ACCEPTED MANUSCRIPT, UNCORRECTED, NOT TO BE USED FOR CITATIONS, REFERENCING OR TAXONOMIC PURPOSES

Molecular Phylogenetics and Evolution

Nuclear phylogenomics, but not mitogenomics, resolves the most successful Late Miocene radiation of African mammals (Rodentia: Muridae: Arvicanthini)

--Manuscript Draft--

Manuscript Number:	MPE-D-20-00104R1
Article Type:	Research Paper
Keywords:	Late Miocene; radiation; Anchored phylogenomics; Rodentia; Tropical Africa; complete mitochondrial DNA
Corresponding Author:	Josef Bryja Institute of Vertebrate Biology of the Czech Academy of Sciences Studenec, CZECH REPUBLIC
First Author:	Ondřej Mikula
Order of Authors:	Ondřej Mikula
	Violaine Nicolas
	Radim Šumbera
	Adam Konečný
	Christiane Denys
	Erik Verheyen
	Anna Bryjová
	Alan R. Lemmon
	Emily Moriarty Lemmon
	Josef Bryja
Abstract:	The tribe Arvicanthini (Muridae: Murinae) is a highly diversified group of rodents (ca. 100 species) and with 18 African genera (plus one Asiatic) represents probably the most successful adaptive radiation of extant mammals in Africa. They colonized a broad spectrum of habitats (from rainforests to semi-deserts) in whole sub-Saharan Africa and their members often belong to most abundant parts of mammal communities. Despite intensive efforts, the phylogenetic relationships among major lineages (i.e. genera) remained obscured, which was likely caused by the intensive radiation of the group, dated to the Late Miocene. Here we used genomic scale data (377 nuclear loci; 581,030 bp) and produced the first fully resolved species tree containing all currently delimited genera of the tribe. Mitogenomes were also extracted, and while the results were largely congruent, there was less resolution at basal nodes of the mitochondrial phylogeny. Using a newly developed algorithm for subsampling of most informative loci, we also performed a fossil-based divergence dating. The results suggest that the African radiation started early after the colonization of Africa by a single arvicanthine ancestor from Asia during the Messinian stage (ca. 7 Ma), and was likely linked with a fragmentation of the pan-African Miocene forest. Some lineages remained in the rain forest, while many others successfully colonized broad spectrum of new open habitats (e.g. savannas, wetlands or montane moorlands) that appeared at the beginning of Pliocene. One lineage even evolved partially arboricolous life style in savanna woodlands, which allowed them to re-colonize equatorial forests. We also discuss delimitation of genera in Arvicanthini and propose corresponding taxonomic changes.
Suggested Reviewers:	Molly M. McDonough, PhD junior researcher mollymcdonough@gmail.com a specialist on African rodents, using phylogenomic approaches
	Ara Monadjem aramonadjem@gmail.com Leading specialist in the diversity and phylogeny of African rodents

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

	Peter Taylor peter.taylor.univen@gmail.com Leading specialist in phylogeny of African rodents
Opposed Reviewers:	
Response to Reviewers:	

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Dear editors,

Thank you very much for assessment of our manuscript MPE-D-20-00104 "*Nuclear phylogenomics, but not mitogenomics, resolves the most successful Late Miocene radiation of African mammals (Rodentia: Muridae: Arvicanthini)*". Following the comments of two reviewers we have performed its revision and we are now resubmitting the revised version to the Molecular Phylogenetics and Evolution. This version includes changes that are specified in details below (the comments of two reviewers are in `Courier font`, while our responses are in **Times New Roman** font).

Most comments of reviewer 2 were relatively minor and easy to correct. On the other hand, reviewer 1 provided numerous suggestions that would require substantial re-analysis of the data. The same reviewer provided an identical review during our submission of a previous version of this work to Systematic Biology (he/she even left "Systematic Biology" at several places in the text of the review; in fact, the review is a copy-paste of their review of the previous version). The manuscript was rejected by Systematic Biology based on this copy-paste review, despite two very positive additional reviews. Because we do not agree with several comments of ref. 1, we provide a detailed rebuttal letter below. While some suggestions for re-analysis were very useful and we re-analysed those data, other suggestions seem irrelevant and give the impression of conflict of interests (direct competition of the reviewer with our work).

We submit the version of the text with tracked changes as requested. We hope that performed corrections will be sufficient for re-evaluation of our manuscript and publication of our work in MPE journal.

Yours sincerely (on behalf of all co-authors)

Josef Bryja

Reviewer #1: This manuscript presents a phylogenomic analysis for one the most species-rich radiations of mammals centered in Africa, the murine rodent tribe Arvicanthini. The data include mitogenomes and 377 independent nuclear loci comprising nearly 600,000 bp from 40 species including all named genera in Arvicanthini. The data are compelling and the Arvicanthini is a compelling continental radiation ideal for demonstrating the utility and application of many locus, phylogenomic analyses. However, to make this study most valuable to Systematic Biology readers the phylogenomic analyses need much deeper interrogation. The only discussion of the phylogenomic analyses is a short comparison of mitogenomes and nuclear loci. Most of the discussion is devoted to a visual interpretation of biogeographic and ecological traits on the single best topology, most of which is not relevant to a broad audience. There also are no formal analysis of these traits, which should use posterior distributions of topologies. In addition, much of this section is speculation, including over equivocal alternatives (e.g. the biogeographic origin of Arvicanthini). The nomenclatural implications for genera are valid, but of little relevance to most Sys. Bio. readers. They also are not resolved and many have been noted previously without phylogenomic data. More needs to be done to demonstrate the information content and utility of the phylogenomic data and/or provide guidance in best practices for analysing these data. Alternatively, a formal analysis of signatures of adaptive radiation in Arvicanthini leveraging the phylogenomic information content, e.g. reticulation, hard v.

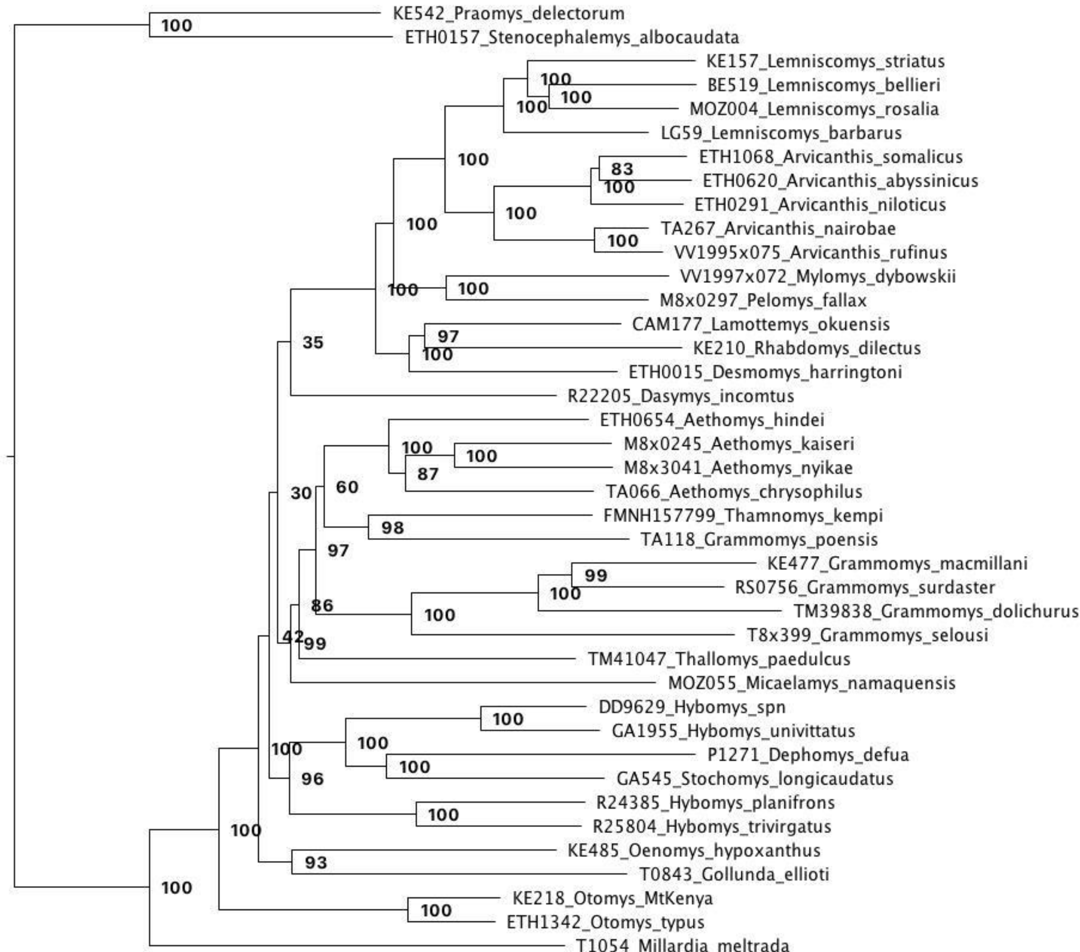
soft polytomies, etc… could make the study more relevant to Syst. Bio. readers.

Response: The previous version of the manuscript was rejected in Systematic Biology based on this review. However, we agree only partially with it, as we specify in details below. Furthermore, the reviewer states that the paper is not relevant to Systematic Biology (and the editors of Syst Biol followed this opinion); after its rejection we discussed with co-authors and we found Molecular Phylogenetics and Evolution as the most suitable journal for this manuscript. The reviewer just copy-paste the review from Syst Biol and say nothing about the relevance of the manuscript for Mol Phyl Evol.
Specific responses:
(1) Discussion on evolutionary history of Arvicanthini. This part of Discussion is based on the comparison of completely resolved phylogeny (for the first time in this group) with known palaeoecological data. In our opinion, **this is relevant to broad audience of Mol Phyl Evol**, especially to readers interested in evolution of mammals and African biodiversity in general. There is no formal analysis of discussed traits (mostly preferences of particular habitat types), but we think that it is not necessary - the discussed patterns are so obvious on the phylogenetic tree that they do not require any formal tests. We agree that some statements are more speculative than others (and we do not conceal it), but in our opinion even the speculative explanations can be proposed in the Discussion.
(2) The nomenclatural implications are based on the fully resolved phylogeny, for the first time based on genomic data, and **they are highly relevant to Mol Phyl Evol** (one of the main aims of the journal is "the development of phylogenetically more accurate taxonomic classifications")

Major Comments
(1) With this scale of data, the branch support values from all analyses (ASTRAL, MrBayes, RAxML) are likely to be inflated and could be positively misleading. Figure 1 reports only one node with less than complete support in the nuclear analyses, which should be somewhat surprising for this rapid African radiation. Additional, branch support values should be considered, especially those that incorporate resampling approaches (e.g. UFBOOT2) should be implemented. Consider IQTREE for implementing alternative bootstrap resampling methods not available in RAxML. See Roycroft et al., 2019 for a dissection of misleading branch support values with comparable data and taxon sampling. Roycroft, E. J., Moussalli, A., & Rowe, K. C. (2019). Phylogenomics Uncovers Confidence and Conflict in the Rapid Radiation of Australo-Papuan Rodents. Systematic Biology. See also Giarla, T. C., & Esselstyn, J. A. (2015). The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. Systematic Biology, 64(5), 727-740.

Response: Contrary to the reviewer, we do not see any reason for inflation of posterior probabilities estimated in MrBayes and also the suggested papers mention this problem only in the context of concatenated analyses. Similarly, we are not aware of any work demonstrating ASTRAL's local PPs to be inflated with increasing scale of data (in other words, demonstrating estimates of local PPs to be statistically inconsistent). The problem could possibly arise with RAxML concatenated analysis, so we re-analyzed concatenated datasets in IQTREE (v. 2.1.2) to check for it. The IQTREE results are, however, congruent with those of RAxML. In the concatenated nuclear data set, the maximum likelihood topology is identical to that of RAxML and the unbiased bootstrap values (UFBOOT2) of all but a single node are 100. The only exception is the node showing *Thallomys* as sister of *Grammomys*+*Thamnomys*, where RAxML bootstrap is 96 and IQTREE bootstrap is 97. In

mitogenome analysis, the maximum likelihood topologies of IQTREE and RAxML are identical and bootstrap values correlated, as apparent form comparison of the below picture (IQTREE) to the right panel of our Figure 1 (RAxML). We therefore decided to keep the RAxML results, but we are open to editorial suggestions and we can easily replace RAxML by (almost identical) IQTREE results.



(2) In general, the phylogenomic information should be leveraged much more to understand the nature of recalcitrant nodes. Is there evidence for well-supported conflict among gene trees at particularly nodes (consider the delta log-likelihood approach)? Are there signatures of reticulation in the data? Are summary coalescent (ASTRAL) and concatenation (MrBayes, RAxML) methods in total agreement? What are the implications of this? See comments above about branch support.

Response: Indeed, topology of the ML concatenated and summary coalescent tree of nuclear data is in total agreement. This is the case in spite of rapid radiation at the basis of Arvicanthini, which could create an anomaly zone and cause the concatenated tree to differ from the summary coalescent tree. Topological differences are present only between mtDNA and nuclear trees. There are six of them, two involving poorly resolved mitochondrial relationships. In three others, nuclear topology was dominant in gene trees. These include recent splits within *Lemniscomys* and *Arvicanthis*), where mitochondrial introgression should be considered. We mentioned it in the section “4.1 Anchored phylogenomics vs. complete mtDNA”. Finally, the position of *Oenomys* was found very variable, which is now also discussed in the paper along with a possibility of a hard polytomy separating *Golunda*,

Oenomys and the rest of Arvicanthini. For the well supported conflict we report percentage of nuclear- and mitochondrial-congruent loci. We limited analysis of recalcitrant nodes to this enumeration because we do not deal with difference between results of different methods or differences between different nuclear genomic data sets, which is a subject of edge testing analyses (Walker et al. 2018 Syst. Biol. 67:916-924, Smith et al. 2020 Syst. Biol. 69:579–592). Our focus was also not on phylogenetic patterns of different mitochondrial genes (where delta log-likelihood could be applied).

(3) That many independent loci have more power to resolve a phylogeny than a single locus, mitogenome or not, is not surprising and has been long-recognised (e.g. Steppan et al. 2005 among many others). In this study, the comparison of the mitogenome and nuclear phylogenies is rather cursory and makes the verbal, non-statistical, argument that the mitogenome is wrong. This does not provide any new insight. Do poorly resolved nodes in the mitogenome data have signatures of conflict among any of the nuclear loci? How many nuclear loci support the alternative and well-supported topologies apparent in the mtDNA data (e.g. sister relationship of Oenomys and Golunda)? Steppan, S. J., Adkins, R. M., Spinks, P. Q., & Hale, C. (2005). Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. Molecular phylogenetics and evolution, 37(2), 370-388.

Response: Indeed, theory clearly predicts a single non-recombining locus (no matter how long) has less power to resolve a species tree. It is not so common, however, to demonstrate validity of this theoretical prediction with mitogenomes and hundreds of nuclear loci (obtained from the very same specimens). Furthermore, even when focus is specifically on the mitochondrial tree it is far from clear when the phylogenetic signal of mtDNA vanishes in particular group. After all, phylogenetic studies based solely on mitogenomes are relatively common, in spite of inherent limits of single-locus analyses, and this fact makes our comparison of nuclear and mitochondrial tree useful.

