheses, by saying: "As for the sampling size (i.e., 17 out of 27 nominal species studied), the high diversity of evolutionary lineages found in our phylogenetic analyses indicates that our sampling design likely recovered most major evolutionary lineages within this cryptic group." It is correct, that there is a high probability that the authors cover all older/major lineages. However, at the same time, many more recent species may have been missed. This seems very plausible since a high amount of cryptic species is observed in Lake Ohrid, e.g. see *Pseudohoratia ohridana*. If more recent nodes accumulate, net diversification rate may show a shift in more recent times, e.g. during glacial cycles (Lindhorst et al. 2015). The oldest node referring to the taxon *P. ohridana* is even at 1 my of age. This taxonomic coverage is problematic and its influence on the diversification rate estimates has to be addressed more thoroughly in the discussion. However, distinguishing such a pattern from a "normal" pull-of-the-present effect due to speciation only processes will get challenging.*

Response: We will expand the discussion regarding taxonomic coverage: "As for the sampling size (i.e., 17 out of 27 nominal species studied), the high diversity of evolutionary lineages found in our phylogenetic analyses indicates that our sampling design likely recovered most major evolutionary lineages within this group. Our sampling includes the majority of genera described except for *Dolapia* (which some authors included in the genus *Ohrigocea*) and *Zaumia*. We also included a variety of morphotypes collected at various types of habitats and type localities across the lake and its surroundings (see Fig. 2). Moreover, the diversification-rate analysis does account for incomplete sampling and we did infer a single rate shift in two of the random trees tested in the present study. Thus the method used seems to have enough power to detect deviations from a constant diversification rate in our data set."

Discussing changes in diversification rates for recent diversification events, however, remains problematic given potential biases resulting from effects such as incomplete lineage sorting. In fact, for assessing such young events, coalescent methods might be more appropriate for phylogenetic inference but are beyond the scope of the present study. For instance, Morlon et al. (2010) developed coalescent-based methods to estimate gradual changes in diversification rate. However, no tests for distinct shifts in diversification rates are available in this framework but a pattern of sudden rate change is expected by catastrophic environmental events. We therefore refrained in our manuscript from including coalescent methods and extensively discussing young events of rate shifts. Moreover, we would be happy to include a cautionary note in our revised version pointing out some of the problems associated with such events.

5) The authors state that "we did not test for substitutional saturations as both genes have been suggested to be not saturated within the family Hydrobiidae (Wilke et al., 2001, 2013)". However substitutional saturation is dataset- and sampling-dependant and a general statement cannot be made here. Please test your dataset for substitutional saturation as this is easily done but at the same time may heavily bias the reconstruction of the phylogenetic tree hypothesis by lowering the phylogenetic information content of the data.

Response: We performed saturation tests for the COI gene fragment using DAMBE. Only little saturation is observed suggesting that this partition can be used for further (molecular-clock) analyses (see below). We did not perform a saturation test for the 16S dataset because 16S is known to be more conservative than COI (see also Wilke et al., 2001, 2013).

Therefore, we will rephrase this sentence in the revised manuscript: “We tested the COI dataset for nucleotide substitution saturation using the test by Xia and Xie (2001) as implemented in DAMBE 5.0.23 (Xia and Xie, 2001). The value for the proportion of invariant sites ($P_{inv} = 0.46$) was obtained from the jModelTest output for the preferred GTR+I+ Γ model. The observed saturation was significantly lower than the critical values ($p < 0.001$), suggesting that this partition can be used for further (molecular-clock) analyses. Note that we did not test for saturation in the 16S dataset as, within the family Hydrobiidae, the 16S gene is more conservative than the COI gene (Wilke et al., 2001, 2013).”

DAMBE output for the XIA saturation test for COI using the PINV value (0.46) obtained from jModelTest for the preferred substitution model GTR+I+G.

| NumOTU | Iss | Iss.cSym | T | DF | P | Iss.cAsym | T | DF | P |
|--------|-------|----------|--------|-----|--------|-----------|--------|-----|--------|
| 4 | 0.117 | 0.793 | 32.297 | 246 | 0.0000 | 0.759 | 30.650 | 246 | 0.0000 |
| 8 | 0.132 | 0.748 | 26.935 | 246 | 0.0000 | 0.636 | 22.023 | 246 | 0.0000 |
| 16 | 0.135 | 0.715 | 25.575 | 246 | 0.0000 | 0.505 | 16.320 | 246 | 0.0000 |
| 32 | 0.138 | 0.698 | 24.915 | 246 | 0.0000 | 0.371 | 10.361 | 246 | 0.0000 |

6) Briefly explain the methodological procedure of the diversification rate analyses performed in TreePar in more detail, as it is used as an additional argument besides the LTT-plots. The analytical difference should become more clear.

Response: We expanded the Methods section: “However, generating LTT plots and detecting changes in the slope is an explorative approach and might lead to misinterpretations (see Stadler, 2011). Therefore, potential shifts in diversification rates through time were analyzed using the R package TreePar v. 3.3 (Stadler, 2015) by testing a maximum of three shifts for ten trees randomly sampled from the posterior distribution. This package implements a birth-death shift model (Stadler, 2011) that allows changes in speciation and extinction rates along a phylogeny for a given timeframe and for pre-defined time intervals. Shifts were analyzed along the pruned tree with default settings and a time interval of 0.1 Ma. The sampling fraction was set to $17/27 = 0.63$ according to the actual number of species included divided by the number of nominal species described accounting for incomplete sampling. Results (log likelihoods of different runs, i.e., constant diversification rate vs. 1 shift, 1 shift vs. 2 shifts, and 2 shifts vs. 3 shifts allowed) were compared by applying likelihood ratio tests in order to examine whether shifts in rates explain the tree significantly better than a constant diversification rate (indicated by p values > 0.95 ; see Stadler, 2011, 2015).”

