Tempo and mode of evolution of oryzomyine rodents (Rodentia, Cricetidae, Sigmodontinae): A phylogenomic approach

Alexandre Reis Percequillo a,b,*, Joyce Rodrigues do Prado a, Edson Fiedler Abreu a, Jeronimo Dalapicolla b,c, Ana Carolina Pavan a, Elisandra de Almeida Chiquito a,d, Pamella Brennand a, Scott J. Steppan e, Alan R. Lemmon f, Emily Moriarty Lemmon f, Mark Wilkinson b

ABSTRACT

The tribe Oryzomyini is an impressive group of rodents, comprising 30 extant genera and an estimated 147 species. Recent remarkable advances in the understanding of the diversity, taxonomy and systematics of the tribe have mostly derived from analyses of single or few genetic markers. However, the evolutionary history and biogeography of Oryzomyini, its origin and diversification across the Neotropics, remain unrevealed. Here we use a multi-locus dataset (over 400 loci) obtained through anchored phylogenomics to provide a genome-wide phylogenetic hypothesis for Oryzomyini and to investigate the tempo and mode of its evolution. Species tree and supermatrix analyses produced topologies with strong support for most branches, with all genera confirmed as monophyletic, a result that previous studies failed to obtain. Our analyses also corroborated the monophyly and phylogenetic relationship of three main clades of Oryzomyini (B, C and D). The origin of the tribe is estimated to be in the Miocene (8.93–5.38 million years ago). The cladogenetic events leading to the four main clades occurred during the late Miocene and early Pliocene and most speciation events in the Pleistocene. Geographic range estimates suggested an east of Andes origin for the ancestor of oryzomyines, most likely in the Boreal Brazilian region, which includes the north bank of Rio Amazonas and the Guiana Shield. Oryzomyini rodents are an autochthonous South America radiation, that colonized areas and dominions of this continent mainly by dispersal events. The evolutionary history of the tribe is deeply associated with the Andean cordillera and the landscape history of Amazon basin.

1. Introduction

Within the rodent subfamily Sigmodontinae, the tribe Oryzomyini represents an impressive group of rats and mice, comprising 30 extant genera and an estimated 147 species, about 30% of the generic and specific diversity of the subfamily (Burgin et al., 2018). This group has a broad geographic distribution, encompassing nearly the entire range of the subfamily as a whole, from southern South America to Southeastern North America and the Galapagos Islands (Prado and Percequillo, 2018). Besides the wide geographical distribution, these rodents also occupy...
different environments, ranging from forests, shrublands, grasslands and marshes, to lowlands and high montane areas (Weksler, 2006).

The tribe Oryzomyini is widespread in the Neotropical region, but its genera and species are unevenly distributed, occupying different areas west of the Andes, Andes, and east of Andes regions (Prado and Percequillo, 2013). Based on patterns of species diversity and spatial distribution, Prado et al. (2015) recovered four areas of endemism for the tribe in South America, associated with the northern Andes, Guianas, Atlantic Forest and Galápagos. This pattern is shared with other groups of rodents such as the echimyids (Fabre et al., 2017), that have areas of high species richness in the forests of eastern Brazil (Atlantic Forest), Guyana region (Lowland Amazonian Forest) and the Northern Andes (Montane Forests).

The temporal origin of this diversity has been imprecisely estimated from the Miocene to the Pliocene (from 11 to 3.4 million years ago [mya]; González et al., 2020; Leite et al., 2014; Parada et al., 2015, 2013; Stepan and Schenk, 2017). The geographical origin of this diversity is also controversial. One of the first biogeographic hypotheses for the origin of Oryzomyini was based on patterns of species diversity and pointed to the northern Andes as the Area of Origin Differentiation (AOD), with subsequent dispersal and further diversification throughout South America (Reig, 1986). The original-area hypothesis was re-evaluated by Weksler (2006) who concluded that “the application of the progression rule to the current phylogeny suggests a South American ancestral area for oryzomyines” (Weksler, 2006:88), but with the caveat that “a more precise location of the oryzomyine ancestral area in South America requires delimitation of smaller units for analysis than the three general categories provided here” [i.e., trans-, cis- and Andean] (Weksler, 2006:89). Leite et al. (2014:7) hypothesized that the ancestral range for the tribe is ambiguous (i.e. its inference not restricted to a single area), the most recent common ancestor of the tribe “extended its distribution into Amazonia via dispersal”, and most genera have ancestral distributions in eastern South America. Schenk and Steppan (2018:421), on the other hand, found in their biogeographic reconstructions that Oryzomyini initially diversified in the “Guiana Highlands/Amazon basin” region. This last result was similar to that obtained by Maestri et al. (2019), whom suggested an Amazonian origin for this tribe. From the mainland, oryzomyine rodents invaded remote and volcanic islands, in both Atlantic and Pacific oceans, namely the Fernando de Noronha and Galápagos archipelagos (respectively), and also the Antilles (Brace et al., 2015; Machado et al., 2014), suggesting that long-distance dispersal was a frequent process in the history of this group.

Recently, there have been remarkable advances in the understanding of the morphological, generic and species diversity, systematics and biogeography of the tribe Oryzomyini (Chiquito et al., 2014; Leite et al., 2014; Parada et al., 2013, 2015; Percequillo et al., 2011b; Pine et al., 2012; Prado et al., 2015; Prado and Percequillo, 2018; Schenk and Steppan, 2018; Steppan and Schenk, 2017; Weksler, 2003, 2006; Weksler et al., 2006; Prado and Percequillo, 2013). Phylogenetic results recovered the Oryzomyini as a monophyletic group, organizing its diversity in four (A, B, C and D; Fabre et al., 2012; Hanson and Platt II, 2019; Percequillo et al., 2011b; Weksler, 2003, 2006) or three main clades (B, C and D; Parada et al., 2013, 2015; Schenk and Steppan, 2018; Steppan and Schenk, 2017). This uncertainty is in the deep nodes of the tree, regarding the relationships between Scolomys and Zygodontomys, which are either recovered as a single clade or split into two lineages. Most of the studies used concatenated mitochondrial cytochrome b (cyt b) and nuclear interphotoreceptor retinoid-binding protein (IRBP) sequences taken from the same individuals. Exceptions are Hanson and Platt II (2019) which included one additional marker (int2 of the alcohol dehydrogenase gene, Adh1-12) and Schenk and Steppan (2018) and Steppan and Schenk (2017), which included other markers (exon 11 of breast cancer 1 gene, BRCA1; exon 10 of the growth hormone receptor, GHR; and single exon of the recombination activation gene 1, RAG1; besides cyt b and IRBP) and specimens in their dataset. Despite minor differences in genetic markers and taxonomic sampling these studies have supported conflicting relationships among genera within clades. This suggests that most conflicts in relationships and values of support are likely a result of the limitations on the markers employed, the rapid diversification of the oryzomyine radiation, the taxonomic scope of analyses and/or the use of different optimality criteria. Thus, the small number of loci (IRBP and cyt b) repeatedly employed may not provide sufficient information to confidently resolve evolutionary relationships that were hypothesized to have occurred in a short span of time (as being suggested by Fabre et al. (2012), Schenk and Steppan (2018), and Steppan and Schenk (2017)).