(4) How does this phylogenomic dataset compare to that in Roycroft et al., 2019? Worth discussing as both are used to resolve adaptive radiations in Murinae and have comparable sampling?

Response: We do not see any reason to discuss our phylogenomic dataset ("anchored phylogenomics") with that used by Roycroft et al. ("exon capture"). The rodent taxa, used phylogenomic approaches and biogeographical regions are completely different. We do not understand what should be the matter of such comparison.

(5) There are numerous grammatical errors throughout the text that would need to be edited in a revised manuscript.

Response: The English was reviewed by native speakers and we hope that the text is now at publishable level.

Minor Comments

(1) Consider using the ASM Mammal Diversity Database for current counts of species diversity (https://mammaldiversity.org)

Response: We used an alternative reference, the recent Handbook of the Mammals of the World, published in 2017.

(2) Consider the murine tribes defined in Rowe, K. C., Achmadi, A. S., Fabre, P. H., Schenk, J. J., Steppan, S. J., & Esselstyn, J. A. (2019). Oceanic islands of Wallacea as a source for dispersal and diversification of murine rodents. Journal of Biogeography, 46(12), 2752-2768.

Response: The suggested paper does not bring any new view/definition of murine tribes, especially those in Africa. As a reference for the number of the tribes, we used the newest Handbook of the Mammals of the World (2017), which we feel is highly relevant and includes updated information.

(3) L. 515 Why assume that the loci are selectively neutral? Why not test that?

Response: We agree AHE loci should be investigated in this respect, but this is beyond the scope of the present paper. Encouragingly, in UCEs (i.e. similar type of genomic elements), distribution of gene trees was found similar to that expected under neutral coalescent process (Faircloth et al. 2012).

(4) L. 519-521. What are the interesting hypotheses worth testing? How will they challenge the generic classification of Arvicanthini?

Response: We do not understand this comment. There is nothing about testing of hypotheses and generic classification at L. 519-521.

(5) Is it 'savanna' or 'savannah'? Both are used in this paper.

Response: Corrected - we use savanna everywhere.

Reviewer #2: This is an interesting paper that makes use of +300 loci to fully resolve the phylogenetic relationships of the Arvicanthini rodents. The analyses are well performed and I have relatively minor comments on the manuscript that I hope the authors find useful.

Response: Thank you for the very positive review of our manuscript. We also thank for the constructive comments that helped to increase its quality. We comment on them below.

Comments
The paper is in need of dedicated editing by a native English speaker (e.g. the Lemmon's). There are many word constructions that can be improved: e.g. line 57, "calibration of a molecular clock"; line 70 "They colonized the whole of sub-Saharan Africa"; lines 87 "markers used" rather than "used markers"; line 88, "rapid radiation" rather than "intense radiation"; line 546, "worthy of testing" rather than "worth of testing"; and other similar instances through the paper.

Response: We thank you very much for these suggestions that we accepted. The text was also checked by native English speakers and we hope the text is now properly edited.

Line 66 - are both the words 'currently' and 'modern' needed? I think you can delete modern.

Response: "modern" was replaced by "extant".

Lines 67-70. I would add the number of species in ( ) after each genus name.

Response: We added this information and relevant reference.

Lines 77 and part of the discussion. Why are the Arvicanthini an adaptive radiation? They are a radiation but no data is presented in this manuscript to suggest it is an adaptive radiation in the traditional sense? They could just as easily be a non-adaptive radiation. I think the wording needs to be tightened.

Response: The radiation of Arvicanthini appeared soon after their arrival to Africa from Asia. They very quickly evolved forms able to live in very diverse environments (from semi-deserts to tropical rain forests), which is described in Introduction and is in agreement with the definition of adaptive radiation. However, to make is less stringent, we put the word "adaptive" at r. 80 in parentheses.

Line 186. Which version of ASTRAL was used? ASTRAL III has some significant improvements over earlier versions.

Response: We used ASTRAL II, but re-analysis in ASTRAL III (v 5.7.3) produced identical results.

Line 193. There is an extra) at end of sentence.

Response: Corrected.

ASTRAL analyses. It is unclear to me what the advantage is of using a Bayesian approach to estimate the individual gene trees and then having to come up with a way to choose a tree as input for the ASTRAL analyses. Why not use RAXML to estimate individual gene trees and use these as input to ASTRAL? This is what most studies that make use of ASTRAL use and is less subject to user selection of a Bayesian tree, which may introduce bias into the species tree estimation.

Response: We do not see any theoretical reason why should be Bayesian maximum clade credibility tree less accurate than the maximum likelihood tree. Actually, the prevalent use of ML gene trees as inputs for ASTRAL may be due to technical limitations as it is not always easy to get access to computational resources allowing hundreds or thousands of parallel Bayesian analyses.

Line 234. Why were the outgroups removed in the StarBeast2 analyses?

Response: The root position inference is an integral part of StarBEAST analysis and there is no need of outgroups, therefore.

What tree prior was used in the StarBeast2 analyses?

Response: We used birth-death prior and the information is now included in the text.

Lines 248-275. This method to choice a subset of loci for the dating analyses is interesting and not a method I am familiar with. Can citations be provided to the original description and validation of this method?

Response: The method was developed for this paper, but, finally, we abandoned it. Our dating is now based on 231 loci that have no missing sequences and comply with the strict

clock model of molecular evolution. This information is updated in the revised version of the manuscript.

```
Lines 248-275. I realize this would be more work, but it would be
interesting to see whether the estimates at nodes change if the StarBeast2
analyses were repeated with the next most informative 40 loci. This would
provide some confidence in the resulting dating of key nodes discussed
later in the manuscript.
```

Response: Thank you for the suggestion. We explored the issue thoroughly, ending up with the solution mentioned in the previous response (and in the revised text). Our exploration included analyses of loci selected by the original subset method, fifty randomly chosen strict clock loci, all 231 strict clock loci and all 269 loci with no missing sequences (under strict clock model).

```
Line 289 - make it clear that this sentence is referring to mtDNA ---
"unambiguous mtDNA alignments".
```

Response: Corrected.

```
Line 382 - I understand that sentiment of "slightly useless" given the
challenge with determining the origin of the group, but I would suggest
this be reworded.
```

Response: We used "*even more complex*".

```
Section 4.3 - See earlier comments about Adaptive radiation.
```

Response: See our explanation above. Anyway, we put "adaptive" in parentheses in the title of the section and we deleted it in the first paragraph to tone down the statement that the radiation was adaptive.

```
Data availability.
The anchored tag data should be submitted to the NBCI short read archive
and the accession number provided. I believe this is a requirement of
publishing in MPE. I would also suggest that the alignments of each locus
be added to the GitHub depository to enable others to repeat and build upon
the current analyses.
```

Response: We will discuss this issue with the editors of MPE. In the revised version of the manuscript, we provide the link to the alignments of all anchored phylogenomic markers at the permanent open access repository of the Czech Academy of Sciences, but we are open to any other suggestion.

Response to Reviewers:

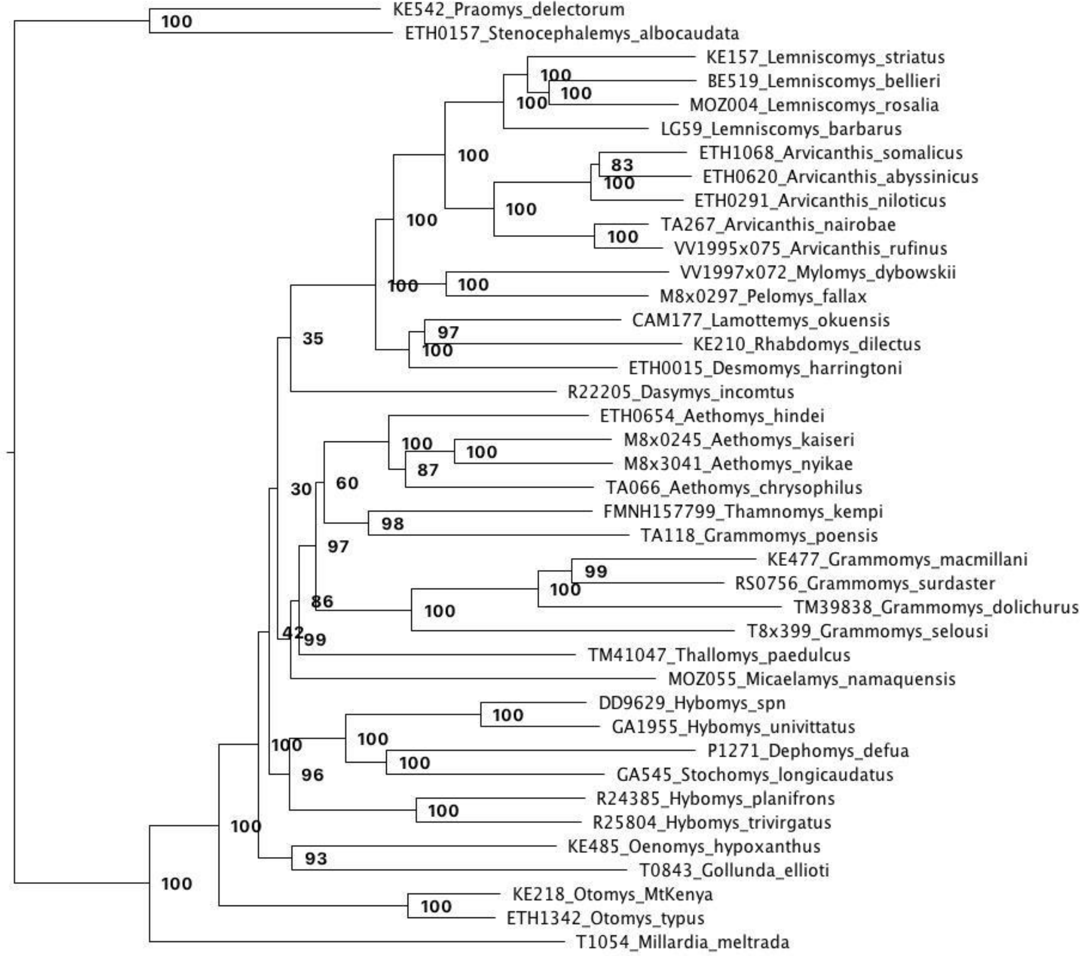
Reviewer #1: This manuscript presents a phylogenomic analysis for one the most species-rich radiations of mammals centered in Africa, the murine rodent tribe Arvicanthini. The data include mitogenomes and 377 independent nuclear loci comprising nearly 600,000 bp from 40 species including all named genera in Arvicanthini. The data are compelling and the Arvicanthini is a compelling continental radiation ideal for demonstrating the utility and application of many locus, phylogenomic analyses. However, to make this study most valuable to Systematic Biology readers the phylogenomic analyses need much deeper interrogation. The only discussion of the phylogenomic analyses is a short comparison of mitogenomes and nuclear loci. Most of the discussion is devoted to a visual interpretation of biogeographic and ecological traits on the single best topology, most of which is not relevant to a broad audience. There also are no formal analysis of these traits, which should use posterior distributions of topologies. In addition, much of this section is speculation, including over equivocal alternatives (e.g. the biogeographic origin of Arvicanthini). The nomenclatural implications for genera are valid, but of little relevance to most Sys. Bio. readers. They also are not resolved and many have been noted previously without phylogenomic data. More needs to be done to demonstrate the information content and utility of the phylogenomic data and/or provide guidance in best practices for analysing these data. Alternatively, a formal analysis of signatures of adaptive radiation in Arvicanthini leveraging the phylogenomic information content, e.g. reticulation, hard v. soft polytomies, etc… could make the study more relevant to Syst. Bio. readers.

Response: The previous version of the manuscript was rejected in Systematic Biology based on this review. However, we agree only partially with it, as we specify in details below. Furthermore, the reviewer states that the paper is not relevant to Systematic Biology (and the editors of Syst Biol followed this opinion); after its rejection we discussed with co-authors and we found Molecular Phylogenetics and Evolution as the most suitable journal for this manuscript. The reviewer just copy-paste the review from Syst Biol and say nothing about the relevance of the manuscript for Mol Phyl Evol.
Specific responses:
(1) Discussion on evolutionary history of Arvicanthini. This part of Discussion is based on the comparison of completely resolved phylogeny (for the first time in this group) with known palaeoecological data. In our opinion, **this is relevant to broad audience of Mol Phyl Evol**, especially to readers interested in evolution of mammals and African biodiversity in general. There is no formal analysis of discussed traits (mostly preferences of particular habitat types), but we think that it is not necessary - the discussed patterns are so obvious on the phylogenetic tree that they do not require any formal tests. We agree that some statements are more speculative than others (and we do not conceal it), but in our opinion even the speculative explanations can be proposed in the Discussion.
(2) The nomenclatural implications are based on the fully resolved phylogeny, for the first time based on genomic data, and **they are highly relevant to Mol Phyl Evol** (one of the main aims of the journal is "the development of phylogenetically more accurate taxonomic classifications")

Major Comments
(1) With this scale of data, the branch support values from all analyses (ASTRAL, MrBayes, RAxML) are likely to be inflated and could be

positively misleading. Figure 1 reports only one node with less than complete support in the nuclear analyses, which should be somewhat surprising for this rapid African radiation. Additional, branch support values should be considered, especially those that incorporate resampling approaches (e.g. UFBOOT2) should be implemented. Consider IQTREE for implementing alternative bootstrap resampling methods not available in RAxML. See Roycroft et al., 2019 for a dissection of misleading branch support values with comparable data and taxon sampling. Roycroft, E. J., Moussalli, A., & Rowe, K. C. (2019). Phylogenomics Uncovers Confidence and Conflict in the Rapid Radiation of Australo-Papuan Rodents. Systematic Biology. See also Giarla, T. C., & Esselstyn, J. A. (2015). The challenges of resolving a rapid, recent radiation: empirical and simulated phylogenomics of Philippine shrews. Systematic Biology, 64(5), 727-740.

Response: Contrary to the reviewer, we do not see any reason for inflation of posterior probabilities estimated in MrBayes and also the suggested papers mention this problem only in the context of concatenated analyses. Similarly, we are not aware of any work demonstrating ASTRAL's local PPs to be inflated with increasing scale of data (in other words, demonstrating estimates of local PPs to be statistically inconsistent). The problem could possibly arise with RAxML concatenated analysis, so we re-analyzed concatenated datasets in IQTREE (v. 2.1.2) to check for it. The IQTREE results are, however, congruent with those of RAxML. In the concatenated nuclear data set, the maximum likelihood topology is identical to that of RAxML and the unbiased bootstrap values (UFBOOT2) of all but a single node are 100. The only exception is the node showing *Thallomys* as sister of *Grammomys+Thamnomys*, where RAxML bootstrap is 96 and IQTREE bootstrap is 97. In mitogenome analysis, the maximum likelihood topologies of IQTREE and RAxML are identical and bootstrap values correlated, as apparent form comparison of the below picture (IQTREE) to the right panel of our Figure 1 (RAxML). We therefore decided to keep the RAxML results, but we are open to editorial suggestions and we can easily replace RAxML by (almost identical) IQTREE results.



