Received: 5 August 2021 | Revised: 22 September 2021 | Accepted: 10 November 2021

DOI: 10.1111/1755-0998.13555

RESOURCE ARTICLE

MOLECULAR ECOLOGY RESOURCES WILEY

A rodent anchored hybrid enrichment probe set for a range of phylogenetic utility: From order to species

Max R. Bangs | Scott J. Steppan

Department of Biological Sciences, Florida State University, Tallahassee, Florida, USA

Correspondence
Max R. Bangs, Department of Biological Sciences, Florida State University, Tallahassee, FL, USA.
Email: mbangs@bio.fsu.edu

Funding information
National Science Foundation, Grant/Award Number: DEB-1754748

Abstract

Rodents are the largest order of mammals and contain several model organisms important to scientific research in a variety of fields, yet no large set of genomic markers have been designed for this group to date, hindering evolutionary studies into relationships of the group as a whole. Here we present a genomic probe set designed and optimized for rodents with a protocol that is easy to replicate with little laboratory investment. This design utilizes an anchored hybrid enrichment approach specifically targeting rodents to generate longer loci with a higher substitution rate than existing vertebrate probes to provide utility at various taxonomic levels. Using a test set of rodents from all five suborders, we successfully obtained alignments for 416 of the 418 target loci with an average of 1379 bp per locus and a total alignment of more than half a million base pairs. This genomic data set performed well in all phylogenetic analyses, especially in recent phylogenetic splits, with ample parsimony-informative sites within genera and even within species, showing more than four times as many single nucleotide polymorphisms per locus than a recent vertebrate ultraconserved elements study. Additional support is provided in resolving deeper clades in Rodentia. By providing this probe design, we hope that more laboratories can easily generate data for answering questions in rodents from species delimitation to understanding relationships among families in rapid radiations.

KEYWORDS
anchored hybrid enrichment, phylogenetics, phylogenomics, Rodentia, target enrichment

1 | INTRODUCTION

Rodents constitute the largest order of mammals, make up more than 40% of all mammal species, are found all over the world, and have adapted to diverse habitats from lush forest to dry deserts and even the highest elevations of any animal. They also include many of the most important model organisms among vertebrates for biomedical research. Our understanding of evolutionary relationships has improved greatly with numerous multilocus studies over the last 20 years or more (e.gFabre et al., 2012; Meredith et al., 2011; Montgelard et al., 2008; Steppan & Schenk, 2017; Swanson et al., 2019). Resolving those relationships is critical whether conducting comparative biomedical studies or using their great diversity to understand evolution. Despite recent phylogenetic efforts, a few regions of the rodent tree remain uncertainly resolved and larger data sets are needed to achieve the desired resolution. However, modern phylogenomic approaches have only had limited application (Swanson et al., 2019) despite the recognition of their value (D'Elía et al., 2019; Lessa et al., 2014). Because of the increasing number of nonmodel species being included in comparative studies, there is also a need to resolve currently uncertain nodes that exist at a variety of levels, ranging from deep splits among suborders to those among or within species.

Here we report on a 418-locus, anchored hybrid enrichment (AHE) probe set designed specifically for rodents and to be informative across all phylogenetic scales of the order. First, we describe the probe set and then evaluate its informativeness at two different scales, from the root ~69 million years ago (Swanson et al., 2019) to the intraspecific level less than 1.0 million years ago (*Peromyscus leucopus*; Steppan & Schenk, 2017). Finally, we focus on two of the most debated nodes: the root node between Ctenohystrica, Sciuromorpha and the mouse-related clade "MRC" (Huchon et al., 2007; Figure 1, trees A–C), and the relationship among the suborders of the MRC (Myomorpha, Anomaluromorpha and Castorimorpha; Figure 1, trees 1–3).

2 | METHODS

2.1 | Testing previous anchored hybrid enrichment probe set

Before designing new probes, the *Vertebrate 512 Loci* probe set (Lemmon et al., 2012) was tested on seven *Apomys* (Philippine forest mice; Murinae) samples generated by the Center for Anchored Hybrid Enrichment at Florida State University following the protocols for library preparation and data filtering of Lemmon et al. (2012). All probes that generated sequences for at least one sample or from the *Mus musculus* (house mouse; Murinae) sample used in the initial testing of the *Vertebrate 512 Loci* probe set (Lemmon et al., 2012) were selected. From this list all probes were blasted to the *Mus musculus* genome (mm10) and if any probe had more than one BLAST hit with an e-score >0.0001 the locus was dropped. All remaining probes were retained.