The tribe is also characterized by an impressive phenotypic variation in external, skeletal and dental traits (Carleton, 1980; Carleton and Musser, 1989; Steppan and Schenk, 2018; Weksler, 2006), that has been valuable and informative for the definition of taxa at the ranks of species (Percequillo et al., 2008; Prado and Percequillo, 2018) and genus (Weksler et al., 2006; Percequillo et al., 2011b). Morphology-only analytical approaches also recovered the monophyly of clades A and C but not B and D as clades (Weksler, 2006). Combined analyses of morphology and molecules recover all four groups as monophyletic with attendant synapomorphies mostly related to cranial and dental characters (Weksler, 2006, Appendix 3: p.122).

The development of Next Generation Sequencing (NGS) methods and Multispecies Coalescent Models (MSC; see Johnson et al., 2013; Knowles and Kubatko, 2011; Lemmon et al., 2012; Rannala and Yang, 2003) are important methodological and conceptual developments for phylogenetics. Despite some disagreements over the appropriateness of some coalescent approaches (mainly regarding the adequacy of shortcut methods; Edwards et al., 2016; Gatesy and Springer, 2014; Springer and Gatesy, 2016), the use of multiple genomic loci with either concatenated and/or coalescent analyses is increasingly being employed to address controversial sections of the tree of life (e.g., Bryson et al., 2016; Crawford et al., 2012; Faircloth et al., 2015; McCormack et al., 2012; Streicher et al., 2016). Neither approach has yet been employed to infer phylogenetic relationships for the tribe Oryzomyini.

Here we use a multi-locus dataset obtained through anchored phylogenomics (Lemmon et al., 2012) to investigate the phylogeny of Oryzomyini and to address the potential of this large dataset to resolve conflicting relationships within this rapid-diversifying group of rodents. We also tackle some outstanding questions including: (1) will a phylogenomic approach recover the major clades that are constantly obtained using small scale sequence data? (2) how are these clades related? (3) what is the ancestral area of origin and differentiation of this tribe? (4) which events played important roles in the diversification of the tribe? Therefore, we aim to propose some hypotheses of the phylogeny, tempo and mode of evolution of the species of Oryzomyini and compare these with data on the geological history of South America in order to discuss the biogeographic history of this group in the continent.

2. Material and methods

2.1. Sampling

We selected 25 genera and 47 species of the tribe Oryzomyini, representing a fairly comprehensive sampling of generic diversity that includes 83% of the extant genera, and an evenly distributed sampling of species, with about 34% of known species included (see Table S1, Supplementary Material, for a complete list of specimens examined with information on the collections studied). We also included 7 species of other Sigmodontinae as outgroups (Table S1, Supplementary Material), representing five tribes, Akodontini, Phyllotini, Sigmodontini, Thomomysini and Wiedomyini, as well as one incertae sedis lineage, the genus Abrawayaomys (Patton et al., 2015).
2.2. Probe design

Following Ruane et al. (2015; snakes), Tucker et al. (2017; lizards), and Prum et al. (2015; birds), we improved the vertebrate Anchored Hybrid Enrichment (AHE) target loci of Lemmon et al. (2012) for optimal use in mammals. We first identified the genomic coordinates in the human genome (hg19) corresponding to the coordinates of the extended anchor regions of *Gallus gallus* (galGal4) obtained by Prum et al. (2015) using the UCSC liftover tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver). The corresponding genomic sequences were then extracted and aligned using MAFFT v7.023b (Katoh and Standley, 2013) to that of the regions used by Prum et al. (2015) for probe design. After inspecting the alignments and masking any misaligned regions in Geneious R9 (Biomatters Ltd.; Kearse et al., 2012), 120 bp probes were tiled uniformly across the human sequences at 1.5× density: conserved regions targeted for AHE enrichment were identified using alignments across divergent species. Once the alignments were constructed, improved and the conserved regions were identified, probe sequences of length of 120 bp were generated by sliding a window approach (as described in Lemmon et al., 2012: 729) across the human sequence from the alignment. A new probe was generated starting every 90 bp in the sequences (Probe sequences, alignments and raw sequence data are available in DRYAD, at https://doi.org/10.5061/dryad.zkh189394).

2.3. Data Generation

Sequence data were generated and analyzed at the Florida State University Center for Anchored Phylogenomics (www.anchoredphylogeny.com), following methods described in Lemmon et al. (2012), Prum et al. (2015), and Hamilton et al. (2016). Extracted DNA was quantified using Qubit and run on gels to assess initial DNA quality. Approximately 500 ng of input DNA was sheared to 200–500 bp fragments using a Covaris ultrasonicator. Indexing Illumina library preparations were performed on a Beckman Coulter FXp liquid-handling robot. Libraries were then pooled in groups of 16 samples and enriched using an enrichment kit produced by Agilent technologies that contained the probes described above. Enriched libraries were pooled for sequence on two HiSeq 2500 lanes, using a paired-end 150 bp protocol with 8 bp (single) indexing. Sequencing was performed at the Translational Laboratory in the College of Medicine at Florida State University.

2.4. Data processing

Reads passing the Illumina CASAVA high-chastity filter were demultiplexed using the 8-bp indexes (with no mismatches tolerated). Overlapping read pairs were merged using the Bayesian approach developed by Roktya et al. (2012). Adapters were removed and sequencing errors in overlapping regions were corrected during the process. Reads were assembled using the quasi-de novo approach described by Hamilton et al. (2016), with *Homo sapiens* serving as the reference during assembly. Consensus sequences were called from assembly clusters containing at least 100 reads. Orthology among consensus sequences for each locus was determined using a neighbor-joining approach that utilized alignment free pairwise distances computed using the % of shared kmers (short sequences), with one sequence per individual being allowed in each orthologous group (for more details, see Hamilton et al. 2016). Putative orthologous sets containing more than 50% of the individuals were utilized downstream as independent loci. Sequences were aligned using MAFFT v7.023b (Katoh and Standley, 2013) and trimmed/masked to reduce the amount of missing data (for details, see Hamilton et al., 2016).