(2)   In general, the phylogenomic information should be leveraged much more to understand the nature of recalcitrant nodes. Is there evidence for well-supported conflict among gene trees at particularly nodes (consider the delta log-likelihood approach)? Are there signatures of reticulation in the data? Are summary coalescent (ASTRAL) and concatenation (MrBayes, RAxML) methods in total agreement? What are the implications of this? See comments above about branch support.

Response: Indeed, topology of the ML concatenated and summary coalescent tree of nuclear data is in total agreement. This is the case in spite of rapid radiation at the basis of Arvicanthini, which could create an anomaly zone and cause the concatenated tree to differ from the summary coalescent tree. Topological differences are present only between mtDNA and nuclear trees. There are six of them, two involving poorly resolved mitochondrial relationships. In three others, nuclear topology was dominant in gene trees. These include recent splits within *Lemniscomys* and *Arvicanthis*), where mitochondrial introgression should be considered. We mentioned it in the section "4.1 Anchored phylogenomics vs. complete mtDNA". Finally, the position of *Oenomys* was found very variable, which is now also discussed in the paper along with a possibility of a hard polytomy separating *Golunda*, *Oenomys* and the rest of Arvicanthini. For the well supported conflict we report percentage of nuclear- and mitochondrial-congruent loci. We limited analysis of recalcitrant nodes to this enumeration because we do not deal with difference between results of different methods or differences between different nuclear genomic data sets, which is a subject of edge testing analyses (Walker et al. 2018 Syst. Biol. 67:916-924, Smith et al. 2020 Syst. Biol. 69:579–

592). Our focus was also not on phylogenetic patterns of different mitochondrial genes (where delta log-likelihood could be applied).

(3) That many independent loci have more power to resolve a phylogeny than a single locus, mitogenome or not, is not surprising and has been long-recognised (e.g. Steppan et al. 2005 among many others). In this study, the comparison of the mitogenome and nuclear phylogenies is rather cursory and makes the verbal, non-statistical, argument that the mitogenome is wrong. This does not provide any new insight. Do poorly resolved nodes in the mitogenome data have signatures of conflict among any of the nuclear loci? How many nuclear loci support the alternative and well-supported topologies apparent in the mtDNA data (e.g. sister relationship of Oenomys and Golunda)? Steppan, S. J., Adkins, R. M., Spinks, P. Q., & Hale, C. (2005). Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. Molecular phylogenetics and evolution, 37(2), 370-388.

Response: Indeed, theory clearly predicts a single non-recombining locus (no matter how long) has less power to resolve a species tree. It is not so common, however, to demonstrate validity of this theoretical prediction with mitogenomes and hundreds of nuclear loci (obtained from the very same specimens). Furthermore, even when focus is specifically on the mitochondrial tree it is far from clear when the phylogenetic signal of mtDNA vanishes in particular group. After all, phylogenetic studies based solely on mitogenomes are relatively common, in spite of inherent limits of single-locus analyses, and this fact makes our comparison of nuclear and mitochondrial tree useful.

(4) How does this phylogenomic dataset compare to that in Roycroft et al., 2019? Worth discussing as both are used to resolve adaptive radiations in Murinae and have comparable sampling?

Response: We do not see any reason to discuss our phylogenomic dataset ("anchored phylogenomics") with that used by Roycroft et al. ("exon capture"). The rodent taxa, used phylogenomic approaches and biogeographical regions are completely different. We do not understand what should be the matter of such comparison.

(5) There are numerous grammatical errors throughout the text that would need to be edited in a revised manuscript.

Response: The English was reviewed by native speakers and we hope that the text is now at publishable level.

Minor Comments

(1) Consider using the ASM Mammal Diversity Database for current counts of species diversity (https://mammaldiversity.org)

Response: We used an alternative reference, the recent Handbook of the Mammals of the World, published in 2017.

(2) Consider the murine tribes defined in Rowe, K. C., Achmadi, A. S., Fabre, P. H., Schenk, J. J., Steppan, S. J., & Esselstyn, J. A. (2019). Oceanic islands of Wallacea as a source for dispersal and diversification of murine rodents. Journal of Biogeography, 46(12), 2752-2768.

Response: The suggested paper does not bring any new view/definition of murine tribes, especially those in Africa. As a reference for the number of the tribes, we used the newest Handbook of the Mammals of the World (2017), which we feel is highly relevant and includes updated information.

(3) L. 515 Why assume that the loci are selectively neutral? Why not test that?

Response: We agree AHE loci should be investigated in this respect, but this is beyond the scope of the present paper. Encouragingly, in UCEs (i.e. similar type of genomic elements), distribution of gene trees was found similar to that expected under neutral coalescent process (Faircloth et al. 2012).

(4) L. 519-521. What are the interesting hypotheses worth testing? How will they challenge the generic classification of Arvicanthini?

Response: We do not understand this comment. There is nothing about testing of hypotheses and generic classification at L. 519-521.

(5) Is it 'savanna' or 'savannah'? Both are used in this paper.

Response: Corrected - we use savanna everywhere.

Reviewer #2: This is an interesting paper that makes use of +300 loci to fully resolve the phylogenetic relationships of the Arvicanthini rodents. The analyses are well performed and I have relatively minor comments on the manuscript that I hope the authors find useful.

Response: Thank you for the very positive review of our manuscript. We also thank for the constructive comments that helped to increase its quality. We comment on them below.

Comments
The paper is in need of dedicated editing by a native English speaker (e.g. the Lemmon's). There are many word constructions that can be improved: e.g. line 57, "calibration of a molecular clock"; line 70 "They colonized the whole of sub-Saharan Africa"; lines 87 "markers used" rather than "used markers"; line 88, "rapid radiation" rather than "intense radiation"; line 546, "worthy of testing" rather than "worth of testing"; and other similar instances through the paper.

Response: We thank you very much for these suggestions that we accepted. The text was also checked by native English speakers and we hope the text is now properly edited.

Line 66 - are both the words 'currently' and 'modern' needed? I think you can delete modern.

Response: "modern" was replaced by "extant".

Lines 67-70. I would add the number of species in ( ) after each genus name.

Response: We added this information and relevant reference.

```
Lines 77 and part of the discussion.  Why are the Arvicanthini an adaptive
radiation? They are a radiation but no data is presented in this manuscript
to suggest it is an adaptive radiation in the traditional sense? They could
just as easily be a non-adaptive radiation. I think the wording needs to be
tightened.
```

Response: The radiation of Arvicanthini appeared soon after their arrival to Africa from Asia. They very quickly evolved forms able to live in very diverse environments (from semi-deserts to tropical rain forests), which is described in Introduction and is in agreement with the definition of adaptive radiation. However, to make is less stringent, we put the word "adaptive" at r. 80 in parentheses.

```
Line 186. Which version of ASTRAL was used?  ASTRAL III has some
significant improvements over earlier versions.
```

Response: We used ASTRAL II, but re-analysis in ASTRAL III (v 5.7.3) produced identical results.

```
Line 193. There is an extra ) at end of sentence.
```

Response: Corrected.

```
ASTRAL analyses. It is unclear to me what the advantage is of using a
Bayesian approach to estimate the individual gene trees and then having to
come up with a way to choose a tree as input for the ASTRAL analyses. Why
not use RAXML to estimate individual gene trees and use these as input to
ASTRAL? This is what most studies that make use of ASTRAL use and is less
subject to user selection of a Bayesian tree, which may introduce bias into
the species tree estimation.
```

Response: We do not see any theoretical reason why should be Bayesian maximum clade credibility tree less accurate than the maximum likelihood tree. Actually, the prevalent use of ML gene trees as inputs for ASTRAL may be due to technical limitations as it is not always easy to get access to computational resources allowing hundreds or thousands of parallel Bayesian analyses.

```
Line 234. Why were the outgroups removed in the StarBeast2 analyses?
```

Response: The root position inference is an integral part of StarBEAST analysis and there is no need of outgroups, therefore.

```
What tree prior was used in the StarBeast2 analyses?
```

Response: We used birth-death prior and the information is now included in the text.

```
Lines 248-275. This method to choice a subset of loci for the dating
analyses is interesting and not a method I am familiar with. Can citations
be provided to the original description and validation of this method?
```

Response: The method was developed for this paper, but, finally, we abandoned it. Our dating is now based on 231 loci that have no missing sequences and comply with the strict clock model of molecular evolution. This information is updated in the revised version of the manuscript.

Lines 248-275. I realize this would be more work, but it would be interesting to see whether the estimates at nodes change if the StarBeast2 analyses were repeated with the next most informative 40 loci. This would provide some confidence in the resulting dating of key nodes discussed later in the manuscript.

Response: Thank you for the suggestion. We explored the issue thoroughly, ending up with the solution mentioned in the previous response (and in the revised text). Our exploration included analyses of loci selected by the original subset method, fifty randomly chosen strict clock loci, all 231 strict clock loci and all 269 loci with no missing sequences (under strict clock model).

Line 289 - make it clear that this sentence is referring to mtDNA --- "unambiguous mtDNA alignments".

Response: Corrected.

Line 382 - I understand that sentiment of "slightly useless" given the challenge with determining the origin of the group, but I would suggest this be reworded.

Response: We used "*even more complex*".

Section 4.3 - See earlier comments about Adaptive radiation.

Response: See our explanation above. Anyway, we put "adaptive" in parentheses in the title of the section and we deleted it in the first paragraph to tone down the statement that the radiation was adaptive.

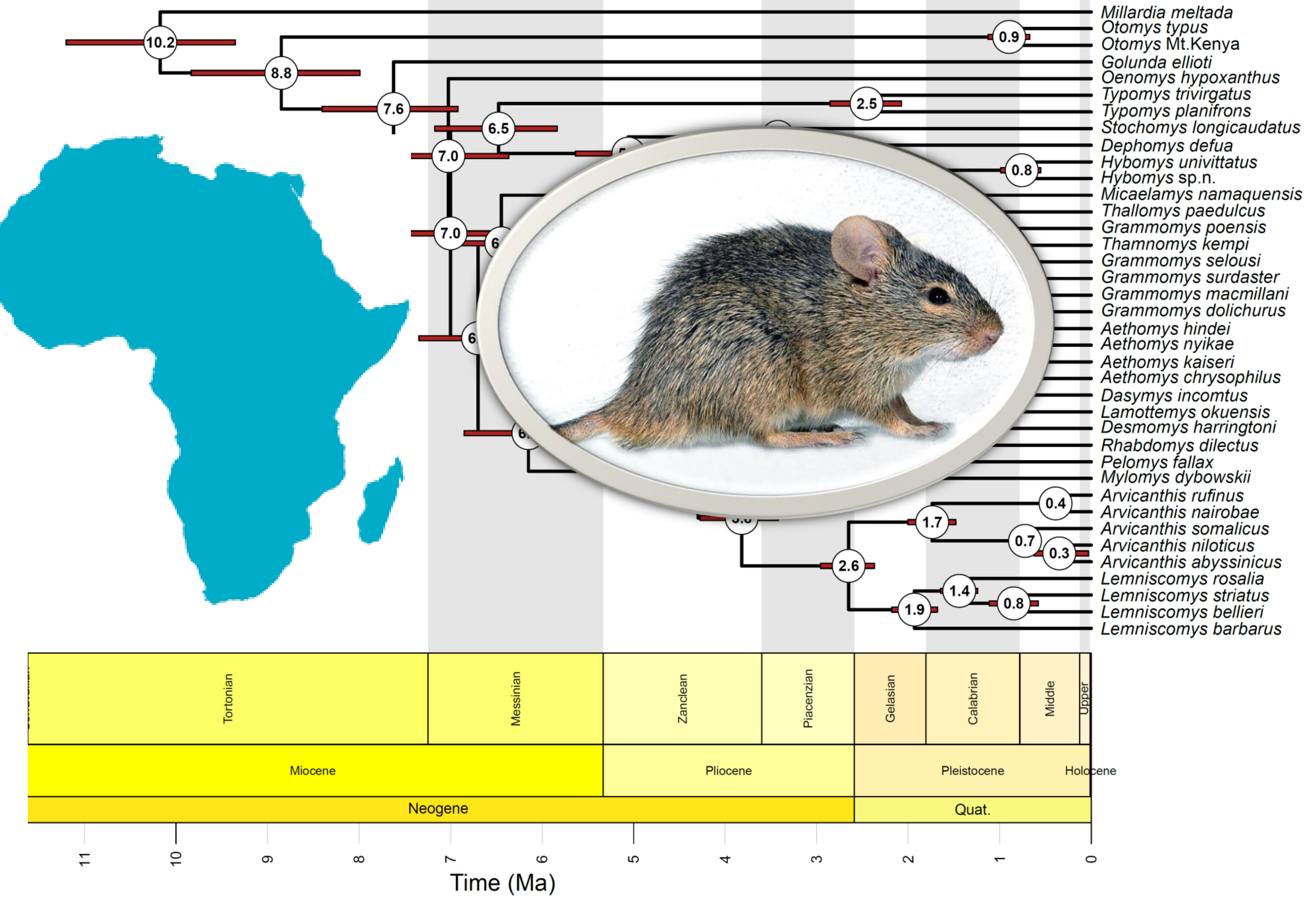
Data availability.
The anchored tag data should be submitted to the NBCI short read archive and the accession number provided. I believe this is a requirement of publishing in MPE. I would also suggest that the alignments of each locus be added to the GitHub depository to enable others to repeat and build upon the current analyses.

Response: We will discuss this issue with the editors of MPE. In the revised version of the manuscript, we provide the link to the alignments of all anchored phylogenomic markers at the permanent open access repository of the Czech Academy of Sciences, but we are open to any other suggestion.