2.2 | Rodent-specific probe design

Additional probes were designed by referencing six rodent genomes: *Mus musculus* (Myomorpha/Muridae; mm10), *Rattus norvegicus* (Norway rat; Myomorpha/Muridae; rn5), *Heterocephalus glaber* (naked mole-rat; Hystricomorpha/Heterocephalidae; hetGla2), *Cavia porcellus* (guinea pig; Hystricomorpha/Caviidae; cavPor3), *Dipodomys ordii* (Ord's kangaroo rat; Castorimorpha/Heteromyidae; dipOrd1) and *Ictidomys tridecemlineatus* (thirteen-lined ground squirrel; Sciuromorpha/Sciuridae; speTri2). These six genomes span four of the five rodent suborders and were included in the UCSC Genomics Institute mouse 60-way genome alignment, which was subsampled using the MAFSPECIESSUBSET tool to generate a whole genome alignment for these six rodent genomes.

Probes were designed using this multigenome alignment as input for MRBAIT (Chafin et al., 2018) using the following parameters: (i) the locus had to be present in all six species, (ii) no gaps allowed, (iii) no low-complexity regions, (iv) minimum length of 200 bp, (v) 120 bp tiling probes, and (vi) fewer than 12 single nucleotiode polymorphisms (SNPs) per probe. All loci passing filtering were then used for probe design by designing 21 tiling probes for each of five species (excluding *Rattus norvegicus* to avoid using two species in the same family) resulting in a total of 105 probes per locus with 110- to 116 -bp overlap between tiled 120-bp probes. All probes were then blasted to the *Mus musculus* genome (mm10) and if any probe had more than one BLAST hit with an e-score >0.0001 the locus was dropped. Probes were also designed for two loci commonly used in rodent phylogenetics (*IRBP* and *RAG1*) as above targeting the most conserved regions in both genes.

2.3 | Sample selection and library generation

In order to test the probes, libraries were generated using 12 samples that span all five rodent suborders including one from Hystricomorpha (*Cavia porcellus*), three from Sciuromorpha (*Graphiurus murinus* [woodland dormouse], *Sciurus stramineus* [Guayaquil squirrel] and *Marmota olympus* [Olympic marmot]), one from Anomaluromorpha (*Pedetes capensis* [springhare]), one from Castorimorpha (*Castor canadensis* [beaver]) and six from Myomorpha (*Allactaga sibirica* [Siberian jerboa], *Calomyscus* sp. [mouse-like hamster], *Phyllotis xanthopygus* [yellow-rumped leaf-eared mouse], *Peromyscus leucopus* [white-footed mouse], *Apomys microdon* [small Luzon forest mouse] and *Mus cookii* [Cook's mouse]). With

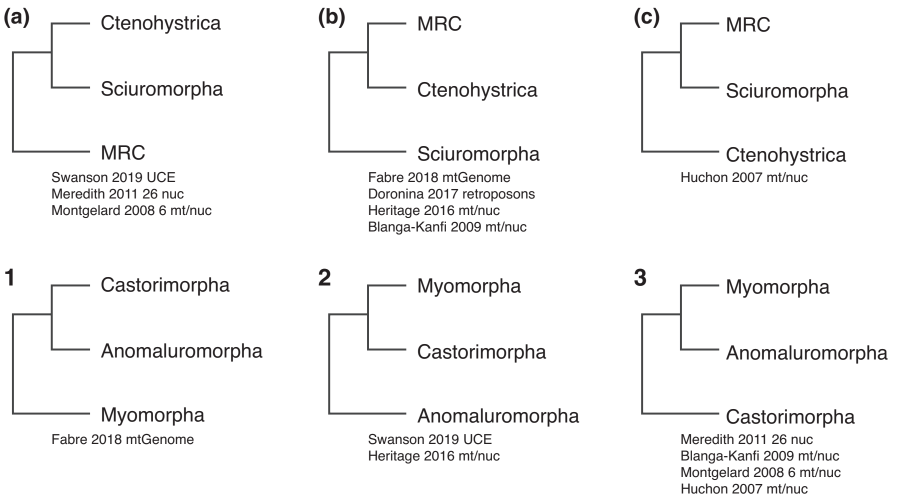

FIGURE 1 The three possible resolutions of the two most contested nodes deep in Rodentia and the recent studies supporting each one. A–C are for the root node between Ctenohystrica (containing Ctenodactylidae, Diatomyidae and Hystricognathi), Sciuromorpha (containing Sciuridae, Aplodontinae and Gliridae) and the mouse-related clade "MRC" (containing Myomorpha, Anomalomorpha and Castorimorpha). 1–3 are for the root node of the MRC between Myomorpha, Anomalomorpha and Castorimorpha

the exception of *Cavia porcellus* and *Castor canadensis*, all samples come from museum specimens previously extracted for other studies (Schenk et al., 2013; Steppan et al., 2004; Steppan & Schenk, 2017), and museum voucher numbers can be found in Table 1. The other two samples were freshly extracted from skin clips of teaching specimens using an Omega EZNA Tissue DNA Kit following the manufacturer's protocol (Omega Bio-tek).