2.5. Phylogenetic inferences

Phylogenetic relationships were estimated using a concatenation (supermatrix) approach and a coalescent-based species tree (supertree) approach. For the former, alignments were concatenated across loci, then used as input for maximum likelihood (ML) phylogeny estimation using RAxML (v2.2.3; Stamatakis, 2006) and IQ-Tree (v2.1.1; Minh et al., 2020). RAxML was run with default parameter with GTR + G as the model of nucleotide substitution partitioned by locus; branch support for RAxML analysis was accessed using 100 rapid bootstrap replications. We performed IQ-Tree analysis in the CIPRES Science Gateway (Miller et al., 2010) also with default parameters and the GTR + G model of substitution partitioned by locus; branch support was assessed through 1,000 ultrafast bootstrap replicates (UFBoot2; Hoang et al., 2018). Individual gene trees were estimated in RAxML in the same fashion as the supermatrix analysis and the resulting gene trees used as input for species tree analysis using ASTRAL (v.4.10.0; Mirarab and Warnow, 2015). In ASTRAL, support was assessed using 100 rapid bootstrap replications.

2.6. Date estimation

Divergence times were estimated using standard models of evolution implemented in BEAST v. 2.5.1 (Bouckaert et al., 2019). Considering that date estimation analysis using the complete dataset was intractable, and that there is a concern in the literature (e.g., Smith et al., 2018) regarding model mis-specification due to the extent of topological and rate heterogeneity across genes, we implemented the SortaDate approach (Smith et al., 2018). This approach filters a subset of loci according to three criteria: (1) clock-likeness (determined from root-to-tip variance), (2) reasonable tree length, and (3) the least topological conflict with the inferred species tree. We chose the top 40 loci to conduct the divergence time estimation and modeled a concatenated alignment with a single (GTR) substitution model.

We used a log-normal relaxed clock with a Yule tree prior and performed Markov Chain Monte Carlo runs for 300 million generations, sampling every 60,000 generations. We confirmed that effective sample sizes were above the threshold of 200 in Tracer v1.7.1 (Rambaut et al., 2018). A time tree was obtained with TreeAnnotator v2.5.1 (Bouckaert et al., 2019) using the maximum clade credibility tree from all post burn-in trees without posterior limits for each node, and node heights as mean heights.

We constrained four calibration points in the phylogeny. The first three employing a log-normal distribution for (i) the fossil records of genus *Holochilus* (mean of 1 ± 1.25 my, and offset 0.8; 5% quantile = 0.85 and 95% quantile = 4.38; Pardini et al., 2002; Pardini and Teta, 2011), (ii) the fossil records of genus *Oligoryzomys* (mean of 1 ± 1.25 my, and offset 0.8; 5% quantile = 0.85 and 95% quantile = 4.38; Pardini et al., 2002; Voglino and Pardinas, 2005), and (iii) the divergence between *Phyllostis* and *Phaenomys* (mean of 1 ± 1.25 my, and offset 4.0; 5% quantile = 4.06 and 95% quantile = 7.58; Goncalves et al., 2020). The last calibration point (iv) employed a normal distribution based on the origin of the Galapagos archipelago about 3 mya (5% quantile = 1.36 and 95% quantile = 4.64; Grehan, 2001), and was constrained to the most recent common ancestor of *Nesoryzomys* and *Aegialomyx*.

2.7. Biogeographic analysis

We investigated the biogeographic history of *Oryzomyini* employing the R package BioGeoBEARS (Matzke, 2013). We estimated ancestral ranges of inferred nodes and investigated how historical events (vicariance, anagenetic dispersal, and founder-event speciation or jump dispersal, i.e., a “jump” of the lineage to a new area of occurrence, outside of the node ancestral range; Matzke, 2013) may have shaped the clade-dispersive pattern of the tribe *Oryzomyinae*. We compared three different models of range evolution on a phylogeny under the Maximum Likelihood criterion, with or without the founder-event speciation as an additional cladogenesis event (+J; Matzke, 2014): (1) Dispersal-Extinction-Cladogenesis (DELIKE; based on Ree et al., 2005), (2)
We, therefore, decided to perform the statistical model comparison (AIC), Klaus and Matzke (2020:5-6) argued that Ree and Sanmartín comparisons with other models using the Akaike Information Criteria Sanmartín (2018) that the parameter et al., 2019; Schenk and Steppan, 2018 ). Despite the criticism of Ree and
occupy up to three distinct areas (Andes) or six distinct areas (Domin
preliminary tests (results not shown) with differential dispersal rates
code the species as present in the referred area. For the second analysis (hereafter “Dominion”), taxa were assigned to ten distinct areas across the Neotropical region, slightly modified from the biogeographic dominions of Morrone (2014: 24, Fig. 12): (1) MEX: Mexican transition zone, that includes the highlands of southern Mexico and transitional areas along the Pacific and Atlantic coasts; (2) MES: Mesoamerican dominion, that encompasses the lowlands of Central America, from Hondurans northwards, along the Pacific and Atlantic coastal regions; (3) PAC: Pacific dominion, that includes the northern Andes, from northern Peru to Venezuela and part of Central America; (4) GAL: Galápagos archipelago; (5) BBRA: Boreal Brazilian dominion, that abridges the Weh, and also the transitional habitats on Patagonia. We assigned the same dispersal probability among areas (adjacent or not) because preliminary tests (results not shown) with differential dispersal rates among areas produced the same results but with substantially increased computation time. Reconstructions of ancestral nodes were allowed to occupy up to three distinct areas (Andes) or six distinct areas (Dominion), which correspond to the number of areas occupied by the species and clades of oryzomyines with the largest geographic ranges in each case. We only presented areas with ancestral range probabilities higher than 0.3.

3. Results

3.1. Summary of phylogenomic dataset

Our concatenated matrix comprised 402 loci representing 598,786 sites (of which 141,960 are variable and from the total variable sites 86,839 are informative) and a total of 44,908,950 bases, with only 11.6% of missing data. The average length of the 402 loci was 1,489 base pairs, ranging from 255 to 2,623 bp. We gathered genomic information from 25 genera and 47 species within the tribe Oryzomyini, as well as 7 additional sigmodontine species used as outgroups. We were not able to generate adequate data for five oryzomyini genera, named Amphinec-tomys, Handleyomys, Microkudontomys, Sigmodontomys and Zygodontomys.

3.2. Systematics of the tribe Oryzomyini

Species tree (ASTRAL) and supermatrix (RAxML and IQ-Tree) analyses produced topologies with strong support for most branches (Fig. 1; a complete tree including the outgroups is available in the Supplementary Material, Fig. S1). All genera represented by more than one species were confirmed as monophyletic. The analyses recovered the monophyly of the tribe Oryzomyini and most of the genera arranged in three clades, with the genus Scolomys (only representative of clade A, preceding the clade of the monophyly of clade A; sensu Weksler, 2003, 2006) as the sister to all other taxa, and clade B as sister to clades C + D (names of the clades as presented by Weksler, 2003, 2006). Within clade B, our analysis recovered the genera “Handleyomys” (for explanation on quotation marks on the genus name see Taxonomic Implications below; see also Weksler, 2015; Weksler et al., 2006), Mindomys, Nephelomys, Oecomys, Transandinomys, Euryoryzomys and Hylaemys. Clade C comprised the four genera Oligoryzomys, Neacomys, Oreoryzomys and Microrzomy. Clade D was the most genus inclusive group, comprising Ereomoryzomys, Drymoreomys, Cerradomys, Lundsomys, Sooretamys, Holochilus, Pseudoryzomys, Nectomys, Oryzomys, Nesoryzomys, Aegialomys, Tanyuromys and Melanomys. Within clade D are the only topological differences between coalescent and concatenated analyses (Fig. 1), with the conflicting relationships poorly supported in both approaches: with Astral, Cerradomys appears as sister to all other genera (except for Ereomoryzomys and Drymoreomys, that form a clade that is sister to all other members of clade in both approaches), and the clade formed by Lundomys and Sooretamys as sister to all remaining taxa, whereas with RAxML and IQ-Tree Lundomys was recovered as sister to all other genera, whose interrelationships are (Sooretamys (Cerradomys (all genera))).