Highlights

- fully resolved phylogeny of a highly diversified group of African mammals
- comparison of "anchored phylogenomics" and mitogenomics
- mechanisms of adaptive radiation starting in the Messinian stage (ca. 7 Ma)
- delimitation of genera in Arvicanthini and corresponding taxonomic changes

Graphical Abstract



Manuscript File

1 **Nuclear phylogenomics, but not mitogenomics, resolves the most successful Late Miocene
2 radiation of African mammals (Rodentia: Muridae: Arvicanthini)**
3

4 Ondřej Mikula [a,b], Violaine Nicolas [c], Radim Šumbera [d], Adam Konečný [e], Christiane Denys [c], Erik
5 Verheyen [f,g], Anna Bryjová [a], Alan R. Lemmon [h], Emily Moriarty Lemmon [i], and Josef Bryja [a,e,*]
6
7 [a] Institute of Vertebrate Biology of the Czech Academy of Sciences, 603 65, Brno, Czech Republic
8 [b] Institute of Animal Physiology and Genetics of the Czech Academy of Sciences, 602 00, Brno,
9 Czech Republic
10 [c] Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum national d'Histoire naturelle,
11 CNRS, Sorbonne Université, EPHE , Université des Antilles, CP51, 75005 Paris, France;
12 [d] Department of Zoology, Faculty of Science, University of South Bohemia, 370 05, České
13 Budějovice, Czech Republic
14 [e] Department of Botany and Zoology, Faculty of Science, Masaryk University, 611 37, Brno, Czech
15 Republic
16 [f] Royal Belgian Institute for Natural Sciences, Operational Direction Taxonomy and Phylogeny,
17 1000, Brussels, Belgium
18 [g] Evolutionary Ecology Group, Biology Department, University of Antwerp, 2020, Antwerp,
19 Belgium
20 [h] Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee,
21 FL 32306-4295, United States
22 [i] Department of Biological Science, Florida State University, 319 Stadium Drive, PO Box 3064295,
23 Tallahassee, FL 32306-4295, United States
24

25 ***Correspondence**: Josef Bryja, Institute of Vertebrate Biology of the Czech Academy of Sciences,

26 Research Facility Studenec, Studenec 122, 675 02 Koněšín, Czech Republic; E-mail: bryja@ivb.cz

27

28 ***Running head*:** Phylogenomics of Arvicanthini rodents

29

ABSTRACT

The tribe Arvicanthini (Muridae: Murinae) is a highly diversified group of rodents (ca. 100 species) and with 18 African genera (plus one Asiatic) represents probably the most successful adaptive radiation of extant mammals in Africa. They colonized a broad spectrum of habitats (from rainforests to semi-deserts) in whole sub-Saharan Africa and their members often belong to most abundant parts of mammal communities. Despite intensive efforts, the phylogenetic relationships among major lineages (i.e. genera) remained obscured, which was likely caused by the intensive radiation of the group, dated to the Late Miocene. Here we used genomic scale data (377 nuclear loci; 581,030 bp) and produced the first fully resolved species tree containing all currently delimited genera of the tribe. Mitogenomes were also extracted, and while the results were largely congruent, there was less resolution at basal nodes of the mitochondrial phylogeny. Using a newly developed algorithm for subsampling of most informative loci, we also performed a fossil-based divergence dating. The results suggest that the African radiation started early after the colonization of Africa by a single arvicanthine ancestor from Asia during the Messinian stage (ca. 7 Ma), and was likely linked with a fragmentation of the pan-African Miocene forest. Some lineages remained in the rain forest, while many others successfully colonized broad spectrum of new open habitats (e.g. savannas, wetlands or montane moorlands) that appeared at the beginning of Pliocene. One lineage even evolved partially arboricolous life style in savanna woodlands, which allowed them to re-colonize equatorial forests. We also discuss delimitation of genera in Arvicanthini and propose corresponding taxonomic changes.

Keywords: Late Miocene, radiation, anchored phylogenomics, Rodentia, tropical Africa, complete mitochondrial DNA

53 **1. Introduction**

54 The murid rodents (Rodentia: Muridae) are evolutionarily the most successful group of mammals
55 in the Old World, with 816 currently recognized species (Wilson et al., 2017). Their phylogeny is
56 relatively well known thanks to recent analyses of large multi-locus genetic datasets and
57 calibration of a molecular clock based on multiple paleontological records (e.g. Steppan and
58 Schenk, 2017; Aghová et al., 2018). Among five subfamilies, Murinae form the majority of murid
59 rodents (ca. 80%; Wilson et al., 2017). They evolved in *ca* 15 major clades (= tribes) (Steppan and
60 Schenk, 2017) with very unequal distribution of species diversity (a single species in Micromyini
61 vs. 185 species in Rattini; Wilson et al., 2017). Five murine tribes (Otomyini, Arvicanthini,
62 Malacomyini, Murini, Praomyini) are indigenous in sub-Saharan Africa (Lecompte et al., 2008) and
63 they constitute the most species-rich group of African mammals.

64

65 The tribe Arvicanthini (Lecompte et al., 2008; Denys et al., 2017) is the most speciose tribe of
66 African rodents with 18 currently recognized ~~modern~~ extant African genera and the Asiatic genus
67 *Golunda* (Denys et al., 2017; Missoup et al., 2018; Table S1 in SM1). Some genera are species-rich
68 and widely distributed (e.g. *Lemniscomys* - 11 species, *Aethomys* - 9 species, *Grammomys* - 14
69 species), while others have low diversity and restricted ranges, ~~(~~ e.g. one species of *Lamottemys*
70 from Mt. Oku in Cameroon, or two species of *Desmomys* from Ethiopian highlands (Denys et al.
71 2017). They colonized the whole of sub-Saharan Africa (two species have also isolated populations
72 in Maghreb and Egypt), where they live in a broad spectrum of habitats; from lowland and
73 montane rainforests through various types of open habitats (marshlands, savannas, woodlands)
74 to semi-deserts. The first radiation of the tribe occurred in Late Miocene after arrival of
75 arvicanthine ancestor(s) from Asia (7-9 Mya; Aghová et al. 2018) and was likely related to intensive
76 climatic changes and a spread of open habitats (Lecompte et al., 2008). Most modern genera of

Arvicanthini appeared almost simultaneously and they can serve as a model for understanding the evolutionary process of (adaptive) radiation.

A reliable phylogenetic reconstruction is required for deciphering mechanisms of such successful radiation. However, despite intensive efforts, the phylogenetic relationships among many genera of Arvicanthini are still uncertain. Previous studies employed mitochondrial (Ducroz et al., 2001) or the combination of a limited number of mitochondrial and nuclear sequences (Lecompte et al., 2008; Missoup et al., 2016; Bryja et al., 2017; Steppan and Schenk, 2017; Aghová et al., 2018; Missoup et al., 2018). Even if these studies agreed e.g. on the monophyly of the so-called *Hybomys* division (sensu Musser and Carleton, 2005) or the sister relationship of *Lemniscomys* and *Arvicanthis*, numerous (especially deeper) nodes on the phylogenetic tree remained unresolved. They are either unsupported or have conflicting topologies dependent on the ~~used~~ markers used, which may be the outcome of the ~~intensive~~ rapid radiation of the Arvicanthini in Late Miocene (Aghová et al., 2018).

Increasing the amount of genetic data frequently allows resolution of even the most problematic phylogenetic relationships. One such approach is based on sequencing of complete mitochondrial genomes ("mitogenomics"), instead of single mitochondrial genes; this helped to reconstruct e.g. the phylogeny of primates (Pozzi et al., 2014) or sharks and rays (Amaral et al., 2018). However, because of the absence of recombination, the mitochondrial DNA should be still considered as a single locus and reconstructed phylogenies represent only "single gene" trees. To address this problem, recent phylogenomic approaches target markers derived from moderately conserved regions, mostly exons and surrounding introns (Lemmon et al., 2012), or ultraconserved genomic elements and their flanking regions (McCormack et al., 2012), which allow to infer a species-tree

101 that accounts for discord among hundreds of independent loci at nuclear DNA (Lemmon and
102 Lemmon, 2013). These regions are enriched in genomic libraries by hybridization and then
103 sequenced by high-throughput sequencing. They can be analysed even from old museum material
104 (e.g. McCormack et al., 2016) and they allowed solving the notoriously difficult nodes in phylogeny
105 of birds (Prum et al., 2015) or placental mammals (McCormack et al., 2012).
106
107 Here we used the so-called anchored phylogenomic approach (Lemmon et al., 2012) to infer the
108 most reliable phylogenetic tree for the Arvicanthini. This is the first multi-locus analysis including
109 all extant African genera of this clade, as well as the Asian genus *Golunda*. With the resolved
110 topology of the tribe in hand, we estimated the time-frame, during which this tribe radiated and
111 assessed its evolutionary history in the context of environmental changes since Late Miocene. As
112 a by-product of sequencing of anchored loci, we assembled also complete mtDNA from all
113 samples and we compared the ability of anchored phylogenomics vs. mitogenomics in
114 phylogenetic reconstruction of a fast mammalian radiation.
115
116 **2. Material and methods**
117 *2.1 Taxon sampling*
118 The final dataset analysed in this study includes 40 genotyped specimens (= one individual per
119 species; Table S1) representing all 18 nominal genera of the African Arvicanthini, as well as the
120 closely related Asian genus *Golunda* belonging to the same tribe (Denys et al., 2017; Missoup et
121 al., 2018). Two species of the tribe Otomyini and one of Millardiini were chosen as the closest
122 relatives of Arvicanthini and two species of the tribe Praomyini were used as more distant
123 outgroups within the subfamily Murinae (Lecompte et al., 2008; Aghová et al., 2018) (Table S1).
124

125 *2.2 Anchored hybrid enrichment (AHE) data collection and assembly of nuclear dataset*

126 Probe design and data collection were performed by the Center for Anchored Phylogenomics
127 (www.anchoredphylogeny.com). Following Ruane et al. (2015; snakes), Tucker et al. (2016;
128 lizards), and Prum et al. (2015; birds), we improved the vertebrate AHE target loci of Lemmon et
129 al. (2012) for optimal use in mammals. We first identified the genomic coordinates in the human
130 genome (hg19) corresponding to the coordinates of the extended anchor regions of *Gallus gallus*
131 (galGal4) obtained by Prum et al. (2015) using the UCSC liftover tool (http://genome.ucsc.edu/cgi-
132 bin/hgLiftOver). The corresponding genomic sequences were then extracted and aligned using
133 MAFFT v7.023b (Katoh and Standley 2013) to that of the regions used by Prum et al. (2015) for
134 probe design. After inspecting the alignments and masking any misaligned regions in Geneious R9
135 (Biomatters Ltd.; Kearse et al. 2012), 120 bp probes were tiled uniformly across the human
136 sequences at 1.5x density. Genomic DNA was extracted using the Invisorb® Spin Tissue Mini Kit
137 (Stratec, Germany). After extraction, indexed libraries were prepared on a Beckman Coulter FXP
138 liquid-handling robot following Lemmon et al. (2012) and Prum et al. (2015). Libraries were then
139 pooled at equal concentrations in three groups of ~14 samples and enriched using an Agilent
140 SureSelect XT kit containing the probes described above. Enriched library pools were then
141 sequenced on one paired-end 150 bp lane (43 Gb of raw data) of an Illumina HiSeq 2500
142 sequencer at the Translational lab in the Florida State University.

143

144 In order to increase read accuracy and length, paired reads were merged prior to assembly
145 following Rokyta et al. (2012), which also removes adapter sequences. Following the approaches
146 of Prum et al. (2015) and Hamilton et al. (2016), a quasi-de novo assembly approach was taken
147 using *Homo sapiens* as the reference. Assembly clusters derived from fewer than 175 reads were
148 removed from further analysis in order to reduce the possible effects of low level contamination

149 and mis-indexing. Orthology was established among the consensus sequences recovered at each
150 of the target loci using the pairwise sequence distances among the consensus sequences (see
151 Hamilton et al., 2016 for details). After orthologous sequences were then aligned using MAFFT
152 v7.023b (Katoh and Standley 2013; with --genafpair and --maxiterate 1000 flags utilized), the
153 alignments were trimmed/masked to remove poorly aligned regions (following Hamilton et al.
154 2016; with the following parameters: MinGoodSites=14, MinPropSame=0.4, and
155 MissingAllowed=20). Finally, trimmed alignments were inspected in Geneious and any remaining
156 misaligned regions were masked.

157

158 *2.3 Assembly and alignment of mitogenomes*

159 Mitochondrial DNA is usually highly prevalent in genomic DNA extractions and it still persists even
160 in genomic libraries enriched for particular conserved loci. As a by-product of AHE approach, we
161 therefore used the raw data of Illumina reads to assembly the complete mitogenomes of 40
162 analysed taxa. Heavy-strand protein-coding genes (12 genes) and genes for non-coding RNA (two
163 ribosomal RNAs and 22 transfer RNAs) were extracted from the complete mitochondrial
164 sequences in Geneious according to the annotated references of complete mtDNAs of *Apodemus*
165 *draco* (GenBank accession number KP694301) and *A. chevrieri* (HQ896683) from the relatively
166 closely related tribe Apodemini (Murinae). Following Pozzi et al. (2014), we excluded the D-loop
167 sequences because of alignment difficulties (highly variable non-coding sequences), and ND6
168 gene because it is encoded on the mitochondrial L-strand which has a different nucleotide
169 composition from the H-strand, and has been shown to have poor phylogenetic signal (Gissi et al.,
170 2000). Protein-coding genes were individually aligned based on their corresponding amino acid
171 translations using Muscle 3.8 (Edgar, 2004) implemented in AliView 1.18 (Larsson, 2014). Two
172 genes for ribosomal RNA (12S-rDNA and 16S-rDNA) and 22 genes for transfer RNA were aligned

173 separately by online version of MAFFT v. 7 (http://mafft.cbrc.jp/alignment/server/) using the
174 algorithm Q-INS-i, which considers secondary structure of RNA and is recommended for a global
175 alignment of highly divergent non-coding RNAs (Katoh and Toh, 2008). The resulting alignments
176 of genes for both rRNAs and tRNAs were analysed by Gblocks 0.91b (Castreana, 2000). Gblocks
177 removes all poorly aligned regions in a dataset, which has been shown to be particularly effective
178 in phylogenetic studies including very divergent sequences (Talavera and Castreana, 2007).
179 Gblocks was run with the options "Minimum Length Of A Block" = 5, and "Allowed Gap Positions"
180 = "With Half".

181

182 *2.4 Phylogenetic analysis of nuclear loci - species tree in ASTRAL*

183 Multispecies coalescent (MSC) provides sound foundation for species tree inference as it models
184 incomplete lineage sorting and hence discordance between gene trees (Degnan and Rosenberg,
185 2009). However, joint estimation of species tree and gene trees becomes too computation
186 expensive with large numbers of loci (Ogilvie et al., 2017). For this reason we inferred species tree
187 by ASTRAL II (v. 4.11.2, Mirarab and Warnow, 2015) – a summary method analysing topologies of
188 pre-estimated gene trees by breaking them into a multi-set of quartet trees and searching for
189 species tree inducing quartet tree topologies that are most frequently observed in the multi-set
190 (Mirarab and Warnow, 2015). The gene trees were obtained in separate Bayesian analyses using
191 MrBayes v. 3.2.6 (Ronquist et al., 2012). They were inferred as unrooted with uniform prior
192 probability over tree topologies. Branch lengths were unconstrained by clock assumptions and we
193 used exponential prior (μ=10.0) for each of them. Integral to the analysis was sampling of time
194 reversible nucleotide substitution models (Huelsenbeck et al., 2004) by reversible jump Markov
195 Chain Monte Carlo~~)~~. Gamma-distributed rate variation (discretized into eight categories) was
196 assumed among sites. The template of MrBayes block in the *nexus* file is available as SM2. ASTRAL

197 accepts just a single tree per gene and thus it was necessary to find a tree representing the whole
198 posterior sample obtained from MrBayes. It was defined as a maximum bipartition credibility tree
199 (MBCT), i.e. the tree with maximum product of its bipartitions' posterior probabilities (cf.
200 Drummond and Bouckaert, 2015, p. 94). Branch lengths are not used in ASTRAL and thus only
201 MBCT topology was calculated in package 'phangorn' (Schliep, 2011) for R (R Core Team, 2019).
202 In general, the gene trees were not fully resolved and the poorly supported bipartitions could
203 mislead ASTRAL. Therefore, the bipartitions with posterior probability (PP) < 0.90 were collapsed,
204 creating a polytomy in the tree. Calculation of PPs was done in 'ape' (Paradis and Schliep, 2018).
205