All 12 samples were normalized to 1 μg of DNA in 50 μl of Qiagen elution buffer using the broad-range Qubit kit to quantify concentrations. Samples were sonicated before library preparation using an M220 ultrasonicator (Covaris) for 50 s at 50-W peak incident power, 10% duty factor and 1000 cycles per burst, in order to yield an average size of 500 bp. Instead of using the custom library preparation of Lemmon et al. (2012), we utilized the NEBNext Ultra II DNA library prep kit (New England BioLabs) and followed the manufacturer's protocol. By doing so we eliminated the upfront cost of bulk buying of oligos and library preparation enzymes and therefore library preparation can be done by laboratories that might only want to do so for a few samples and do not want to invest in large quantities of library preparation materials. Each individual was uniquely indexed using the NEBNext Multiplex Dual Index Oligos for Illumina (New England BioLabs) and then pooled after gel verification of library success using a 1.5% agarose gel. Libraries were quantified using a Qubit broad-range kit and pooled at equal concentrations and dehydrated using a SpeedVac Vacuum Concentrator (ThermoFisher) to produce a concentration of 750 ng of DNA in 3.5 μl of Qiagen elution buffer. This pooled library was enriched using the Agilent SureSelect RNA probe enrichment kit following manufacturer's protocol. The final enriched library was verified using a TapeStation 2200 (Agilent Technologies) and qPCR before sequencing on a lane of an Illumina MiSeq pair-end 150-bp v3 sequencer (Illumina) at the Florida State University DNA Sequencing Facility. All cost and product information for the protocol above can be found in Table S1.

2.4 | Sequence processing and alignment

Reads were processed using the SECAPR pipeline (Andermann et al., 2018), which in brief first removed adaptor sequences, palindromic sequences and poor-quality reads (Q-score < 20) using TRIMMOMATIC (Bolger et al., 2014), then assembles reads using ABYSS (Simpson et al., 2009), before finally pulling out target sequences using a reference generated from the mouse (mm10) genome. This reference included the probe target sites along with 500 bp upstream and downstream of each target. All target sites that were within 1000 bp of each other were combined into a single sequence.

Two variables were tested for pulling out target sequences in SECAPR: (i) the percentage identity for a match and (ii) the minimum coverage of the target. Increasing values for these variables decreases the chance for retrieving the correct match, while decreasing these values increases the chance for pulling out multiple matches for a target due to incorrect matches. Values of 30%, 40%, 60% and 80% were tested for the minimum coverage and values of 95%, 90%, 85%, 80% and 75% were tested for the percentage identity for a match. Ultimately, we ended up choosing a percentage identity of 75% and a percentage coverage of 60% based on maximizing retrieval of target sequences while minimizing loss of sequences due to paralogue filtering (Table S2). Sequences were then aligned using MUSCLE (Edgar, 2004), remapped using a consensus sequence as reference and finally realigned as suggested by Andermann et al. (2018). All bases with less than 10× read depth were removed using SECAPR in order to exclude potential low-quality areas of the sequences.

In order to expand taxonomic sampling, additional sequences were extracted *in silico* from the NCBI reference representative genomes database. These included two lagomorphs as outgroups and 22 additional rodents. The rodents included three species that have whole genomes on NCBI for which we sequenced an additional individual using our probe set (*Cavia porcellus*, *Castor canadensis* and *Peromyscus leucopus*) as well as multiple species within two genera (*Mus* and *Peromyscus*). Sequencing multiple congeners allowed us to examine the amount of within-species and within-genus diversity these target sites contain. Our goal was to create a probe set that can be used at many taxonomic levels (e.g., order, family, genus, species) for systematic analysis, and thus having replicates of species, as well as members of the same genus, and members at higher levels of taxonomy allow us to examine the utility of these probes at multiple levels. Sequences were retrieved by preforming a BLAST search against NCBI reference representative genomes using the reference created from the mouse (mm10) genome. Sequences were then added to the samples sequenced in this study and aligned using MUSCLE. All alignments were then manually inspected using BIOEDIT version 7.0.0 (Hall, 2004) and trimmed to remove poorly aligned regions using the *-gappyout* function of TRIMAL version 1.2 (Capella-Gutiérrez et al., 2009). After inspection, alignments were concatenated using SEQUENCEMATRIX (Vaidya et al., 2011).