3.3. Tempo and mode of Oryzomyini evolution

The log-normal relaxed clock with a GTR model of nucleotide substitution recovered estimates for the origin of the tribe Oryzomyini during the Miocene (8.93–5.38 mya; Table 1; Fig. 2; Fig. S1, Table S2, Supplementary Material). The date estimates for the origins of the main internal clades were during the Pliocene, with the origin of clade B between 5.86 and 3.46 mya, clade C at 4.85–2.76 mya, and clade D at 5.76–3.52 mya, with some genera originating in the Pliocene and some in the Pleistocene (Table 1; Figs. 2 and 3; Fig. S1, Table S2, Supplementary Material). Speciation events occurred mostly during the Pleistocene (Figs. 2 and 3; Fig. S1, Table S2, Supplementary Material).

The likelihood-ratio test (LTR) showed that the addition of the founder-event speciation (parameter + J) in our datasets only significantly improved the BAYAREALIKE model in the Dominion analysis (Table 2). In addition, values of ΔAIC < 2 selected only the BAYAREALIKE + J model in Dominion analysis and DECLIKE and DECLIKE + J.

We selected the best model of each analysis to show and discuss the results.
models in the Andes analysis (Table 3). Given that LRTs point to a lack of difference in the statistical support for DECLIKE and DEC + J models in the Andes analysis and neither differences were found in the estimated ancestral areas (results not shown), we have chosen to present and discuss the simplest DECLIKE model as the result of the Andes analysis. This model also fits our data better than other models based on AIC comparisons (LnL = 68.11, AICwt = 0.62; Fig. 2, Table 3, Table S3, Fig S2, Fig S3, Supplementary Material).

Geographic range estimates for the Andes analysis suggested an east of Andes origin for the ancestor of Oryzomyini with the highest probability score (86.2%), and also for the ancestors of all early nodes, as in Clades B, C + D, C and D (73.6, 96, 88.2, 95.9% as the highest probabilities for these four nodes, respectively). This analysis also hypothesizes the occurrence of vicariance (two events) and dispersal (16 events, 5 jump and 11 anagenetic) in clades B, C and D in the history of oryzomyine rodents. The vicariant events were (1) the split between species of Nephelomys in clade B, and (2) between Nectomys and a clade containing Oryzomys, Typhonomys, Melanomys, Nesoryzomys and Aegialomys in clade D. The five jump dispersals were (1) the "Handleyomys" colonization of Central America, (2) the dispersal of the clade formed by Nephelomys and Mindomys to the Andes and west of Andes, (3) the invasion of Transandinomys to the Andes, all in clade B, (4) the

Fig. 1. RAxML (and IQ-Tree) phylogenetic tree of tribe Oryzomyini, indicating the clades recovered; the ASTRAL tree recovered the same topology, with the exception of the alternative relationships shown on the lower right hand. All internal branches have 100% bootstrap support in RAxML, IQ-Tree and ASTRAL analyses, except when annotated (RAxML/IQ-Tree/ASTRAL); these sections of tree with support below 100% also represent incongruence between RAxML (and IQ-Tree) and ASTRAL analyses. Major clades are indicated on the right hand, within the clade areas, there are pictures of species that are representatives of each major clade: Scolomys melanops (photo: P.R.O. Roth); clade B, Nephelomys albiguilaris (photo: Pedro Peloso); clade C, Oligoryzomys maccrossae (photo: E.F. Abreu); clade D, Holochilus brasiliensis (photo: P.R.O. Roth).
colonization of the ancestor of *Microryzomys* and *Oreoryzomys* of the Andes in clade C, and (5) the jump of *Eremoryzomys* to Andes from the east of Andes in clade D. The 11 anagenetic dispersals were in many directions, mainly with species dispersing in genera (1) "*Handleyomys*" from west of Andes to Andes, (2) *Oecomys* from east of Andes to Andes and Andes + west of Andes, (3) *Nectomys* from east of Andes to Andes, (4) *Melanomys* from west of Andes to Andes [2 events], (5) *Aegialomys* from west of Andes to Andes [2 events]; there were also anagenetic events in deeper portions of the phylogeny, with the (6) dispersal of the ancestor of *Nephelomys* to east of Andes, (7) dispersal of ancestor of the clade (*Nectomys* to *Aegialomys*) to west of Andes, and (8) of the ancestor of *Tanyuromys* to Andes, from the west of Andes.

Model testing of Dominion analysis points to BAYAREALIKE + J as the model best explaining the geographic range evolution in Oryzomyini (LnL = -210.8, AICwt = 1; Fig. 3, Table 2, Table S4, Fig S2, Fig. S4, Supplementary Material). This model was not able to reconstruct the ancestral area for the tribe and for the ancestor of clades B and C (occupancy probabilities were inferior to the threshold of 0.3), possibly because our sampling is incomplete for representatives of the species-group taxa, precluding well supported inferences. However, some candidate regions for the ancestral area of the tribe (with probabilities below 30%) are: BBRA plus SBRA, with 19% probability (Table S4, Supplementary Material) and the combination of BBRA, SBRA and AMAZ, with 15.3%; and for the clades C and D, the most likely candidate area is CHA/PAR, with 29% probability. The ancestral range recovered in this analysis for the most recent common ancestor of clades B, C and D

<table>
<thead>
<tr>
<th>Clades</th>
<th>Mean</th>
<th>Median</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oryzomyini</td>
<td>6.88</td>
<td>6.73</td>
<td>8.93</td>
<td>5.38</td>
</tr>
<tr>
<td>B(C + D)</td>
<td>5.57</td>
<td>5.42</td>
<td>7.11</td>
<td>4.38</td>
</tr>
<tr>
<td>C + D</td>
<td>4.90</td>
<td>4.77</td>
<td>6.27</td>
<td>3.82</td>
</tr>
<tr>
<td>B</td>
<td>4.51</td>
<td>4.42</td>
<td>5.86</td>
<td>3.46</td>
</tr>
<tr>
<td>D</td>
<td>4.51</td>
<td>4.4</td>
<td>5.76</td>
<td>3.52</td>
</tr>
<tr>
<td>C</td>
<td>3.67</td>
<td>3.58</td>
<td>4.85</td>
<td>2.76</td>
</tr>
</tbody>
</table>