206 *2.5 Bayesian phylogeny of mitogenomes*

207 We used PartitionFinder v. 2 (Lanfear et al., 2016) to simultaneously detect partitions and the
208 most suitable substitution models for different parts of mtDNA. Using AICc criterion, the best
209 scheme supported 12 partitions (the partitioned *nexus* input file is in SM3). Bayesian analysis of
210 evolutionary relationships was performed in MrBayes v. 3.2.6, employing Markov Chain Monte
211 Carlo (MCMC) simulations of posterior probability. Three heated and one cold chain were
212 employed in an analysis with 12 partitions, and runs were initiated from random trees. Two
213 independent runs were conducted with 20 million generations each and trees and parameters
214 were sampled every 1000 generations. Convergence was checked using TRACER v1.5 (Rambaut
215 and Drummond, 2007). For each run, the first 20% of sampled trees were discarded as burn-in.
216 Bayesian posterior probabilities were used to assess branch support of the maximum clade
217 credibility tree with common ancestor node heights.
218

219 *2.6 Maximum likelihood estimation of mtDNA and nuclear phylogenies*

220 The statistical methods used here for the species tree and mitochondrial tree inference are
221 computationally demanding and we therefore applied also complementary and much faster
222 maximum likelihood inference in RAxML 8.2.10 (Stamatakis, 2014). The nuclear and mitochondrial
223 datasets were analysed separately. Individual loci in both datasets were concatenated and hence
224 assumed to share the same phylogeny, which is realistic only in physically linked mitochondrial
225 loci, but not in unlinked nuclear loci. Because simpler models are not available in RAxML, the
226 GTR+G model was used for all partitions, which were allowed to differ in their substitution
227 parameters. For mtDNA the partitions were defined as described above and in nuclear data set
228 every locus corresponded to a single partition. The robustness of the nodes was evaluated by the
229 rapid bootstrap procedure (Stamatakis et al., 2008) with 1000 replications.
230

231 *2.7 ~~Sub-sampling of loci and t~~Time-calibrated phylogeny*

232 The time-calibrated history of divergences between arvicanthine species was inferred in
233 StarBEAST2 (Ogilvie et al., 2017). The species tree topology as well as gene tree topologies were
234 fixed to the estimates obtained by ASTRAL and MrBayes, respectively, but the branch lengths
235 were allowed to vary. We assumed species tree to arise in a constant rate birth-death process
236 (Gernhard, 2008) with uninformative priors put on its parameters. Outgroups used to root the
237 trees were excluded in this analysis. Time information was injected into the species tree by two
238 fossil-based constraints on ages of specific ancestral nodes. Firstly, 9.2 million years (Ma) old
239 †*Karnimata darwini* (Kimura et al., 2015) constrained the age of the root, i.e. the most recent
240 common ancestor (MRCA) of Millardiini/Otomyini/Arvicanthini clade. Secondly, †*Aethomys* sp.
241 and †*Arvicanthis* sp. fossils from 6.1 Ma old site Lemudong-o' (Manthi, 2007) constrained the age
242 of MRCA of *Aethomys* and *Arvicanthis*. The fossils were taken from the set proposed for subfamily
243 Murinae by Aghová et al. (2018), but the latter two were used more conservatively to account for

244 a possibility they represent just members of lineages leading to the particular genera. The
245 calibration densities were uniform: 9.2-11.2 Ma for the root and 6.1-11.2 for
246 *Aethomys*/*Arvicanthis*. The maximum age 11.2 Ma was motivated by the fossil of †cf. *Karnimata*
247 from Nagri Formation, Siwalik Group, interpreted to be close to the split of lineages leading to
248 extant *Mus* and *Arvicanthis* (Kimura et al., 2015; Aghová et al., 2018).
249
250 For the time calibration analysis we considered just 269 out of 377 nuclear loci, namely those
251 successfully sequenced in all species and not having outgroup and ingroup species intermixed in
252 single-gene topologies. In each of these loci we compared strict and uncorrelated lognormal
253 relaxed clock (Drummond et al., 2006) using Bayes factors calculated in RevBayes v. 1.1.0 (Höhna
254 et al., 2016). For the analysis, we retained 231 loci where the strict clock model was supported.
255 Nucleotide substitution model parameters were fixed to the averages of posterior samples
256 obtained from the MrBayes analyses.
257
258 Two independent runs of the analysis were conducted to check for convergence. The pooled
259 posterior sample was represented by the Maximum Clade Credibility (MCC) tree with the
260 mean common ancestor node heights (Drummond and Bouckaert, 2015).~~To reduce the~~
261 ~~computational time we performed the time calibration analysis on just 39 out of nuclear 377 loci.~~
262 ~~First, we retained only 270 loci that were successfully sequenced in all ingroup species. Then, 39~~
263 ~~loci were selected to have best resolved MrBayes trees and to represent a variety of observed~~
264 ~~gene tree topologies. The subsampling aims to select loci with strong, yet diverse phylogenetic~~
265 ~~signal, because gene tree discordance is a natural phenomenon informative about speciation~~
266 ~~history and bias towards loci with identical gene trees might bias species tree inference as well.~~
267 ~~The selection procedure starts with a summary of posterior probabilities of particular clades~~

268 across loci. This is provided by matrix L of dimension $l \times c$ where l is the number of loci and c is
269 the total number of distinct clades observed at least once in the posterior sample of any gene
270 tree. The entries of L are posterior probabilities of particular clades at particular loci. Gene tree
271 discordance is summarized by matrix $\mathrm{P} = \frac{1}{p}\mathrm{LL}'$, where p is the number of nodes in a gene tree.
272 This matrix is of dimension $l \times l$ and its entries can be interpreted as probabilities of observing
273 identical clade in two gene trees picked from posterior samples of different loci (off-diagonal
274 terms) or in two gene trees from the same posterior sample (diagonal terms). For convenience, P
275 can be rearranged so that its diagonal entries form a decreasing sequence. Then, one can consider
276 square submatrices of increasing size $k = 1 \ldots l$ (i.e. 2×2submatrix, 3×3 submatrix and so on
277 up to the full $l \times l$ matrix). The mean of diagonal elements in any such submatrix ($S = \sum_{i=1}^{k} P_{ii}$)
278 measures strength of phylogenetic signal in particular subset of k loci, while one minus the mean
279 of off-diagonal elements ($D = 1 - \frac{1}{(k^2-k)/2}\sum_{i=1}^{k-1}\sum_{j=i+1}^{k} P_{ij}$) measures its diversity. As the aim is
280 to maximize both strength and diversity of signal at once, the quantity considered is product of
281 strength and diversity, $Q = S * D$. The product Q increases steeply, albeit erratically, at the
282 beginning but just a little afterwards (not shown). The purpose of subsampling was to retain the
283 minimum number of loci whose phylogenetic signal is sufficiently strong and diverse to represent
284 the whole set of loci successfully sequenced in all species. To achieve this objective we retained
285 loci corresponding to the starting (overall increasing) part of the curve. We chose to approximate
286 behaviour of Q by a smooth curve (cubic regression spline) and take the maximum third order
287 difference between successive smoothed values as a breakpoint. The corresponding R script is
288 available in SM4. In each of the selected loci we compared strict and uncorrelated lognormal
289 relaxed clock (Drummond et al., 2006) using Bayes factors calculated in RevBayes v. 1.0.10 (Höhna
290 et al., 2016). The strict clock model was supported in 36 out of 39 loci. Given the exclusion of three
291 other loci left results virtually unchanged (not shown), we applied strict clock to all 39 loci, each

292 ~~of them having its own substitution rate. Nucleotide substitution model parameters were fixed to~~
293 ~~the averages of posterior samples obtained from the MrBayes analyses.~~
294
295 **3. Results**
296 *3.1 Summary of collected data sets*
297 The nuclear phylogenomic analysis was based on 377 successfully sequenced loci ranging in length
298 from 436 to 2,565 bp (median 1,644 bp). The total length of concatenated alignment for 40 taxa
299 was 581,030 bp. Some sequences were incomplete or missing and thus the data set contained
300 from 22 to 40 sequences for particular loci. Overall, 3.9% sequences and 5.6% bp were missing.
301 For the same 40 individuals we produced the unambiguous mtDNA alignments for 12 protein-
302 coding genes (10,891 bp), two rRNA-coding genes (2,419 bp) and 22 tRNA-coding genes (1,467
303 bp). These alignments were concatenated into final mitogenomic alignment of 14,777 bp,
304 equivalent to approximately 91% of the rodent mitochondrial genome.
305
306 *3.2 Phylogenetic reconstructions based on multilocus nuclear data and complete mtDNA*
307 The results of phylogenetic analyses are summarized in Fig. 1. After rooting by Praomyini, the
308 nuclear species tree shows *Millardia meltada* as the sister lineage to Otomyini+Arvicanthini and
309 hence the tribe Otomyini as the sister lineage of Arvicanthini. The only living Asian species of
310 Arvicanthini, *Golunda ellioti*, is in sister relationship to all African taxa. In Africa, the basal split is
311 between *Oenomys* and remaining genera, where we recognize four major clades, named here
312 *Hybomys*, *Aethomys*, *Dasymys* and *Arvicanthis* clades (Fig. 1, Table S1). The phylogeny is almost
313 fully resolved (PP=1.00), just the position of *Thallomys* has slightly lower support (PP=0.92). The
314 same topology was obtained from the ML analysis of concatenated loci and also the bootstrap

support (BS) was maximum for all nodes, except for *Thallomys*–*Thamnomys*/*Grammomys* node with BS=96~~2~~%.

On the contrary, the phylogenetic tree based on mitochondrial genomes shows low support for relationships between *Hybomys*, *Aethomys*, *Dasymys* and *Arvicanthis* clades, as well as within the *Aethomys* clade. The topology is generally similar to the nuclear tree, but differs in the following points: (i) *Golunda* and *Oenomys* form a strongly supported (PP=1.00) monophyletic clade rather than subsequent offshoots, (ii) *Rhabdomys* is supported (PP=0.94, BS=98%) as the sister of *Lamottemys* rather than of *Desmomys*, (iii) there are two differences in topology within genera *Arvicanthis* and *Lemniscomys*. For each of these conflicts we examined the number of MrBayes gene trees, whose topology was congruent with either nuclear or mitochondrial tree. The position of *Oenomys* on gene trees varied considerably. In equal share of 9% gene trees *Oenomys* was sister to "non-*Golunda* Arvicanthini" (nuclear topology), *Golunda* (mitochondrial topology) or to the rest of arvicanthini, but in smaller proportions of gene trees it was found sister to many different clades. *Rhabdomys* was sister to *Desmomys* (nuclear topology) in 40% of gene trees, while in 21% it was sister to *Lamottemys* (mitochondrial topology). Within *Arvicanthis* and *Lemniscomys*, the percentage was 41% to 26% and 61% to 18%, respectively, always in favour of the nuclear topology.

3.3 Diversification in the historical context

The split between the tribe Arvicanthini and its sister tribe Otomyini is dated to the Tortonian stage of Late Miocene (~~median estimate = 8~~9.~~8~~0 Ma) (Fig. 2). Still in the same stage, the Asian *Golunda* diverged from the ancestor of all African arvicanthines (7.9~~6~~ Ma). The most intensive African radiation, when a majority of modern genera appeared, ~~is dated to~~overlaps with the

Messinian stage of Late Miocene (the times of most recent ancestors, TMRCAs, clustered in the period between 7.~~0~~6 and 5.~~1~~3 Ma). The next intensive diversification period is dated to Lower-Pliocene (4.~~3~~7 - 3.~~4~~7 Ma) with intergeneric splits between *Stochomys*/*Dephomys*, *Grammomys poensis*/*Thamnomys*, *Lamottemys*/*Desmomys*/*Rhabdomys* and the oldest diversifications within the *Arvicanthis* clade. The first intrageneric divergences (in *Aethomys*) are dated in the same period. The oldest splits within other genera (*Typomys*, *Grammomys*) are dated to the ~~beginning of Pleistocene~~end of Pliocene, which overlaps with the divergence between the youngest genera of Arvicanthini, i.e. *Pelomys*/*Mylomys* and *Arvicanthis*/*Lemniscomys* (Fig. 2).

4. Discussion

4.1 Anchored phylogenomics vs. complete mtDNA

We demonstrated that complete mtDNA is less powerful and less reliable in resolving phylogenetic relationships than the anchored nuclear loci; this resulted in lack of resolution of some of the deep nodes (dated to 6.~~1~~3-7.~~0~~6 Ma) in the Bayesian analysis of complete mtDNA. This is probably due to higher substitution rates in mtDNA, which makes it largely saturated by mutations on larger timescales. Also, mtDNA was much shorter (14,777 bp compared to 581,030 bp) and its analysis cannot benefit from modelling of gene tree discordance and its potential to bring additional information about phylogenetic relationships at the species level. Finally, variation in mtDNA may be more affected by selection due to prevalence of coding sequences.

There are also some differences in topology of mtDNA and nuclear trees. The relationships inferred by anchored phylogenomics have higher credit here because mtDNA tree may differ from the species tree due to incomplete lineage sorting and, especially on shallow scales (e.g. within *Arvicanthis*, *Lemniscomys*), also due to mitochondrial introgression. Notably, the incongruent

363 nodes are usually poorly supported by at least one phylogenetic method at mitochondrial tree.
364 Gene trees were predominantly congruent with the nuclear species tree topology dominated in
365 all but one conflicting relationships. The exception was *Oenomys*, whose position in gene trees
366 was very variable and its confidential placement by ASTRAL was apparently driven by distribution
367 of quartet subtrees rather than by prevalence of fully congruent bipartitions in input unrooted
368 trees. The subsequent dating analysis estimated close to zero branch length separating *Oenomys*
369 from non-*Golunda* arvicanthines and so hard polytomy may be suspected here. Taken together,
370 ~~As a result~~ we will base the following discussion on the species tree obtained by anchored
371 phylogenomics.
372
373 *4.2 Evolutionary origin of Arvicanthini*
374 The tribe Arvicanthini (Denys et al., 2017) forms a strongly supported monophyletic group, sister
375 to Otomyini and more distantly to Millardiini. The monophyly of the tribe has been repeatedly
376 recognized in previous phylogenetic analyses based on few genetic segments of mitochondrial
377 and nuclear DNA (Ducroz et al., 2001; Steppan et al., 2004, 2005; Lavrenchenko and Verheyen,
378 2005; Lecompte et al., 2008; Rowe et al., 2008; Schenk et al., 2013; Missoup et al., 2016, 2018;
379 Steppan and Schenk, 2017; Aghová et al., 2018), although no multi-locus genetic study has
380 integrated all nominal genera of the tribe; even the most complete study of Missoup et al. (2018)
381 used only mtDNA for *Thamnomys*. Our phylogenomic analysis with complete sampling of all
382 genera not only confirms monophyly of Arvicanthini, but for the first time fully resolves
383 phylogenetic relationships among genera within the tribe, which now allows the reconstruction
384 of their (adaptive) radiation.
385

All recent African genera of Arvicanthini form a monophyletic group, sister of Indian *Golunda*. The basal position of *Golunda* in Arvicanthini was for the first time suggested by Lecompte et al. (2008), while in many other multi-locus phylogenies, *Golunda* formed a sister group of *Oenomys* (e.g. Steppan and Schenk, 2017; Aghová et al., 2018). The latter topology was likely affected by mtDNA variation, because sister relationship of the two genera was revealed also in our mitogenomic phylogeny (Fig. 1). The successive sisters of Arvicanthini (with one Indian and one African lineage) are African Otomyini and Indian Millardiini. It seems therefore reasonable to ask whether these rodents diversified in Asia and then twice colonized Africa by ancestors of Otomyini and African Arvicanthini or vice versa (i.e. diversification of the group in Africa and re-colonization of Asia by *Golunda* and the ancestor of Millardiini). The origin of murine rodents, subfamily Murinae, in Asia is well supported (e.g. Aghová et al., 2008; Schenk et al., 2013; and references therein). Lecompte et al. (2008) pointed out that multiple sister relationships between Asian and African clades (Praomyini - Murini; Malacomyini - Apodemini; Arvicanthini/Otomyini - Millardiini, respectively) suggest that each of the African lineages was differentiated prior to their dispersal into Africa. This should happened around the same time as a part of broader episode of faunal interchange (11-10 Ma), which is in a very good fit with the fossil record (see references in Lecompte et al., 2008, and review in Winkler et al., 2010). Applying this logic to the resolved phylogeny of Arvicanthini, we can speculate that even the ancestors of African Arvicanthini and Otomyini diverged already in Asia and arrived to Africa independently in Late Miocene (ca. 9-7 Ma), leaving *Golunda* (as the only surviving lineage of already differentiated Arvicanthini) in Asia.