2.5 | Phylogenetic analysis

Phylogenies were generated using both concatenated maximum-likelihood (IQ-TREE version 1.6.1; Nguyen et al., 2015) and multispecies coalescent (ASTRAL-III; Zhang et al., 2018) methods. IQ-TREE was run using the partition concatenated alignment for GENESITE bootstrap resampling with 1000 ultrafast bootstrap replicates (UFBOOT2; Hoang et al., 2018) using the CIPRES online portal (Miller et al., 2010). This method reduces the chance for overestimation of support for large genome data sets compared to other maximum-likelihood methods (Hoang et al., 2018). Individual gene trees were estimated in RAXML (version 8.2.12; Stamatakis, 2014) for each of the 416 loci with 1000 rapid bootstraps. All nodes with less than 50% bootstrap support were collapsed and resulting trees were used as input into a multispecies coalescent analysis in ASTRAL-III.

TABLE 1 Voucher information and summary of sequencing success including the number of aligned bases in the final alignment, number of loci generated and the percentage of missing loci for each sample

Species	Voucher	Aligned sequence (bp)	Number of loci	% Missing loci
Allactaga sibirica	USNM 449152	362,495	390	6.7
Apomys microdon	USNM 458919	404,746	406	2.9
Calomyscus sp.	MVZ 191923	466,643	412	1.4
Castor canadensis		283,250	385	7.9
Cavia porcellus		365,389	381	8.9
Graphiurus murinus	SP 6067	398,405	393	6.0
Marmota olympus	UWBM 76262	264,660	379	9.3
Mus cookii	USNM 583802	425,376	408	2.4
Pedetes capensis	CM 95012	325,299	386	7.7
Peromyscus leucopus	OK 15	467,115	409	2.2
Phyllotis xanthopygus	MVZ 182702	452,100	411	1.7
Sciurus stramineus	LSUMZ M936	404,201	398	4.8
	Average	384,973	397	5.1