Fig. 2. Chronogram of the tribe Oryzomini, constructed using BEAST2 implemented with the SortaDate approach (divergence times are presented on Table 1 and Table S2, Supplementary Material), presenting the results of DECLIKE model implemented on BioGeoBEARS package. Boxes at the nodes indicate the ancestral states and boxes at the terminals indicate the current distribution of the species on east of Andes (black), Andes (light gray), west of Andes (dark gray); two-colored boxes indicate that the model recovered more than one area with more than 30% probability of occurrence (the threshold established here; see material and methods). Arrows indicate anagenetic dispersal events and Vs indicate a vicariant event (see text).
is a large area that includes BBRA, SBRA, and AMAZ (Fig. 3). The ancestor of clade B also occupied the BBRA, SBRA, and AMAZ dominions. The ancestor of clade C occupied exclusively the PAC dominion, and the ancestor of clade D was more likely to have occurred along the CHA and PAR dominions (Fig. 3; see Table S4 for the probabilities).

The estimates of historical events in this analysis (Fig. 3, Table 3, Table S4, Fig S2, Fig. S4, Supplementary Material) suggest 30 dispersal events and no vicariance throughout the evolutionary history of Oryzomyini: the model recovered 13 anagenetic dispersal events and 17 jump dispersal or founder-event speciation in all three clades. The jump dispersals were (1) the *Handleyomys* ancestor from Amazon (BBRA, SBRA and AMAZ) to MES, (2) the ancestor of the clade *Mindomys* +
Nephelomys from Amazon to PAC, (3–4) the ancestor of an internal clade of Hylaecymus (sexuatus, oniscus and acritus) from Amazon to PAR and from PAR back to Amazon, SBRA (H. acritus), (5) the jump of Transandinomys from Amazon to PAC and MES, (6–8) the dispersal of the ancestor of Euryoryzomys from Amazon to eastern South America, CHA/PAR, BBRA and DOM, the jump back of the ancestor of E. emmonsae and E. nitidus to AMAZ, and the colonization of SBRA from BBRA by E. nitidus, (9–10) the founder event of the ancestor of Oligoryzomys entering CHA/PAR from the PAC dominion and the invasion of PAC and BBRA by O. delicatus, (11) the dispersal of the ancestor of Neacomys to AMAZ from PAC, (12) the entrance of E. polius in PAC and SBRA from PAR, (13) the invasion of SBRA/CHA by the ancestor of Cerradomys from CHA/PAR, (14) the jump of the ancestor of the clade including Nectomys to Aegialomys to the PAC from CHA/PAR, (15) the dispersal of the ancestor of genus Oryzomys to Central and Northern America, MEX and MES, from PAC; (16) the colonization of Galápagos by Nesoryzomys from PAC, and (17) the jump dispersal from PAC to the Andean region, SAZ, by Aegialomys ica.

The anagenetic dispersal events were also widespread in all clades, as the (1) dispersion of H. alfaroi to MEX and PAC, (2) dispersion of N. albignaiae to SBRA, (3) the migration of H. megacephalus to CHA/PAR, (4–6) the dispersal of the ancestor of Oecocimus to CHA dominion and two independent colonization events of PAC by O. trinitatis and by O. bicolor, (7) the dispersal of N. amoenus to PAR, BBRA and SBRA, (8) the invasion of high Andes (SAZ) by M. altissimus, (9) the subsequent dispersal of the ancestor to Cerradomys to CHA, (10) the invasion of several areas, BBRA, SBRA, AMAZ, by P. simplex, (11–12) the entrance of the ancestor of Nectomys in BBRA/SBRA and the subsequent invasion of N. squamipes to AMAZ/CHA/PAR, and (13) the entrance of O. couesi in Central and South America, MES and PAC.

4. Discussion

4.1. Systematics of the tribe Oryzomyini

4.1.1. Phylogenetic inferences and comparisons with previous hypotheses

The topologies obtained with the genomic dataset are similar to previous published phylogenies, especially those involving supermatrices (hereafter SM; Fabre et al., 2012; Hanson and Platt II, 2019; Maestri et al., 2017; Schenk and Steppan, 2018; Steppan and Schenk, 2017). Despite the absence of Zygodontomys in our taxa-set, our results recovered Scelomys as a basal lineage within the tribe, similar to other studies. Clades B, C and D are also recovered as monophyletic, as in previous aforementioned SM analyses, with clade B as sister to a clade formed by C and D. The morphological analysis performed by Weksler (2006) recovered the same clades, but with different internal relationships, and revealed that the clade A is sister to clade C, and both sister to clade B, all of them sister to clade D (when polymorphic traits were considered as “composites”) or that clades B, C and D, as well as Scelomys and Zygodontomys were recovered as a basal unresolved polytomy (with polymorphic traits treated as ordered).

Within clade B, there are also topological similarities and differences between our and other SM studies. The sister relationship between Euryoryzomys and Transandinomys was recovered in all studies except for Maestri et al. (2017) who recovered Transandinomys nested within a consequently non-monophyletic Euryoryzomys. Hanson and Platt II (2019), Maestri et al. (2017), Schenk and Steppan (2018), and Steppan and Schenk (2017) also recovered a sister relationship between Occomys and the clade comprised of Euryoryzomys and Transandinomys. The position of Mindomys hammondii was always a controversial point: based on combination of morphological and molecular data, Weksler (2006) hypothesized this species to be sister to all other species of the tribe. Differently, Percequillo et al. (2011b) and Pine et al. (2012) suggested that M. hammondii belonged to clade C, as sister to Oecomys. The analysis performed by Maestri et al. (2017) and Hanson and Platt II (2019) recovered Mindomys within Nephelomys. The genomic data employed here supported a hypothesis that this rare species is sister to the genus Nephelomys. The morphological approach of Weksler (2006) employing maximum parsimony recovered species of “Handleomyus” within clade D or B, depending upon the analytical criteria used for polymorphic characters (when these traits were coded as “composite character-state” or “polymorphic”, as suggested by Wiens, 2000, “Handleomyus” appeared in clade D, and when treated as ordered, or “scaled” on the coding scheme of Wiens, 2000, they were recovered in clade B). In one of morphologic approaches (Weksler, 2006: fig. 35), these genera appear an unresolved polytomy, with some analysis including Amphinecotomys savamia (a taxon that is currently associated to clade D; Steppan and Schenk, 2017). The other morphological analysis (Weksler, 2006: fig. 34) showed some genera, Oecomys and Nephelomys, as monophyletic, but other genera, as Hylaecymus and Euryoryzomys, as polyphyletic.