Palaeontological records, however, do not refute an alternative explanation, making the discussion about African or Asian origin of Arvicanthini ~~slightly useless~~even more complex. During the Middle Miocene (16.0-11.6 Ma), the Mediterranean-Indo-Pacific seaway closed again at the

410 beginning of the Serravallian ca. 13.8 Ma ('Parathethys Salinity Crisis'; Rögl, 1999), and the newly
411 formed land bridge allowed repeated exchanges of terrestrial organisms. We can hypothesize that
412 ancestors of contemporary Arvicanthini occurred more or less continuously across this part of
413 Afro-Asia. Murine fossil records provide clear evidence for connections between the Indomalaya,
414 the Palearctic and the Afrotropics in this period (see references in Aghová et al., 2018). Among
415 them, the conspicuous examples are the oldest records of †*Parapelomys* spp., considered to
416 belong to Arvicanthini (Denys et al., 2017), found synchronously in Africa (8.5 Ma in Chorora,
417 Ethiopia; Geraads, 2001) and in Pakistan (ca. 8.0 Ma; Jacobs and Flynn, 2005). In agreement with
418 that, the Ethiopian as well as Moroccan sites of early Pliocene epoch reportedly contain fossils
419 identified as *Millardia* and *Golunda* (e.g. WoldeGabriel et al., 1994; Wynn et al., 2006), currently
420 limited to the Indian subcontinent. The hypothesized sister genus of *Golunda*, †*Saidomys*, was in
421 late Miocene also widely distributed in both northern Africa and Asia (Winkler, 2002; Patnaik,
422 2014). All these records suggest that we cannot simply define the place of diversification of
423 *Golunda* and other arvicanthines and the direction of colonization. *Golunda* (but similarly also
424 *Millardia*) may be just viewed as phylogenetic relict of a Miocene faunal interchange that have
425 disappeared from Africa, while all other ancient Arvicanthini (†*Parapelomys*, †*Saidomys*) went to
426 the extinction globally (Lecompte et al., 2008). Fossils already assignable to extant African
427 arvicanthine genera date from Late Miocene through Early Pliocene, around 7-5 Ma (e.g.
428 *Aethomys* or *Arvicanthis*, see review in Winkler et al., 2010 and Denys and Winkler 2015). This is
429 in good agreement with our molecular dating, which suggests that the first radiation of African
430 Arvicanthini occurred in the Messinian epoch of Late Miocene, when major lineages within this
431 clade have already diversified.
432
433 *4.3 Ancestral traits, mechanisms of (adaptive) radiation*

The tribe Arvicanthini, with its high number of species and diverse ecological adaptations and lifestyles represents the most successful ~~adaptive~~ radiation of rodents in the African continent. African Arvicanthini are monophyletic, suggesting that the ~~adaptive~~ radiation started from a single ancestor lineage. The first two offshoots are represented by the genus *Oenomys* and the strongly supported *Hybomys* clade. All species from these two clades inhabit Guineo-Congolian zone, suggesting that the radiation of the tribe in Africa started in a forest. African forests developed by the late Cretaceous, and during the Middle Miocene climatic optimum they extended coast to coast across the equatorial zone (Maley, 1996; Morley and Kingdon, 2013). It is therefore likely that murine newcomers from Asia in late Miocene first adapted (or already came adapted) to this most widespread ecosystem. Two other tribes of murine rodents that entered Africa also in Late Miocene ~~–~~are either restricted to Guineo-Congolian forests (Malacomyini) or have there the highest evolutionary diversity (Praomyini), which provides additional support for this hypothesis (Aghová et al., 2018; our unpubl. genomic data).

The most characteristic features of the climate in tropical Africa since the Late Miocene is its increasing variability and overall aridification (Ségalen et al., 2007; Potts, 2013). The shrinking of forests was linked to development of more open savanna-like ecosystems, evidenced e.g. by spread of C4 grasses. They appear in East Africa in the Mid-Late Miocene and between 8 and 6 Ma they already represent significant part of diet of grazing mammals (Cerling et al., 1997). Arvicanthine rodents were among the most successful mammals (together with large ungulates) that colonized these newly emerging ecological niches. It is especially true for the *Arvicanthis* clade, whose members often dominate small mammal assemblages in the non-forested sub-Saharan habitats. First, the genera *Arvicanthis* and *Lemniscomys* are closely related with numerous common morphological, ecological and behavioural traits (Ducroz et al., 2001). Their

458 TMRCA was estimated to 2.69 Ma, which is much more recent than previously thought (Aghová
459 et al., 2018). They are widespread and often very abundant in various savannahs, except South
460 Africa. The second subclade includes *Mylomys* and *Pelomys*, poorly known taxa preferring open
461 moist habitats in the forest-savanna mosaic, mostly along the equator (Wilson et al 2017). Finally,
462 the third subclade includes *Desmomys*, *Lamottemys* and *Rhabdomys* (in agreement with Ducroz
463 et al., 2001; Lavrenchenko and Verheyen, 2005; Lecompte et al., 2008; Missoup et al., 2016). The
464 composition of this cluster seems rather surprising at first because *Desmomys* is endemic to
465 Ethiopian highlands, while *Lamottemys* is endemic to Mount Oku in Cameroon and *Rhabdomys* is
466 widespread in various open habitats of Eastern and Southern Africa. However, they share multiple
467 morphological traits and all of them prefer relatively humid montane habitats (e.g. Ducroz et al.,
468 2001; Denys et al., 2014; Missoup et al., 2016). It is only south of the Limpopo river that
469 *Rhabdomys* becomes widespread in arid habitats, a fact that might be related to the absence of
470 competition from *Arvicanthis* and *Lemniscomys* (Ducroz et al., 2001). The distribution pattern in
471 this subclade reinforces the hypothesis of recurrent connections between western and eastern
472 African mountains in mid-Pliocene i.e. the period that was characterized by warm and wet climate
473 in Africa (Feakins and deMenocal, 2010). As a consequence, the moist montane environments
474 expanded and facilitated a trans-continental dispersal of their inhabitants (see Taylor et al. 2014
475 for another example in rodents). We showed for the first time (contrary to Missoup et al. 2016)
476 that *Lamottemys* diverged first in this subclade (ca. 3.74.0 Ma), which is understandable given the
477 geographical distance between eastern and western African mountains, and in agreement with
478 the pollen records indicating an abrupt change in forest cover ca. 3.3 Ma (Bonnefille et al., 2004).
479 The genera *Rhabdomys* and *Desmomys* are ecological vicariants in East African mountains (Wilson
480 et al., 2017) and they split soon after the divergence of *Lamottemys*. Recent phylogenetic studies
481 repeatedly suggested that Ethiopian highlands served as a source from which other eastern

482 African mountains were colonized through forest corridors that were predicted on both sides of
483 the Great Rift Valley during humid periods of Plio-Pleistocene (e.g. Bryja et al., 2014; Šumbera et
484 al., 2018; Krásová et al., 2019). It is therefore easy to imagine that the ancestor of *Desmomys* and
485 *Rhabdomys* inhabited a large north-south belt of East African montane grasslands and moorlands
486 and the two genera definitely diverged after the end of warm and wet mid-Pliocene period.
487
488 As already suggested by the tree topologies in recent multi-locus phylogenetic studies (e.g. Schenk
489 et al., 2013; Steppan and Schenk, 2017; Missoup et al., 2018; Aghová et al., 2018), *Dasymys* is the
490 sister to the *Arvicanthis* clade. The genus *Dasymys* has semi-aquatic habits and is specialized to
491 live in marshlands (Wilson et al 2017), which again supports the hypothesis that the origin of the
492 speciose *Arvicanthis* clade is in relatively moist open habitats, which are still occupied by some
493 *Rhabdomys*, *Pelomys*, *Mylomys* or *Desmomys* species. On the other hand, the adaptation to arid
494 (sometimes even semi-desert) environments in some species of *Arvicanthis* and *Lemniscomys* and
495 southern African *Rhabdomys* is probably a trait that evolved more recently (in Pleistocene) in
496 response to the increasing aridity (deMenocal, 1995) and allowed these groups to occupy large
497 areas in both South and North of equator forest blocks.
498
499 One of the most important results of this study is the first unequivocal resolution of phylogenetic
500 relationships of notoriously difficult taxa *Micaelamys*, *Aethomys*, *Thamnomys* and *Thallomys*.
501 Phylogenomic analysis clearly shows that all these genera form together a monophyletic clade
502 with *Grammomys* (named the *Aethomys* clade here, following "*Aethomys* division" of Musser and
503 Carleton (2005)). While two basal offshoots of this clade (i.e. *Micaelamys* and *Aethomys*) consist
504 of taxa with predominantly terrestrial activity in savannas, the three remaining genera form a
505 monophyletic group of at least partially arboricolous taxa (*Thallomys*, *Thamnomys*, *Grammomys*).

506 This suggests that the (pre-)adaptation to climb trees evolved only once in the common ancestor
507 of these genera, already in Late Miocene and probably in woodlands of south-eastern Africa,
508 where most species of *Thallomys* and *Grammomys* are found today and where is also the highest
509 diversity of *Aethomys*/*Micaelamys*. Based on the topology of the phylogenomic tree we can even
510 speculate that the ability to climb - to gain access to resources that are above the ground and to
511 protect themselves against predators - was advantageous for the secondary recolonization of
512 rainforests (shrub/tree floor) in Albertine Rift Mountains and part of Guineo-Congolian region by
513 the clade of (*Thamnomys* + *G. poensis* group) (sensu Bryja et al., 2017; see below for proposed
514 taxonomic changes).
515
516 *4.4 Taxonomic implications, delimitation of genera in Arvicanthini*
517 The resolved phylogeny of the tribe provides the opportunity to revise its generic classification.
518 There are neither rules nor generally accepted consensus about what the mammalian genus
519 should be (Dubois, 1988), contrary to numerous species concepts (e.g. Zachos, 2016), but at least
520 it should consist of a monophyletic group of species characterized by synapomorphic traits. There
521 are at least two cases, where this is not true in current taxonomy of Arvicanthini as reported in
522 the recent Handbook of the Mammals of the World (Wilson et al. 2017). First, Missoup et al.
523 (2018) recently performed phylogenetic analysis of one mitochondrial and two nuclear DNA
524 fragments and found that *Hybomys* (sensu Wilson et al., 2017) is a paraphyletic taxon. Our
525 phylogenomic analysis confirmed this finding and we already follow the generic classification
526 proposed by Missoup et al. (2018), i.e. the split of former *Hybomys* into *Hybomys* and *Typomys*.
527
528 The second clearly paraphyletic genus is *Grammomys* (Fig. 1), where *G*. *poensis* is a sister taxon
529 to *Thamnomys*, but not to the other *Grammomys* species. Many authors have asserted that the

530 species of *Thamnomys* and *Gramommys* are in the same monophyletic group and separable only
531 at the subgeneric level, while others consider them as two clearly distinct genera (see references
532 in Musser & Carleton 2005). The genus *Thamnomys* represents a poorly documented group of
533 species that are either rare or difficult to collect, and with very restricted distribution ranges in
534 rainforests of the Albertine Rift and eastern Congo Basin (Wilson et al., 2017). Because of
535 unavailability of samples, they were not included in phylogenetic studies until very recently. Bryja
536 et al. (2017) for the first time used *Thamnomys* sequences in their multi-locus study, but its
537 position remained unresolved. Even if their overall phylogeny of Arvicanthini was based on four
538 mitochondrial and five nuclear DNA markers, for *Thamnomys* only two mitochondrial fragments
539 were available (see similar results in Missoup et al., 2018). Here we unambiguously showed that
540 the *poensis* group (sensu Bryja et al., 2017) is much closer to *Thamnomys* than to *Grammomys*
541 (Fig. 1). This relationship is supported also by morphological traits on the skull and teeth (Hutterer
542 and Dieterlen, 1984) and we therefore propose to classify the *poensis* group (with two species
543 listed in the most recent compendia, *poensis* and *kuru*; e.g. Monadjem et al., 2015; Wilson et al.,
544 2017) as an internal lineage of *Thamnomys*. The *poensis* group and remaining *Thamnomys*
545 diverged ca. 4 Ma, which is comparable with the first intrageneric splits, e.g. between species of
546 *Aethomys* (Fig. 2). The genera *Thamnomys* (including *poensis* group) and *Grammomys* diverged
547 5.7~~1~~ Ma, i.e. well before all other intra-generic diversification events in Arvicanthini (Fig. 2). This
548 ancient divergence and the fact that these two taxa can be distinguished by several morphological
549 characters are strong arguments to consider *Thamnomys* and *Grammomys* as distinct genera and
550 not as subgenera. According to this findings, the two sister genera, *Thamnomys* (with most species
551 occupying Congo basin and Albertine Rift Mts.) and *Grammomys* (with the highest diversity in
552 montane and coastal forests of East Africa), are the descendants of lineages that became
553 separated by the split of Guineo-Congolian and Eastern-African forests during the

554 Miocene/Pliocene boundary (see more details in Bryja et al., 2017). This situation is analogous to
555 another widespread murine genus *Praomys* sensu lato (the tribe Praomyini), where most diversity
556 is currently found in tropical Guineo-Congolian forests. Especially the *P. jacksoni* complex shows
557 very similar phylogeographic structure to *Thamnomys* (in the new view, i.e. including the *poensis*
558 group), with very high diversity in Albertine rift Mts. and east-west structure of populations north
559 of the Congo River (Mizerovská et al., 2019). On the other hand, Eastern African montane forests
560 are inhabited by the so-called *P. delectorum* group (often found together with *Grammomys*
561 species) that diverged very early during the Praomyini radiation in Miocene/Pliocene boundary
562 and should be excluded from the genus *Praomys* (Missoup et al., 2012).