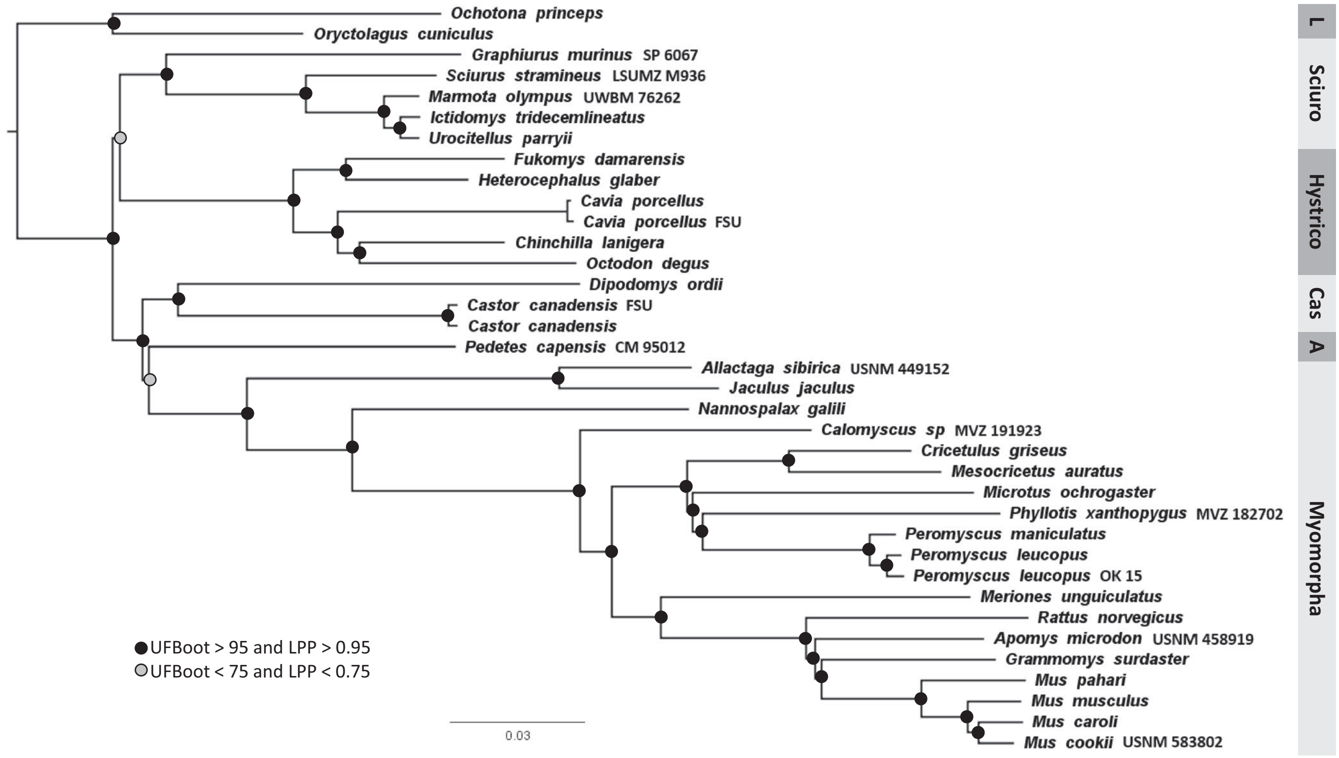

FIGURE 2 Phylogeny for all AHE loci with branch lengths generated by IQ-TREE. Nodes with greater than 95% ultrafast bootstrap support (UFBoot) in IQ-TREE and greater than 0.95 local posterior probably (LPP) in ASTRAL-III are represented with a black dot and all nodes with less than 75% UFBoot and 0.75 LPP are represented with a grey dot. All museum samples used end in museum voucher number and the two samples collected by our laboratory end with "Lab" in order to differentiate newly sequences samples from *in silico* samples. All suborders of rodents are highlighted on the right side along with the outgroup (order Lagomorpha). Abbreviations: "L" for Lagomorpha (outgroup), "Hystrico" for Hystricomorpha, "Sciuro" for Sciuromorpha, "Cas" for Castorimorpha and "A" for Anomaluromorpha

Collapsing the low-support nodes was done to improve accuracy as per Zhang et al. (2018). Support in ASTRAL-III was calculated using local posterior probabilities (LPPs). Statistical tests for resolution of polytomy were performed in ASTRAL-III to assess the ability of these markers to resolve short branches where unresolved polytomies are likely to occur.

3 | RESULTS

3.1 | Rodent probe set

Trial of the vertebrate AHE probes from Lemmon et al. (2012) generated data for 284 probe sites, of which 205 passed BLAST filtering and were used in the design. These 205 probe sets were collapsed in alignments if any sites were within 1000 bp of each other, resulting in 177 loci in the alignments (Table S3). MRBAITS generated 264 new target sites of which 241 passed BLAST filtering and were used in the design; these included probes for *IRBP* and *RAG1*. None of the new target sites were within 1000 bp of each other or the vertebrate probe set, and thus each probe site represents a locus in the generated alignments (Table S3). This resulted in a total of 446 probe sets across 418 target loci.

A total of 21 million raw reads were generated across the 12 samples; 18 million reads passed quality filters with 129,344–272,963 contigs generated per sample. Out of the 418 targeted loci, 416 (99.5%) generated sequences for at least three individuals. On average, 396.5 (94.5%) sequences were generated per sample with a range of 381 (91.1%) to 412 (98.6%) sequences across samples (Table 1). The final alignment generated 570,157 bp with an average of 1371 bp per locus (range 748–3203 bp). After inclusion of *in silico* samples, the final alignment was missing only 5.1% of loci per individual and contained 144,933 parsimony-informative sites, 106,126 singleton sites and 319,098 constant sites.

3.2 | Efficacy at deep nodes: rodent phylogenetics

IQ-TREE and ASTRAL-III produced nearly identical topologies with the only difference being the placement of *Pedetes capensis* (Anomaluromorpha) by ASTRAL-III outside a Castorimorpha plus Myomorpha clade and IQ-TREE placing it sister to Myomorpha with Castorimorpha outside the sister grouping. Support for the node was low in both analyses (61% ultrafast bootstrap support [UFBoot]; 0.47 LPP) and failed the ASTRAL-III polytomy test ($p > .05$). The only other node that failed the polytomy test was at the sister grouping of Sciuromorpha and Hystricomorpha, which was also the only other node with low bootstrap support (67% UFBoot; 0.78 LPP). All other nodes had high support (> 95% UFBoot; >0.95 LPP) for both methods (Figure 2).

3.3 | Efficacy at shallow nodes: Species level

Two genera (*Peromyscus* and *Mus*) had multiple species sampled in order to test the utility of these markers for species-level studies. Pairwise sequence divergence within genera ranged from 1.14% to 3.31%, resulting in an average number of nucleotide differences per locus ranging from 11.5 to 43.7 (Table 2). Two samples of *Peromyscus leucopus* were also included with one coming from *in silico* sampling of the LL strain genome started with mice from North Carolina while the individual sequenced in this study was collected in Oklahoma. The divergence between these samples was 0.45% and on average contained 5.0 nucleotide differences per locus (Table 2). All relationships within the two genera received maximal support (> 100% UFBoot; 1.0 LPP) across both phylogenetic methods (Figure 2). Most individual gene trees correctly resolved all within-genus relationships for *Mus* and *Peromyscus* (68% and 87%, respectively) with moderately high supports (> 70% bootstrap support in RAXML).