Within clade C, all studies have supported a similar topology, except for the study of Maestri et al. (2017) that did not recover Neacomys as monophyletic. Morphology supports a distinct topology in the “composites” analysis, with Neacomys sister to a clade formed by (Microryzomys, Oreoryzomys) Oligoryzomys, while the “ordered” analysis showed a polytomy.

Within clade D, the most noticeable incongruence is the position of Lundomys, which is usually found in a clade with Holochilus and Pseudoryzomys (Fabre et al., 2012; Pine et al., 2012; Schenk and Steppan, 2018; Steppan and Schenk, 2017). Maestri et al. (2017) recovered this genus in a clade along with (Tanyuromys (Lundomys (Melanomys, Sigmodontomys))), a very unusual result. In our analyses, its phylogenetic position was ambiguous, but in neither analysis did this genus appear closely related to Holochilus or Pseudoryzomys. Perhaps unsurprisingly in light of these ambiguities, this was the section of the trees with the lowest level of support (see Fig. 1). Topological incongruences between

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Model</th>
<th>LnL</th>
<th>P</th>
<th>d</th>
<th>e</th>
<th>j</th>
<th>AIC</th>
<th>ΔAIC</th>
<th>AICwt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andes</td>
<td>DECLIKE</td>
<td>–68.11</td>
<td>2</td>
<td>0.13</td>
<td>0.62</td>
<td>0</td>
<td>140.2</td>
<td>0</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>DECLIKE + J</td>
<td>–68.12</td>
<td>3</td>
<td>0.13</td>
<td>0.62</td>
<td>0.0001</td>
<td>140.5</td>
<td>0.3</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>DIVALIKE</td>
<td>–77.53</td>
<td>3</td>
<td>0.073</td>
<td>1.00E–12</td>
<td>0</td>
<td>159.1</td>
<td>18.9</td>
<td>5.10E–05</td>
</tr>
<tr>
<td></td>
<td>DIVALIKE + J</td>
<td>–76.05</td>
<td>3</td>
<td>0.056</td>
<td>1.00E–12</td>
<td>0.032</td>
<td>158.1</td>
<td>17.9</td>
<td>8.20E–05</td>
</tr>
<tr>
<td></td>
<td>BAYAREALIKE</td>
<td>–83.75</td>
<td>2</td>
<td>0.022</td>
<td>0.16</td>
<td>0</td>
<td>171.5</td>
<td>31.3</td>
<td>1.00E–07</td>
</tr>
<tr>
<td></td>
<td>BAYAREALIKE + J</td>
<td>–68.56</td>
<td>3</td>
<td>0.027</td>
<td>0.0005</td>
<td>0.061</td>
<td>143.1</td>
<td>2.9</td>
<td>0.15</td>
</tr>
<tr>
<td>Dominion</td>
<td>DECLIKE</td>
<td>–219.5</td>
<td>2</td>
<td>0.17</td>
<td>1.39</td>
<td>0</td>
<td>443.2</td>
<td>15.30</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>DECLIKE + J</td>
<td>–219.5</td>
<td>3</td>
<td>0.17</td>
<td>1.39</td>
<td>0.0001</td>
<td>445.5</td>
<td>17.30</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>DIVALIKE</td>
<td>–289.2</td>
<td>2</td>
<td>0.010</td>
<td>0.01</td>
<td>0</td>
<td>582.6</td>
<td>154.70</td>
<td>2.9E–34</td>
</tr>
<tr>
<td></td>
<td>DIVALIKE + J</td>
<td>–288.8</td>
<td>3</td>
<td>0.010</td>
<td>0.26</td>
<td>0.0001</td>
<td>584.1</td>
<td>155.90</td>
<td>1.4E–34</td>
</tr>
<tr>
<td></td>
<td>BAYAREALIKE</td>
<td>–231.5</td>
<td>2</td>
<td>0.025</td>
<td>0.31</td>
<td>0</td>
<td>467.2</td>
<td>39.30</td>
<td>3.3E–09</td>
</tr>
<tr>
<td></td>
<td>BAYAREALIKE + J</td>
<td>–210.8</td>
<td>3</td>
<td>0.022</td>
<td>0.02</td>
<td>0.052</td>
<td>428.2</td>
<td>0.00</td>
<td>6.9994</td>
</tr>
</tbody>
</table>

Table 3 Log-likelihood values (LnL) for each one of six biogeographic models tested to explain the biogeographic range evolution on the phylogeny of the tribe Oryzomyini, for Andes and Dominion analyses. Akaike Information Criterion (AIC), AIC differences (ΔAIC), and Akaike weight (AICwt) were used to compare the models. Parameters number (P), and the values for dispersal (d), extinction (e) and founder (j) parameters are also given.

Molecular Phylogenetics and Evolution 159 (2021) 107120
A.R. Percequillo et al.
coalescent and concatenated approaches might be associated to methodological causes, such as gene-tree estimation error associated to the anchored phylogenomics (see Cai et al., 2020), or biological and evolutionary proprieties of the taxa composing this section of the oryzomyine tree, as rapid diversification with high speciation rates (frequently observed in Sigmodontini rodents; Steppan and Schenk, 2017; Schenk and Steppan, 2018) that could potentially generate incomplete sort of lineages. However, before investigating such candidate phenomena affecting the phylogenetic estimation of this portion of the oryzomyine tree, it would be important to obtain a more complete taxonomic sampling for clad D, as well as to implement methods that incorporate more complex models of tree estimation, branch support and topological comparisons.

The morphologic analyses of Weksler (2006) recovered conflicting topologies among themselves and with our results, despite some similarities as the position of E. polius as sister to all other taxa (in both approaches). The “composite” approach is more resolved but recovered “Handleyomys” within clad D (presently in clad B), while “ordered” analysis presents an unresolved topology, with some monophyletic groups, one with Oryzomys, Pseudoryzomys, Lundomys and Holochilus, and the other with Nesoryzomys, Aeagialomys, Melanomys, Nectomys, Sigmodontomys and Tanyromys.

Our results are characterized by better resolution and maximum support for the majority of clades, substantially higher support than in most previously published phylogenies (e.g. Fabre et al., 2012; Maestri et al., 2017; Percequillo et al., 2011b; Pine et al., 2012; Schenk and Steppan, 2018; Steppan and Schenk, 2017; Weksler, 2006, 2003). Taxonomic sampling was similar among these analyses with most of them employing the same sequences from the same individuals (and thus with similar amounts of missing data), with the exception of the studies of Schenk and Steppan (2018) and Steppan and Schenk (2017) who employed distinct datasets, both on the specimens and sequences employed. Maestri et al. (2017) has a comprehensive dataset, but for oryzomyine rodents 56% of their taxonomic dataset was represented by only two genes (mainly CYTB, COI and IRBP) in their supermatrix and only 3.3% of taxa by 8 genes. Thus, most topological conflicts and differences in support values are more likely to be related to the limitations of the markers employed and also to the distinct optimality criteria.