563

564 The genomic differences among currently recognized sister genera of Arvicanthini, reflected as
565 their divergence time (Fig. 2), are very variable. For example the evolutionary distances among
566 species within genera *Aethomys*, *Thamnomys* (including the *poensis* group) and *Grammomys*
567 (even after exclusion of the *poensis* group) are comparable with many intergeneric differences
568 (e.g. *Stochomys*/*Dephomys*, *Lamottemys*/*Desmomys*/*Rhabdomys*) or much higher
569 (*Mylomys*/*Pelomys*, or *Arvicanthis*/*Lemniscomys*). If we assume that most of loci used in this study
570 are selectively neutral, this implies that the extent of genomic differences (and divergence times)
571 is much higher within some genera than between some others. We do not advocate here the split
572 of genetically heterogeneous genera (or the lumping of genetically similar genera), but the
573 outputs of our phylogenomic analysis provide interesting hypotheses worthy of testing by future
574 integrative taxonomic work. They should become a matter for discussions in the mammalogical
575 community and might challenge the present generic classification of Arvicanthini.

576

577 **Acknowledgements**

578 This study was supported by the project of the Czech Science Foundation, no. 18-17398S, and
579 UMR7205 of the MNHN. We also thank the SYNTHESYS programme (BE-TAF-5113 to OM and
580 FRTAF-5799 to JB). For permission to carry out the research and to collect specimens we are
581 obliged to the Kenyan Forest Service and the Kenyan Wildlife Service, the Ethiopian Wildlife
582 Conservation Authority, the National Research Council and Forestry Department in Malawi,
583 Sokoine University of Agriculture in Morogoro (Tanzania), the National Directorate for Protected
584 Areas (DINAC – Mozambique), and Zambian Wildlife Authority. We would like to thank F. Catzeflis,
585 L. Granjon and J. Kerbis Peterhans for providing four tissue samples. For help in the field and with
586 logistics, we acknowledge S. Gryseels, Y. Meheretu, K. Welegerima, A. Ribas, P. Kaňuch, H.
587 Konvičková, F. Sedláček, R. Makundi, A. Massawe, G. Phamphi, J. Mbau, J. Šklíba, M. Lövy, and all
588 local collaborators. The authors thank M. Kortyna, S. Holland, K. Birch, and A. Bigelow for
589 assistance with generating the anchored phylogenomics data set. Access to the National Grid
590 Infrastructure MetaCentrum provided under the programme CESNET LM2015042 is greatly
591 appreciated and so is the access to CIPRES Science Gateway (https://www.phylo.org/). These
592 facilities were used to run most of the phylogenetic analyses.

593

594 **Author contributions**

595 J.B., R.S., A.K., V.N., C.D., and E.V. conceived the study, provided samples and funding, A.R.L. and
596 E.M.L. produced the anchored phylogenomics dataset, A.B. did part of the lab work and
597 assembled mitogenomes, O.M. and J.B. analysed data and drafted the manuscript. All authors
598 contributed to the editing of the manuscript, gave final approval for publication and agreed to be
599 held accountable for the work performed therein.

600

601 **Data availability**

Complete mitochondrial genomes are available in SM3 and in GenBank under accession numbers MN807579-MN807618 (see Table S1 in SM1). Alignments of nuclear loci obtained by anchored phylogenomic approach (as partitioned nexus file) and the Bayesian gene trees used as input for ASTRAL analysis (in newick format) ~~-~~ are available in ~~GitHub~~ the public repository of the Czech Academy of Sciences (http://hdl.handle.net/11104/0312390~~https://github.com/onmikula/genetree_subset~~).

Appendix A. Supplementary material

Supplementary Materials SM1 (Table S1): List of used specimens.

Supplementary Materials SM2: The template of MrBayes block specifying the inference of gene trees in the *nexus* file.

Supplementary Materials SM3: The mitogenomic dataset, and MrBayes block specifying the best partition scheme for mtDNA and substitution models in the *nexus* file.

~~*Supplementary Materials SM4*: The R script performing the selection of anchored loci for the divergence dating analysis.~~

References

Aghová, T., Kimura, Y., Bryja, J., Dobigny, G., Granjon, L., Kergoat, G.J., 2018. Fossils know it best: using a new set of fossil calibrations to improve the temporal phylogenetic framework of murid rodents (Rodentia: Muridae). Mol. Phylogenet. Evol. 128, 98–111.

625 Amaral, C.R.L., Pereira, F., Silva, D.A., Amorim, A., de Carvalho, E.F., 2018. The mitogenomic
626 phylogeny of the Elasmobranchii (Chondrichthyes). Mitochondrial DNA Part A 29, 867–878.
627 Bonnefille, R., Potts, R., Chalie, F., Jolly, D., Peyron, O., 2004. High-resolution vegetation and
628 climate change associated with Pliocene *Australopithecus afarensis*. PNAS 101, 12125–12129.
629 Bryja, J., Šumbera, R., Kerbis Peterhans, J.C., Aghová, T., Bryjová, A., Mikula, O., Nicolas, V., Denys,
630 C., Verheyen, E., 2017. Evolutionary history of the thicket rats (genus *Grammomys*) mirrors the
631 evolution of African forests since late Miocene. J Biogeogr. 44, 182–194.
632 Bryja, J., Mikula, O., Šmbera, R., Meheretu, Y., Aghová, T., Lavrenchenko, L.A., Mazoch, V., Oguge,
633 N., Mbau, J.S., Welegerima, K., Amundala, N., Colyn, M., Leirs, H., Verheyen, E., 2014. Pan-
634 African phylogeny of *Mus* (subgenus *Nannomys*) reveals one of the most successful mammal
635 radiations in Africa. BMC Evol. Biol. 14, 256.
636 Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in
637 phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
638 Cerling, T.E., Harris, J.M., MacFadden, B.J., Leakey, M.G., Quade, J., Eisenmann, W., Ehleringer,
639 J.R., 1997. Global vegetation change through the Miocene/Pliocene boundary. Nature 389,
640 153–158.
641 Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the
642 multispecies coalescent. Trends Ecol. Evol. 24, 332–340.
643 deMenocal, P.B., 1995. Plio-Pleistocene African climate. Science 270, 53–59.
644 Denys, C., Missoup, A.D., Nicolas, V., Fülling, O., Delapré, A., Bilong Bilong, C.F., Taylor, P.J.J.,
645 Hutterer R., 2014. African highlands as mammal diversity hotspots: new records of
646 *Lamottemys okuensis* (Rodentia: Muridae) and other endemic rodents from Mount Oku,
647 Cameroon. Zoosystema 36, 647–690.

648 Denys, C., Lecompte, E., Dubois, A., Taylor, P.J., 2017. Diagnoses and contents of new African and
649 Eurasian Murinae (Rodentia, Muridae) tribes. Dumerilia 7, 82–90.
650 Denys, C., Winkler, A., 2015. Advances in integrative taxonomy and evolution of African murid
651 rodent: how morphological trees hide the molecular forest. In Cox, P., Hautier, L. (eds.):
652 Evolution of the Rodents: Advances in Phylogeny, Palaeontology and Functional Morphology.
653 Cambridge studies in morphology and molecules. New paradigms in Evolutionary Biology,
654 Cambridge University Press, 186–220.
655 Drummond, A.J., Bouckaert, R.R., 2015. Bayesian evolutionary analysis with BEAST. Cambridge
656 University Press.
657 Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating
658 with confidence. PLoS Biology 4, e88.
659 Dubois, A. 1988. The genus in zoology: A contribution to the theory of evolutionary systematics.
660 Mém. Mus. Natn. Hist. Nat. (A) 140, 1–124.
661 Ducroz, J.F., Volobouev, V., Granjon, L., 2001. An assessment of the systematics of arvicanthine
662 rodents using mitochondrial DNA sequences: evolutionary and biogeographical implications.
663 J. Mammal. Evol. 8, 173–206.
664 Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
665 Nucleic Acids Research 32, 1792–1797.
666 Feakins, S.J., deMenocal, P.B., 2010. Global and African regional climate during the Cenozoic. In
667 Werdelin, L., Sanders, W.J. (eds): Cenozoic Mammals of Africa. Berkeley, University of
668 California Press, 45–55.
669 Geraads, D., 2001. Rongeurs du Miocene superieur de Chorora (Ethiopie): Dendromuridae,
670 Muridae et conclusions. Palaeovertebrata 30, 89–109.
671 Gernhard, T., 2008. The conditioned reconstructed process. J. Theor. Biol. 253, 769–778.

672 Gissi, C., Reyes, A., Pesole, G., Saccone, C., 2000. Lineage-specific evolutionary rate in mammalian
673 mtDNA. Mol. Biol. Evol. 17, 1022–1031.
674 Hamilton, C.A., Lemmon, A.R., Lemmon, E.M., Bond, J.E., 2016. Expanding Anchored Hybrid
675 Enrichment to resolve both deep and shallow relationships within the spider Tree of Life. BMC
676 Evol. Biol. 16, 212.
677 Höhna, S., Landis, M.J., Heath, T.A., Boussau, B., Lartillot, N., Moore, B.R., Huelsenbeck, J.P.,
678 Ronquist, F., 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an
679 interactive model-specification language. Syst. Biol. 65, 726–736.
680 Huelsenbeck, J.P., Larget, B., Alfaro, M.E., 2004. Bayesian phylogenetic model selection using
681 reversible jump Markov chain Monte Carlo. Mol. Biol. Evol. 21, 1123–1133.
682 Hutterer, R., Dieterlen, F. 1984. Zwei neue Arten der Gattung *Grammomys* aus Äthiopien und
683 Kenia (Mammalia: Muridae). Stuttgarter Beiträge zur Naturkunde Serie A (Biologie) 374, 1–18.
684 Jacobs, L.L., Flynn, J.J., 2005. Of Mice…again: the Siwalik rodent record, murine distribution, and
685 molecular clocks. In Lieberman, D.E., Smith, R.J., Keller, J. (eds.): Interpreting the Past Essays
686 on Human, Primate, and Mammal Evolution in Honor of David Pilbeam. Brill Academic
687 Publishers, 63–80.
688 Katoh, K., Toh, H., 2008. Improved accuracy of multiple ncRNA alignment by incorporating
689 structural information into a MAFFT-based framework. BMC Bioinformatics 9, 212
690 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
691 improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
692 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A.,
693 Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious
694 Basic: An integrated and extendable desktop software platform for the organization and
695 analysis of sequence data. Bioinformatics 28, 1647–1649.

696 Kimura, Y., Hawkins, M.T.R., McDonough, M.M., Jacobs, L.L., Flynn, L.J., 2015. Corrected
697 placement of *Mus–Rattus* fossil calibration forces precision in the molecular tree of rodents.
698 Sci. Rep. 5, 14444.
699 Krásová, J., Mikula, O., Mazoch, V., Bryja, J., Říčan, O., Šumbera, R., 2019. Evolution of the Grey-
700 bellied Pygmy Mouse group: highly structured molecular diversity with predictable geographic
701 ranges but morphological crypsis. Mol. Phylogenet. Evol. 130, 143–155.
702 Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: new
703 methods for selecting partitioned models of evolution for molecular and morphological
704 phylogenetic analyses. Mol. Biol. Evol. 34, 772–773.
705 Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets.
706 Bioinformatics 30, 3276–3278.
707 Lavrenchenko, L.A., Verheyen, E., 2005. An assessment of the systematics of the genus *Desmomys*
708 Thomas, 1910 (Rodentia: Muridae) using mitochondrial DNA sequences. In Huber, B.A.,
709 Sinclair, B.J., Lampe, K.H. (eds.): African Biodiversity: Molecules, Organisms, Ecosystems.
710 Springer, New York, 363–369.
711 Lecompte, E., Aplin, K., Denys, C., Catzeflis, F., Chades, M., Chevret, P., 2008. Phylogeny and
712 biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a
713 new tribal classification of the subfamily. BMC Evol. Biol. 8, 199.
714 Lemmon, A.R., Emme, S., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-
715 throughput phylogenetics. Syst. Biol. 61, 721–744.
716 Lemmon, E.M., Lemmon A.R., 2013. High-Throughput genomic data in systematics and
717 phylogenetics. Ann. Rev. Ecol. Evol. Syst. 44, 99–121.
718 Maley, J., 1996. The African rain forest – main characteristics of changes in vegetation and climate
719 from the Upper Cretaceous to the Quaternary. Procs. R. Soc. Edinburgh 104B, 31–73.

720 Manthi, F.K., 2007. A preliminary review of the rodent fauna from Lemudong'o, south-western
721 Kenya, and its implication to the late Miocene paleoenvironments. Kirtlandia 56, 92–105.
722 McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C., 2012.
723 Ultraconserved elements are novel phylogenomic markers that resolve placental mammal
724 phylogeny when combined with species tree analysis. Genome Res. 22, 746–754.
725 McCormack, J.E., Tsai, W.L., Faircloth, B.C., 2016. Sequence capture of ultraconserved elements
726 from bird museum specimens. Mol. Ecol. Resour. 16, 1189–1203.
727 Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many
728 hundreds of taxa and thousands of genes. Bioinformatics 31, i44–i52.
729 Missoup, A.D., Nicolas, V., Wendelen, W., Keming, E., Bilong Bilong, C.F., Couloux, A., Atanga, E.,
730 Hutterer, R., Denys, C., 2012. Systematics and diversification of *Praomys* species (Rodentia:
731 Muridae) endemic to the Cameroon Volcanic Line (West Central Africa). Zool. Scri. 41, 327–
732 345.
733 Missoup, A.D., Nicolas, V., Eiseb, S., Chung, E.K., Denys, C., 2016. Phylogenetic position of the
734 endemic Mount Oku rat, *Lamottemys okuensis* (Rodentia: Muridae), based on molecular and
735 morphological data. Zool. J. Lin. Soc. 177, 209–226.
736 Missoup, A.D., Yemchui, G.D., Denys, C., Nicolas, V., 2018. Molecular phylogenetic analyses
737 indicate paraphyly of the genus *Hybomys* (Rodentia: Muridae): Taxonomic implications. J. Zool.
738 Syst. Evol. Res. 56, 444–452.
739 Mizerovská, D., Nicolas, V., Demos, T.C., Akaibe, D., Colyn, M., Denys, C., Kaleme, P., Katuala, P.,
740 Kennis, J., Kerbis Peterhans, J.C., Laudisoit, A., Missoup, A.D., Šumbera, R., Verheyen, E., Bryja,
741 J., 2019. Genetic variation of the most abundant forest-dwelling rodents in Central Africa
742 (*Praomys jacksoni* complex): Evidence for Pleistocene refugia in both montane and lowland
743 forests. J Biogeogr. 46, 1466–1478.