4 | DISCUSSION

A large set of genomic markers has not been designed for rodents to date, hindering evolutionary studies into relationships of the group as a whole. Here we present a genomic probe set designed and optimized for rodents using existing library preparation kits that make the protocol easy to replicate with little laboratory investment. This design utilizes an AHE approach to generate longer loci with a higher frequency of substitution than existing

TABLE 2 Comparison of within-genus variation for (a) *Mus* and (b) *Peromyscus*. Values below the diagonal are pairwise sequence divergence across all loci and above the diagonal are average number of nucleotide differences per locus for each pair of species. Below each table is the total number of SNPs across the within-genus alignments

(a) ***Mus***				
	M. caroli	***M. cookii***	***M. musculus***	***M. pahari***
M. caroli	–	11.5	24.2	43.7
M. cookii	1.14%	–	13.8	27.0
M. musculus	1.79%	1.38%	–	44.2
M. pahari	3.28%	2.61%	3.31%	–
Total SNPs:	21,010			

(b) ***Peromyscus***			
	***P. leucopus* (NC)**	***P. leucopus* (OK)**	***P. maniculatus***
P. leucopus (NC)	–	5.0	8.5
P. leucopus (OK)	0.46%	–	7.1
P. maniculatus	0.85%	0.79%	–
Total SNPs:	4295		

vertebrate ultraconserved elements (UCEs) in the hope that it could be used at a greater range of taxonomic levels within the order.

The *Rodent 418 Loci* probe set presented here had extremely high efficiency, with alignments generated for 416 of the 418 targeted loci (99.5%) and an average of 396.5 loci (94.5%) generated per sample after all filtering. This allowed us to generate an alignment of more than a half a million base pairs with over 140,000 parsimony-informative sites. This efficiency is quite high given that our initial test of the vertebrate AHE probe set (Lemmon et al., 2012) generated 285 out of 512 targeted loci (56%), and a publication that came out during our design and testing (Swanson et al., 2019) using the *Amniote 5060 UCE* probe set (Faircloth et al., 2012) generated 2213 out of the 5060 targeted loci (44%). The high efficiency here is a result of our use of the AHE tiling approach, which reduces the chance of losing targets due to mismatches in target probes, and the fact that this probe set is designed using rodent genomes across all five rodent suborders.

Sequences from this *Rodent 418 Loci* probe set generated a well-supported phylogeny for Rodentia with all nodes showing high support with the exception of two nodes deep in the rodent tree that have been hard to resolve in past studies as well (e.gBlanga-Kanfi et al., 2009; Fabre et al., 2018; Heritage et al., 2016; Meredith et al., 2011; Montgelard et al., 2008; Swanson et al., 2019). Our phylogeny largely matches one generated from UCEs (Swanson et al., 2019). Comparing these two genomic phylogenies shows that the UCE phylogeny generated higher support for the two deeper nodes but with lower support than AHE in the more recent splits. This matches the results in a recent comparison of squamate UCEs and AHE and was attributed to the tradeoff between saturation and polymorphism (Karin et al., 2020). UCEs are more conserved and therefore saturation is less of a problem. There is also less recombination potential in UCEs since they are smaller, which results in a lower chance of chimeric gene tree history (e.g., lower chance of two different gene genealogies being combined in one locus due to recombination). Therefore, for nodes that are deep in the phylogeny with high gene tree discordance (like those at the base of Rodentia), UCEs are slightly more informative, but within families and genera AHE targets have more informative sites and in more recent divergences there has been less time for recombination to create discordance among individual gene trees. Nonetheless, given the high sequencing efficiency, this probe set should even be informative across deeper nodes of placentals as well.

In both data sets, the AHE generated here and the UCEs from Swanson et al. (2019), 21 species were shared because of the use of published genomes in both. When comparing these two data sets for shared species, ours had 30.4% more polymorphism and contained 4.49 times as many SNPs per locus. The difference in polymorphism is highlighted in the most recent split between shared species, *Mus–Rattus*, in which the AHE data had more than twice the divergence of the UCEs (6.2% vs. 2.9%) and more than seven times as many SNPs per locus (85 vs. 12). Therefore, the AHE design presented here provided more utility within rodent suborders while the amniote UCE design provided more utility between suborders, but with limited use at lower taxonomic scales.

We included multiple species for two genera (*Mus* and *Peromyscus*) to test the potential of the *Rodent 418 Loci* probe set for shallow taxonomic scales and examined within-genus pairwise distance and phylogenetic support. All within-genus comparisons had greater than 1% sequence divergence with 11.5–43.7 SNPs on average per locus (Table 2). Even within species there was 0.45% sequence divergence between the two samples of *Peromyscus leucopus* (originating from Oklahoma and North Carolina), highlighting the potential utility of this probe set for species delimitation.