Parsimony results obtained by Weksler (2006) are quite variable, with (1) morphology-only analyses recovering well resolved trees, with no support for the most inclusive clades (Weksler, 2006:62; Fig. 34), or trees with basal polytomies (Weksler, 2006:63, Fig. 35), and no or low support for more or less inclusive clades; (2) molecular analysis producing some polytomies and low to moderate support (Weksler, 2006:64, Fig. 36); and (3) combined analysis, which generated more resolved trees, but with only low to moderate support (figs. 37 and 38, respectively). Nevertheless, most approaches (in Weksler, 2006) supported the existence of three or four clades (B, C and D; clad A, with Scolomys + Zygodontomys not recovered in IRBP-only analysis). In general, the relationships between genera were completely unstable in the analyses performed by Weksler (2006), even on the more resolved trees. The combined matrix (two genes and morphology) of Percequillo et al. (2011b) recovered similar results to those obtained by Weksler (2006), with four clades and slightly lower levels of support for some branches using parsimony analysis.

Probabilistic inferences (Maximum Likelihood) with molecular data also recovered similar results and with higher branch support, but consistently failed to recover clad A. Pine et al. (2012) also included a Bayesian analysis of the combined dataset (morphology and molecular datasets, with an additional gene). They found the four clades (A-D) with Mindomys hammondii appearing within the species of the genus Euryoryzomys, questioning the monophyly of the latter, and Melanomys also appearing to be non-monophyletic. Their maximum likelihood tree was similar to previous ones, but the parsimony analysis produced a very different topology, with lower levels of support and not recovering clades C and D. Fabre et al. (2012), using a dataset assembled from published sequences similar to Pine et al. (2012), obtained a very similar tree, with minimal differences (“Handleyomys infectus” recovered as sister to Nephelomys albigarilis).

In Schenk et al. (2013) analyses with GHR and RAG1 genes recovered the genera Scolomys and Zygodontomys separated, while the gene IRBP recovered a polytomy on this section of the tree, and only with CYTB they composed a clad. The concatenated six gene supermatrix assembled by Steppan and Schenk (2017), analysed with maximum likelihood and Bayesian inference, recovered a monophyletic tribe Orzyomynini, with the three clades B, C, and D and with Zygodontomys arising from the basal split within the tribe, and Scolomys as sister to the remaining three clades. The supermatrix employed by Maestri et al. (2017) recovered similar results to Steppan and Schenk (2017), apart from the sister relationship that was obtained between Scolomys and Zygodontomys.

In summary, using a genomic dataset (over 400 loci) and employing both concatenated and coalescent frameworks, we were able here to present a robust phylogenetic hypothesis for Orzyomynini, supporting our understanding of the main clades, their interrelationships and minimizing many of those uncertainties on generic relationships presented above (as showed by the high support values).

4.1.2. Taxonomic implications

Despite the consistency of support from molecules and morphology (data from Weksler, 2006; Percequillo et al., 2011b) for the clades B, C and D and their geographic coherence, we believe that nomenclatural acts, such as the erection of new formal names for clades B, C and D, as subtribes, are unnecessary and would be inappropriate, leading to needless taxonomic inflation. Nevertheless, some sections of the inferred phylogenetic tree relating to genera Hylaeamys, Neacomys and “Handleyomys” merit taxonomic comment.

It is noteworthy that the length of the branch separating H. yunganus from other species of Hylaeamys is similar to that separating species of the genus Oecomys from species of the clad of Euryoryzomys and Transandinomys, and that between these last two genera. Musser et al. (1998) and Weksler (2006) considered the “yunganus group” (including the formerly Oryzomys yunganus and O. tatei [not included in this study]) as a separate group, distinct from the “capito” or “megacephalus” group. Most external and cranial traits are similar between species of the genus Hylaeamys, apart from the most obvious character that discriminates H. yunganus and H. tatei from other species of the genus, namely the presence of two enamel islands on M2 and the shorter and narrower incisive foramina (see Musser et al., 1998; Weksler, 2006). Other evidence that would sustain the recognition of H. yunganus in a distinct genus is the sympatry between H. yunganus and H. megacephalus on eastern Amazon and to H. perennis in western Amazon, an unusual pattern given that most sister species of sigmodontines exhibit allopatric and/or parapatric, and rarely sympatric, distributions (see Patton et al., 2015; Prado and Percequillo, 2013). A more comprehensive and detailed analysis of the genetic and phenotypic variation of these taxa will allow more conclusive taxonomic decisions.

Our trees (Fig. 1) include one unnamed form of the genus Neacomys from Rio Xingu, Central Brazilian Amazon (Neacomys sp.). Understanding of the diversity of this genus has changed in the last two decades, with descriptions of new species by Hurtado and Pacheco (2017), Patton et al. (2000), Sánchez-Vendízú et al. (2018), Semedo et al. (2020), and Voss et al. (2001). These efforts were geographically focused revisions of species groups within Neacomys, only partially addressing its diversity and complexity. The most recent study (Semedo et al. 2020) analyzed the species diversity along the Central Brazilian Amazon and the authors described three new species, but they did not include samples associated to our Neacomys sp. This taxon might be conspecific with N. vargallosai, but comparisons of CYTB gene sequences (results not shown) reveal a high level of genetic divergence (over 7%); hence we consider it provisionally as a separate entity pending further, more detailed, analysis.

It is also important to clarify our use of the name “Handleyomys”. We
were not able to include samples of *H. infectus* and *H. fuscatus* (the type species of the genus *Handleyomys*, originally included by Voss et al. (2002) in his description of the genus. The inclusion of species of *alforoi* group (including *H. alforoi*, *H. melanotis*, *H. rostratus*, *H. chapmani*, *H. rhabdops*, *H. saturatior* and *H. guerrerensis*; sensu Ramirez-Pulido et al., 2014) within the limits of the genus was provisionally suggested by Wexels et al. (2006), until the description of a new genus to include these species. In 2015, the name “Handleyomys” was used by Gómez-Laverde et al. (2015) and we also employed the same arrangement here, as we lack a new generic name for this taxon.

### 4.2. Tempo and mode: Biogeographic models

In the Andes analysis, in addition to its better statistical fit to the data, the DELIKE model also seems to represent a reasonable biological explanation of the ancestral ranges history based on our knowledge of the system. We have three geographical areas with large extensions, and each one of these areas included the range of several species, especially in the east of Andes. This characteristic allowed different biogeographic scenarios, of dispersal, extinction and vicariance, events that are favored on DELIKE model.