744 Monadjem, A., Taylor, P.J., Denys, C., Cotterill, F.P.D., 2015. Rodents of Sub-Saharan Africa. A
745 biogeographic and taxonomic synthesis. Walter de Gruyter GmbH, Berlin/Munich/Boston.
746 Morley, R.J., Kingdon, J. (eds.), 2013. Mammals of Africa. Volume I: Introductory Chapters and
747 Afrotheria. Bloomsbury.
748 Musser, G.G., Carleton, M.D., 2005. Superfamily Muroidea. In Wilson, D.E., Reeder, D.A.M. (eds.):
749 Mammal Species of the World. A Taxonomic and Geographic Reference. The Johns Hopkins
750 University Press, Baltimore, 894–1531.
751 Ogilvie, H.A., Bouckaert, R.R., Drummond, A.J., 2017. StarBEAST2 brings faster species tree
752 inference and accurate estimates of substitution rates. Mol. Biol. Evol. 34, 2101–2114.
753 Paradis, E., Schliep, K., 2018. ape 5.0: an environment for modern phylogenetics and evolutionary
754 analyses in R. Bioinformatics 35, 526–528.
755 Patnaik R., 2014. Phylogeny of Siwalik murine rodents: Implications for *Mus-Rattus* divergence
756 time. J. Paleontol. Soc. India 59, 15–28.
757 Potts R., 2013. Hominin evolution in settings of strong environmental variability. Quat. Sci. Rev.
758 73, 1–13.
759 Pozzi, L., Hodgson, J.A., Burrell, A.S., Sterner, K.N., Raaum, R.L., Disotell, T.R., 2014. Primate
760 phylogenetic relationships and divergence dates inferred from complete mitochondrial
761 genomes. Mol. Phylogenet. Evol. 75, 165–183.
762 Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R.,
763 2015. A fully resolved, comprehensive phylogeny of birds (Aves) using targeted next
764 generation DNA sequencing. Nature 526, 569–573.
765 R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for
766 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
767 Rambaut, A., Drummond, A.J., 2007. Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer

768 Rögl, F., 1999. Mediterranean and Paratethys. Facts and hypotheses of an Oligocene to Miocene
769 paleogeography (short overview). Geol. Carpathica 50, 339–349.
770 Rokyta, D.R., Lemmon, A.R., Margres, M.J., Arnow, K., 2012. The venom-gland transcriptome of
771 the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13, 312.
772 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L.,
773 Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference
774 and model choice across a large model space. Syst. Biol. 61, 539–542.
775 Rowe, K.C., Reno, M.L., Richmond, D.M., Adkins, R.M., Steppan, S.J., 2008. Pliocene colonization
776 and adaptive radiations in Australia and New Guinea (Sahul): Multilocus systematics of the old
777 endemic rodents (Muroidea: Murinae). Mol. Phylogenet. Evol. 47, 84–101.
778 Ruane, S., Raxworthy, C.J., Lemmon, A.R., Lemmon, E.M., Burbrink, F.T., 2015. Comparing large
779 anchored phylogenomic and small molecular datasets for species tree estimation: An empirical
780 example using Malagasy pseudoxyrhophiine snakes. BMC Evol. Biol. 15, 221.
781 Schenk, J.J., Rowe, K.C., Steppan, S.J., 2013. Ecological opportunity and incumbency in the
782 diversification of repeated continental colonizations by muroid rodents. Syst. Biol. 62, 837–
783 864.
784 Schliep, K.P., 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27, 592–593.
785 Ségalen, L., Lee-Thorp, J., Cerling, T., 2007. Timing of C4 grass expansion across sub-Saharan
786 Africa. J. Hum. Evol. 53, 549–559.
787 Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML Web
788 servers. Syst. Biol. 57, 758–771.
789 Stamatakis, A., 2014. RAxML version 8 a tool for phylogenetic analysis and post-analysis of large
790 phylogenies. Bioinformatics 30, 1312–1313.

791 Steppan, S., Adkins, R., Anderson, J., 2004. Phylogeny and divergence-date estimates of rapid
792 radiations in muroid rodents based on multiple nuclear genes. Syst. Biol. 53, 533–553.
793 Steppan, S.J., Adkins, R.M., Spinks, P.Q., Hale, C., 2005. Multigene phylogeny of the Old World
794 mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial
795 genes compared to nuclear genes. Mol. Phylogenet. Evol. 37, 370–388.
796 Steppan, S.J., Schenk, J.J., 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing
797 diversification rates. PLoS One 12, e0183070.
798 Šumbera, R., Krásová, J., Lavrenchenko, L.A., Mengistu, S., Bekele, A., Mikula, O., Bryja, J., 2018.
799 Ethiopian highlands as a cradle of the African fossorial root-rats (genus *Tachyoryctes*), the
800 genetic evidence. Mol. Phylogenet. Evol. 126, 105–115.
801 Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and
802 ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56, 564–577.
803 Taylor, P.J., Maree, S., Cotterill, F.P.D., Missoup, A.D., Nicolas, V., Denys, C., 2014. Molecular and
804 morphological evidence for a Pleistocene radiation of laminate-toothed rats (*Otomys*:
805 Rodentia) across a volcanic archipelago in equatorial Africa. Biol. J. Linn. Soc. 113, 320–344.
806 Tucker, D.B., Colli, G.R., Giugliano, L.G., Hedges, S.B., Hendry, C.R., Lemmon, E.M., Lemmon, A.R.,
807 Sites, J.W.Jr., Pyron, R.A., 2016. Phylogenomic analysis of tegus and whiptails (Teiidae:
808 Squamata), with a revised taxonomy and a new genus from the West Indies. Mol. Phylogenet.
809 Evol. 103, 75–84.
810 Wilson, D.E., Lacher, T.E.Jr., Mittermeier, R.A. (eds), 2017. Handbook of the Mammals of the
811 World, vol. 7. Rodents II. Lynx Editions, Barcelona.
812 Winkler, A.J., 2002. Neogene paleobiogeography and East African paleoenvironments:
813 contributions from the Tugen Hills rodents and lagomorphs. J. Hum. Evol. 42, 237–256.

814 Winkler, A.J., Denys, C., Avery, M.D., 2010. Rodentia. In Werdelin, L., Sanders, W.J. (eds): Cenozoic
815 Mammals of Africa. Berkeley, University of California Press, 263–304.
816 WoldeGabriel, G., White, T.D., Suwa, G., Renne, P., de Heinzelin, J., Hart, W.K., Heiken, G., 1994.
817 Ecological and temporal placement of early Pliocene hominids at Aramis, Ethiopia. Nature 371,
818 330–333.
819 Wynn, J.G., Alemseged, Z., Bobe, R., Geraads, D., Reed, D., Roman, D.C., 2006. Geological and
820 palaeontological context of a Pliocene juvenile hominin at Dikika, Ethiopia. Nature 443, 332–
821 336.
822 Zachos, F., 2016. Species Concepts in Biology. Historical Development, Theoretical Foundations
823 and Practical Relevance. Springer International Publishing, Switzerland.
824

825 **Figure legends**

826 **Figure 1** Species tree based on 377 nuclear loci (from ASTRAL; in total 581,030 bp) and Bayesian
827 estimate of mitochondrial phylogeny (from MrBayes; 14,777 bp). Nodes are coloured according
828 to posterior probabilities from ASTRAL and MrBayes (squares) and bootstrap support from the
829 maximum likelihood (RAxML) analyses (circles). The colours distinguish categories of statistical
830 support.

831

832 **Figure 2** Divergence dating of the species tree inferred using a multi-species coalescent approach
833 in StarBEAST2. The analysis were based on ~~39~~ 231 loci from the anchored phylogenomic dataset
834 and the molecular clock was calibrated by two fossil constraints (the root and the MRCA of
835 *Aethomys* and *Arvicanthis*). The numbers in circles are TMRCAs of particular clades in million years
836 ago (Ma).

837

Figure1

377 nuclear loci

Complete mitochondrial DNA

Stenocephalemys albocaudata
Praomys delectorum
Praomyini

RAxML
<50
50–69
70–79
80–100
ASTRAL
<0.50
0.50–0.89
0.90–0.94
0.95–1.00

Lemniscomys bellieri
Lemniscomys striatus
Lemniscomys rosalia
Lemniscomys barbarus
Arvicanthis somalicus
Arvicanthis abyssinicus
Arvicanthis niloticus
Arvicanthis nairobae
Arvicanthis rufinus
Mylomys dybowskii
Pelomys fallax
Lamottemys okuensis
Rhabdomys dilectus
Desmomys harringtoni
Dasymys incomtus
Aethomys hindei
Aethomys kaiseri
Aethomys nyikae
Aethomys chrysophilus
Thamnomys kempi
Grammomys poensis
Grammomys macmillani
Grammomys surdaster
Grammomys dolichurus
Grammomys selousi
Thallomys paedulcus
Micaelamys namaquensis
Hybomys sp.n.
Hybomys univittatus
Dephomys defua
Stochomys longicaudatus
Typomys planifrons
Typomys trivirgatus
Oenomys hypoxanthus
Golunda ellioti
Otomys Mt.Kenya
Otomys typus
Otomyinii
Millardia meltada
Millardiini
Arvicanthini
4 coalescent units

Arvicanthis clade
Dasymys clade
Aethomys clade
Hybomys clade
Oenomys clade
Golunda clade

Stenocephalemys albocaudata
Praomys delectorum
Lemniscomys striatus
Lemniscomys bellieri
Lemniscomys rosalia
Lemniscomys barbarus
Arvicanthis somalicus
Arvicanthis abyssinicus
Arvicanthis niloticus
Arvicanthis nairobae
Arvicanthis rufinus
Mylomys dybowskii
Pelomys fallax
Lamottemys okuensis
Rhabdomys dilectus
Desmomys harringtoni
Dasymys incomtus
Aethomys hindei
Aethomys kaiseri
Aethomys nyikae
Aethomys chrysophilus
Thamnomys kempi
Grammomys poensis
Grammomys macmillani
Grammomys surdaster
Grammomys dolichurus
Grammomys selousi
Thallomys paedulcus
Micaelamys namaquensis
Hybomys sp.n.
Hybomys univittatus
Dephomys defua
Stochomys longicaudatus
Typomys planifrons
Typomys trivirgatus
Oenomys hypoxanthus
Golunda ellioti
Otomys Mt.Kenya
Otomys typus
Millardia meltada
Arvicanthini
0.5 substitutions / bp

RAxML
<50
50–69
70–79
80–100
<0.50
0.50–0.89
0.90–0.94
0.95–1.00
MrBayes

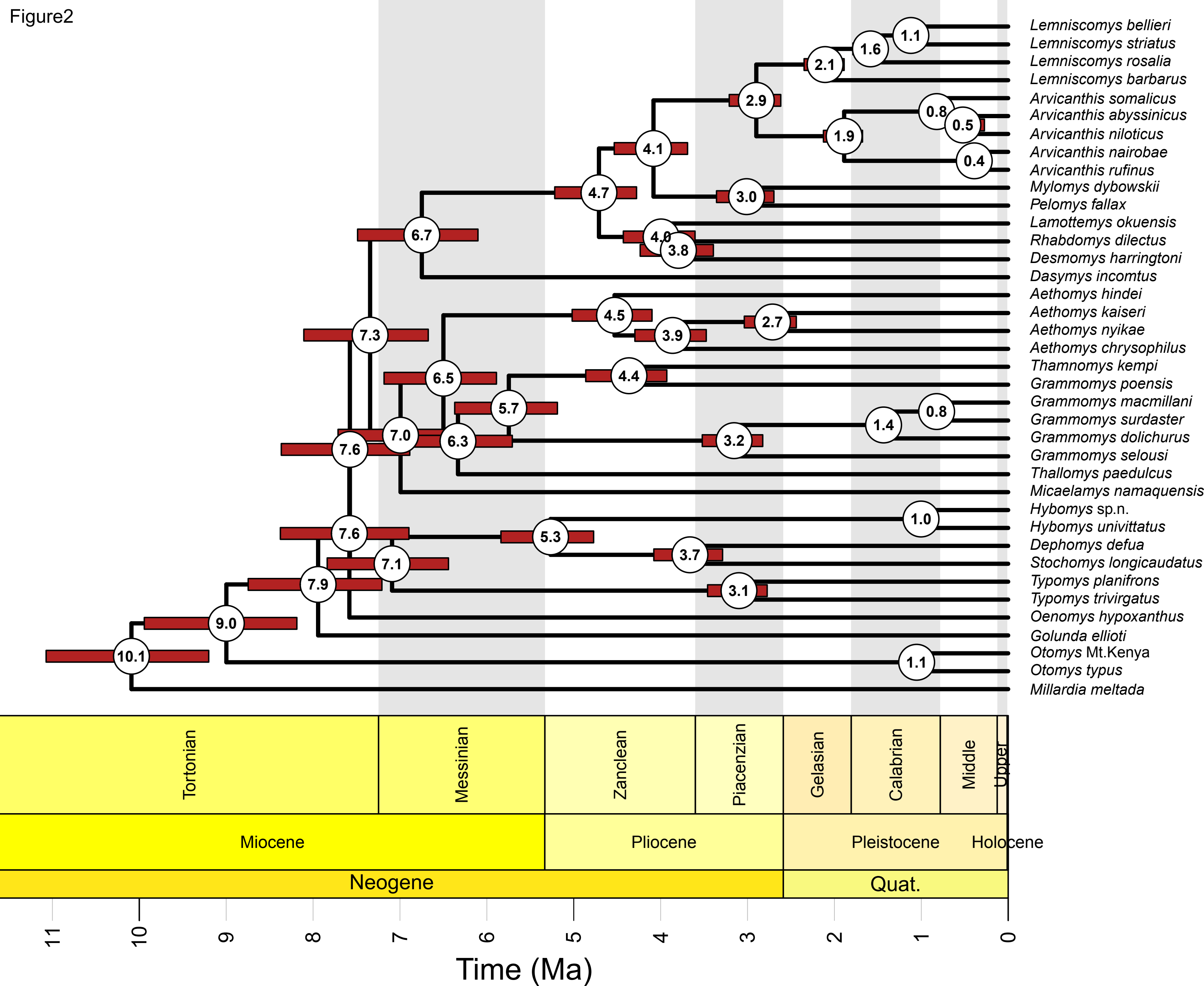


Figure2
Lemniscomys bellieri
Lemniscomys striatus
Lemniscomys rosalia
Lemniscomys barbarus
Arvicanthis somalicus
Arvicanthis abyssinicus
Arvicanthis niloticus
Arvicanthis nairobae
Arvicanthis rufinus
Mylomys dybowskii
Pelomys fallax
Lamottemys okuensis
Rhabdomys dilectus
Desmomys harringtoni
Dasymys incomtus
Aethomys hindei
Aethomys kaiseri
Aethomys nyikae
Aethomys chrysophilus
Thamnomys kempi
Grammomys poensis
Grammomys macmillani
Grammomys surdaster
Grammomys dolichurus
Grammomys selousi
Thallomys paedulcus
Micaelamys namaquensis
Hybomys sp.n.
Hybomys univittatus
Dephomys defua
Stochomys longicaudatus
Typomys planifrons
Typomys trivirgatus
Oenomys hypoxanthus
Golunda ellioti
Otomys Mt.Kenya
Otomys typus
Millardia meltada
Tortonian
Messinian
Zanclean
Piacenzian
Gelasian
Calabrian
Middle
Upper
Miocene
Pliocene
Pleistocene
Holocene
Neogene
Quat.
Time (Ma)

Authors´ Statement

None of the material in this manuscript has been published or is under consideration for publication elsewhere. All data used in the paper are made available. The manuscript has been approved by all the co-authors who agreed to its submission.

Thank you for considering this manuscript, we will look forward to hearing from you.

Yours Sincerely,

Josef Bryja (on behalf of all co-authors)

Click here to access/download
Supplementary Material
Mikula_Arvicanthini_SM1_Table S1.docx

Click here to access/download
Supplementary Material
Mikula_Arvicanthini_SM2_MrBayes block for ASTRAL.nex

Supplementary Material SM3

Click here to access/download

Supplementary Material

Mikula_Arvicanthini_SM3_partitioned complete mtDNA.nex