At deeper taxonomic levels, the *Rodent 418 Loci* phylogeny resulted in a similar topology as the Swanson et al. (2019) phylogeny with both studies finding high support (>95 bootstrap and >0.95 posterior probability) in all nodes except for two; the *Rodent 418 Loci* phylogeny weakly favours the sister placement of Ctenohystrica and Sciuromorpha relative to the MRC (Figure 1, tree A), and the sister placement of Anomaluromorpha and Myomorpha relative to Castorimorpha within the MRC (Figure 1, tree 3). However, the *Rodent 418 Loci* phylogeny did provide strong support for the monophyly of the MRC, another node that is difficult to resolve. While the topology in both studies is largely the same, low taxonomic sampling probably precludes the abilities to get high support at these two difficult-to-resolve, early Palaeocene nodes, an aspect beyond the scope of this study.

5 | CONCLUSION

We present a new probe set design specifically for rodents to be used at all taxonomic levels in the order. These >400 loci proved informative at all levels within suborders with higher success at retrieving targets than existing AHE (Lemmon et al., 2012) and UCE (Swanson et al., 2019) designs. We also describe an easy-to-use protocol for library preparation using existing kits that require little up-front cost for laboratories wanting to use this probe set for a smaller number of samples, increasing the utility for laboratories with limited core resources. Sequences from this probe set can also be combined with many existing sequence data sets since it includes many loci (*N* = 177) used in the vertebrate AHE probe set (Lemmon et al., 2012) and two common nuclear genes (*RAG1* and *IRBP*) used in many rodent phylogenies (e.gFabre et al., 2012; Jansa et al., 2009; Meredith et al., 2011; Pagès et al., 2016; Parada et al., 2013; Steppan & Schenk, 2017). With this probe design, we hope that more laboratories can easily generate data for answering questions from species delimitation to understanding relationships among families or more inclusive taxa in rapid radiations.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at Dryad https://doi.org/10.5061/dryad.x69p8czjx.

DATA AVAILABILITY STATEMENT

All filtered reads have been submitted to the NCBI Short Read Archive (BioProject accession no. PRJNA748050; sample accession nos. SAMN20307343–SAMN20307354). All alignments, probe sequences and supporting information is available on Dryad at https://doi.org/10.5061/dryad.x69p8czjx.

ACKNOWLEDGEMENTS

This work was funded by National Science Foundation grant DEB-1754748 to S.J.S. and with postdoctoral support for M.R.B. Thanks are expressed to United States National Museum, University of Michigan Museum of Zoology, University of Washington Burke Museum, Louisiana State University Museum of Zoology, and Carl Saltzberg for providing samples for this study, and to the Center for Anchored Phylogenomics for information on probe set success in *Apomys*. Also, we would also like to acknowledge the technical support provided by Tyler Chafin (MRBAITS) and the UCSC Genomics Institute (mouse 60-way alignment and parsing tools). Analytical resources were provided by the Arkansas High Performance Computing Center (AHPCC).

CONFLICTS OF INTEREST

None reported.

ORCID

Max R. Bangs https://orcid.org/0000-0002-9506-2815

REFERENCES

Andermann, T., Cano, Á., Zizka, A., Bacon, C., & Antonelli, A. (2018). SECAPR - a bioinformatics pipeline for the rapid and user-friendly processing of targeted enriched Illumina sequences, from raw reads to alignments. *PeerJ*, *6*, e5175. https://doi.org/10.7717/peerj.5175

Blanga-Kanfi, S., Miranda, H., Penn, O., Pupko, T., DeBry, R. W., & Huchon, D. (2009). Rodent phylogeny revised: Analysis of six nuclear genes from all major rodent clades. *BMC Evolutionary Biology*, *9*(1), 1–12. https://doi.org/10.1186/1471-2148-9-71

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Chafin, T. K., Douglas, M. R., & Douglas, M. E. (2018). MrBait: Universal identification and design of targeted-enrichment capture probes. *Bioinformatics*, *34*(24), 4293–4296. https://doi.org/10.1093/bioinformatics/bty548

D'Elía, G., Fabre, P. H., & Lessa, E. P. (2019). Rodent systematics in an age of discovery: Recent advances and prospects. *Journal of Mammalogy*, *100*(3), 852–871. https://doi.org/10.1093/jmammal/gyy179

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Fabre, P. H., Hautier, L., Dimitrov, D., & Douzery, E. J. (2012). A glimpse on the pattern of rodent diversification: A phylogenetic approach. *BMC Evolutionary Biology*, *12*(1), 1–19. https://doi.org/10.1186/1471-2148-12-88

Fabre, P. H., Tilak, M. K., Denys, C., Gaubert, P., Nicolas, V., Douzery, E. J., & Marivaux, L. (2018). Flightless scaly-tailed squirrels never learned how to fly: A reappraisal of Anomaluridae phylogeny. *Zoologica Scripta*, *47*(4), 404–417. https://doi.org/10.1111/zsc.12286

Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, *61*(5), 717–726. https://doi.org/10.1093/sysbio/sys004

Hall, T. (2004). *BioEdit version 7.0.0*. Distributed by the author, website: www. mbio. ncsu. edu/BioEdit/bioedit. html.