7) In the discussion, the authors argue that “the TreePar analysis used does account for incomplete sampling and we did infer two single rate shifts in the present study.” More precisely, a single rate shift each in two independent trees has been identified. Or do you have the assumption of two rate shifts during the evolution of non-pyrgulinid Hydrobiidae in Lake Ohrid? Would it be possible to show even more trees. I am not familiar with the standard procedure, but showing ten trees only seems rather too less.

Response: Thanks for highlighting this point. A single rate shift was found in two out of the ten random trees. We rephrased this sentence: “the TreePar analysis does account for incomplete sampling and we did infer a single rate shift in two of the random trees tested in the present study.”

Ideally, the entire posterior distribution with 20,000 trees could be tested for changes in diversification rates. However, testing multiple random trees as done here is more robust compared to analyzing a single consensus (maximum clade credibility tree) as often performed in different other publications for both diversification rate-analyses and LTT plots.

Referee 2 (T. von Rintelen)

Specific comments:

1) *A few issues should still be addressed, though, in order to further improve the MS. As the first reviewer, A. Weigand, has already pointed out, it would be interesting to see whether the apparent support for the monophyly of the non-pyrgulinid Hydrobiidae in Lake Ohrid by a BPP of 0.87 is also found when using a ML analysis, or, as I would also suggest, a BI analysis without constraints. In contrast to the authors' statement that they ran (initially) unconstrained analyses, I would suggest that any molecular clock analysis, whether relative or calibrated, is constrained in a way by the requirement of ultrametry. Given that BEAST enforces bifurcations, a, say, MrBayes analysis would show whether the topology and support for it are stable.*

Response: We hope that the issue of monophyly vs. non-monophyly of the Lake Ohrid species flock has been sufficiently addressed in the above sections of Referee 1, point 1). We did perform further non-clock analyses using a different and faster phylogenetic method (maximum likelihood; RAxML).

2) *Based on the present tree, I cannot quite agree with the first referee on his second point, as collapsing basal (unsupported) splits would not contradict the onset of diversification in non-pyrgulinid hydrobiids before deep-water conditions set in at Lake Ohrid.*

Response: We agree. See comments above, Referee 1, point 2).

3) *Regarding the inference of rate homogeneity and LTT plots (also discussed by A. Weigand), a brief perusal of Ricklefs 2007 suggests to me that this point certainly warrants some more discussion by the authors, acknowledging potential limitations and alternative hypotheses.*

Response: We will replace the term "rate homogeneity" by "constant diversification rate" and commented on this issue above. See comments above, Referee 1, point 3).

4) *I also concur with the other points raised by the first referee. Anyhow, I am convinced that the concerns raised above can be dealt with either by providing evidence for them being unfounded and/or by providing a more balanced discussion - either way, the data are worthy of publication in this journal.*

Response: We agree and will provide a more balanced discussion highlighting the above issues and limitations in the revised manuscript.

References

- Beaulieu, J. M. and O'Meara, B. C.: Extinction can be estimated from moderately sized molecular phylogenies, *Evolution*, 69(4), 1036–1043, doi:10.1111/evo.12614, 2015.
- Laurent, S., Robinson-Rechavi, M. and Salamin, N.: Detecting patterns of species diversification in the presence of both rate shifts and mass extinctions, *BMC Evol. Biol.*, 15(1), 157, doi:10.1186/s12862-015-0432-z, 2015.
- Morlon, H., Parsons, T. L. and Plotkin, J. B.: Reconciling molecular phylogenies with the fossil record, *Proc. Natl. Acad. Sci.*, 108(39), 16327–16332, doi:10.1073/pnas.1102543108, 2011.
- Morlon, H., Potts, M. D. and Plotkin, J. B.: Inferring the dynamics of diversification: a coalescent approach, *PLoS Biol*, 8(9), e1000493, doi:10.1371/journal.pbio.1000493, 2010.
- Rabosky, D. L.: Extinction rates should not be estimated from molecular phylogenies, *Evolution*, 64(6), 1816–1824, doi:10.1111/j.1558-5646.2009.00926.x, 2010.
- Rabosky, D. L.: Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees, *PLoS ONE*, 9(2), e89543, doi:10.1371/journal.pone.0089543, 2014.
- Salzburger, W., Mack, T., Verheyen, E. and Meyer, A.: Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes, *BMC Evol. Biol.*, 5, 17, doi:10.1186/1471-2148-5-17, 2005.
- Schultheiß, R., Van Bocxlaer, B., Wilke, T. and Albrecht, C.: Old fossils – young species: evolutionary history of an endemic gastropod assemblage in Lake Malawi, *Proc. R. Soc. London B*, 276(1668), 2837–2846, doi:10.1098/rspb.2009.0467, 2009.
- Stadler, T.: Mammalian phylogeny reveals recent diversification rate shifts, *Proc. Natl. Acad. Sci.*, 108(15), 6187–6192, doi:10.1073/pnas.1016876108, 2011.
- Stadler, T.: TreePar: Estimating birth and death rates based on phylogenies. R package version 3.3. <http://CRAN.R-project.org/package=TreePar>, 2015.
- Wilke, T., Albrecht, C., Anistratenko, V. V., Sahin, S. K. and Yildirim, Z.: Testing biogeographical hypotheses in space and time: faunal relationships of the putative ancient Lake Egirdir in Asia Minor, *J. Biogeogr.*, 34, 1807–1821, doi:10.1111/j.1365-2699.2007.01727.x, 2007.
- Xia, X. and Xie, Z.: DAMBE: Data analysis in molecular biology and evolution, *J. Hered.*, 92, 371–373, doi:10.1093/jhered/92.4.371, 2001.