Heritage, S., Fernández, D., Sallam, H. M., Cronin, D. T., Echube, J. M. E., & Seiffert, E. R. (2016). Ancient phylogenetic divergence of the enigmatic African rodent Zenkerella and the origin of anomalurid gliding. *PeerJ*, *4*, e2320.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522. https://doi.org/10.1093/molbev/msx281

Huchon, D., Chevret, P., Jordan, U., Kilpatrick, C. W., Ranwez, V., Jenkins, P. D., Brosius, J., & Schmitz, J. (2007). Multiple molecular evidences for a living mammalian fossil. *Proceedings of the National Academy of Sciences*, *104*(18), 7495–7499. https://doi.org/10.1073/pnas.0701289104

Jansa, S. A., Giarla, T. C., & Lim, B. K. (2009). The phylogenetic position of the rodent genus Typhlomys and the geographic origin of Muroidea. *Journal of Mammalogy*, *90*(5), 1083–1094.

Karin, B. R., Gamble, T., & Jackman, T. R. (2020). Optimizing phylogenomics with rapidly evolving long exons: Comparison with anchored hybrid enrichment and ultraconserved elements. *Molecular Biology and Evolution*, *37*(3), 904–922. https://doi.org/10.1093/molbev/msz263

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology*, *61*(5), 727–744. https://doi.org/10.1093/sysbio/sys049

Lessa, E. P., Cook, J. A., D'Elía, G., & Opazo, J. C. (2014). Rodent diversity in South America: Transitioning into the genomics era. *Frontiers in Ecology and Evolution*, *2*, 39. https://doi.org/10.3389/fevo.2014.00039

Meredith, R. W., Janečka, J. E., Gatesy, J., Ryder, O. A., Fisher, C. A., Teeling, E. C., Goodbla, A., Eizirik, E., Simão, T. L., Stadler, T., & Murphy, W. J. (2011). Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. *Science*, *334*(6055), 521–524.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). *Creating the CIPRES Science Gateway for inference of large phylogenetic trees*. In *2010 Gateway Computing Environments Workshop (GCE)*, pp 1–8.

Montgelard, C., Forty, E., Arnal, V., & Matthee, C. A. (2008). Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. *BMC Evolutionary Biology*, *8*(1), 1–16. https://doi.org/10.1186/1471-2148-8-321

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/molbev/msu300

Pagès, M., Fabre, P. H., Chaval, Y., Mortelliti, A., Nicolas, V., Wells, K., Michaux, J. R., & Lazzari, V. (2016). Molecular phylogeny of South-East Asian arboreal murine rodents. *Zoologica Scripta*, *45*(4), 349–364. https://doi.org/10.1111/zsc.12161

Parada, A., Pardiñas, U. F., Salazar-Bravo, J., D'Elía, G., & Palma, R. E. (2013). Dating an impressive Neotropical radiation: Molecular time estimates for the Sigmodontinae (Rodentia) provide insights into its historical biogeography. *Molecular Phylogenetics and Evolution*, *66*(3), 960–968. https://doi.org/10.1016/j.ympev.2012.12.001

Schenk, J. J., Rowe, K. C., & Steppan, S. J. (2013). Ecological opportunity and incumbency in the diversification of repeated continental colonizations by muroid rodents. *Systematic Biology*, *62*(6), 837–864. https://doi.org/10.1093/sysbio/syt050

Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, *19*(6), 1117–1123. https://doi.org/10.1101/gr.089532.108

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steppan, S. J., Adkins, R. M., & Anderson, J. (2004). Phylogeny and divergence date estimates of murid rodents based on multiple nuclear genes. *Systematic Biology*, *53*, 533–553.

Steppan, S. J., & Schenk, J. J. (2017). Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. *PLoS One*, *12*(8), e0183070. https://doi.org/10.1371/journal.pone.0183070

Swanson, M. T., Oliveros, C. H., & Esselstyn, J. A. (2019). A phylogenomic rodent tree reveals the repeated evolution of masseter architectures. *Proceedings of the Royal Society B*, *286*(1902), 1–9.

Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*, *27*(2), 171–180. https://doi.org/10.1111/j.1096-0031.2010.00329.x

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(6), 153. https://doi.org/10.1186/s12859-018-2129-y

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Bangs, M. R., & Steppan, S. J. (2022). A rodent anchored hybrid enrichment probe set for a range of phylogenetic utility: From order to species. *Molecular Ecology Resources*, 22, 1521–1528. https://doi.org/10.1111/1755-0998.13555