The founder-events might be explained by (1) our incomplete sampling on genus and species level, and (2) the putative extinction of lineages from intermediate areas during or after the cladogenesis events. Regarding (1), the absence of true *Handleyomys* and other species of “Handleyomys” and *Nephelomys* might have affected the recovery of ancestral areas, given that species of these genera are widespread in PAC, BBRA, SBRA and SAZ. In clade C, species of *Oligoryzomys* are widespread in the entire continent, and the genus *Neacomys* is far more diverse and more widely distributed than the representatives included in our analysis, especially in Boreal Brazilian, Southern Brazilian and Pacific areas (see Hurtado and Pacheco, 2017). It is likely that with a complete sampling, with species occurring in adjacent areas, analyses would recover jumps between contiguous areas. Regarding clade D, the absence of *Amphicatomyx* and other species of *Nectomys*, that occur in the BBRA, SBRA, AMAZ and PAC might explain the jump from the CHA/PAR area to the PAC dominion. Considering (2), jumps between areas that are not adjacent, might reflect the extinction of lineages (that may have played an important role on the diversification of murid rodents; Steppan and Schenk, 2017) in the intermediate areas, for instance in the geographical area between BBRA/SBRA (the area of the ancestor of clade formed by *Nephelomys*, *Mindomys*, *Hylaeamys*, *Transandinomys*, *Oecomys* and *Euryoryzomys*) and MES (the area of “Handleyomys”).

### 4.3. Tempo and Mode: The origin of tribe Oryzomyini and diversification its major clades

 Tribe: our results suggest the origin of the tribe Oryzomyini around 6.88 mya in an area east of Andes (Fig. 2), likely in Amazonia, as the most probable state at this node was recovered for an area including BBRA and SBRA dominions (Table S4, Supplementary Material). This is similar to the potential AOD for the tribe found by Prado and Percequillo (2013), and other studies (Gonçalves et al., 2020; Leite et al., 2014; Parada et al., 2013), with dates ranging from 9.64 to 5.1 mya (BEAST) or even 11.2 to 5.5 mya (Leite et al. 2014; Multivitime).

The origin of the tribe Oryzomyini, occurred during the Huayquerian South American Land Mammal Age (ALMA; Flynn and Swisher III, 1995). At this moment the Amazon basin was being formed, with the Acre System dominating the landscape (Hoorn et al., 2010; Wesselingh et al., 2010), and diverse rainforests dominating western Amazon. *Scolomys*, the first genus to differentiate within the tribe, is currently distributed in western Amazon lowland forests, between the Rio Jurua and Rio Japura/Rio Caquetá. Although we did not include the genus *Zygodontomys* in our analysis, this genus is probably the next lineage to differentiate within the tribe. In the Late Miocene, the Vaupés arch was also being formed, separating the “Orinoco-Amazonas divide”: species of the genus *Zygodontomys* are widespread in the open habitats in northern South America (lowlands, highlands and the Andes) and are limited to the north bank of Rio Caquetá in Colombia. Leite et al. (2014:7) suggested that a vicariant event split the “ancestor of both *Scolomys* and *Zygodontomys* in Amazonia” from other Oryzomyini, a plausible explanation not fully tested in our analysis.

Clades B, C and D: Our data and analyses support the inference that the most recent common ancestor of clades B, C and D (Figs. 2 and 3) originated in the late Miocene, around 5.5 mya, and the ancestral area for this node located at the BBRA, SBRA and AMAZ dominions. This result is largely concordant with the reconstruction of Schenk and Steppan (2018). The origin of the three major clades (B, C and D) occurred during the Montehermosan, Chapadmalalan and Uquian SAL-MAs, from early to late Pliocene (Table 1).

Clade B: The origin of clade B is inferred to have happened about 4.51 mya and to be located more likely on the east of Andes (Fig. 2), with an ancestral area in BBRA, SBRA and AMAZ, therefore in the Amazon region (Fig. 3). Within this clade there were movements from i) east to west, with three independent invasions of Andes and west of Andes in Late Pliocene about 2.43 mya (*Transandinomys*) and in early Pleistocene around 2 mya (“*Handleyomys*” and the ancestor of *Nephelomys/Mindomys*), including entrance in Central and North America. There were also ii) invasions of eastern South America during Pleistocene (from 1.91 to 1.03 mya), probably via the Central and Northeastern portion of Brazil, as species of *Hylaemys* occur in Cerrado (*H. megacephalus*) and Atlantic Forest (*H. oniscus* and *H. seuneni*) as well as *E. russatus* that also inhabits the northeastern portion of Atlantic Forest. Later, there were movements from iii) east to west, as lineages of *Euryoryzomys* and *Hylaemys* subsequently invaded Central (*E. emmonsae*) and Western (*E. nitidus* and *H. acrurus*) Amazon in middle Pleistocene around 1.03/1.05 mya, from an eastern source, likely from the central portion of Brazil, in the Atlantic Forest/Cerrado/Eastern Amazon connection (Costa, 2003). Finally, there was also a (iv) generalized dispersal of *Oecomys* to eastern and western areas, all of them from of the ancestral “core” Amazon area, likely from northeastern, central and southern routes (Batilha-Filho et al., 2013; Costa, 2003) in early Pleistocene.

Clade C and D: The divergence of clades C and D is estimated at 4.9 mya, during the late Pliocene. The ancestral range was on the east of Andes region (Fig. 2), but no ancestral area within this region was estimated with high probability on Dominon analysis. Therefore, an unknown event separated the ancestors of clades C and D, and the former lineage started its diversification in the middle Pleistocene, around 3.67 mya, during the Uquian in the east of Andes realm, likely on the Pacific region, with 30.4% probability of occupancy. In clade C, DEC and
BAYAREALIKE recovered a northern east of Andes origin, on highland and lowland habitats, with i) independent invasions of Amazonian lowlands by Neacomyos in early Pleistocene and Oligoryzomys from early Pliocene (separation of the ramus leading to Oligoryzomys from other genera) to Pleistocene, and eastern portion of South America (Oligoryzomys), likely via the northeastern portion of Amazon, as the ancestor of the clade was in this area. ii) Later there was the invasion and subsequent diversification of Microzyomys and Orelzymyines, likely following the formation of the montane Andean habitats in the early Pleistocene, e.g. the Paramo (Hoorn et al., 2010; Wesselingh et al., 2010), one major habitat in which those species currently occur (Carlton, 2015; Percequillo, 2015). Thus, in this clade, there was not an invasion of the northern Andes, but inversely, a colonization from northwestern South America towards eastern areas, more similar to the ideas proposed by (Reig, 1986) based on the AOD of the tribe.

Clade D: All major clades have overlapping distributions in South America (see Prado and Percequillo, 2013), although clade D is more diverse and geographically widespread in west of Andes areas than other clades. This is the group with more genera and a more complex pattern of distribution (Prado and Percequillo, 2015) and its origin is inferred to be in the early Pliocene, from 3.12 to 4.09 mya, on the Monterhermosan SALMA, in the east of Andes region (Fig. 2), most likely in the CHA and PAR areas on eastern South America (Fig. 3: 35.6% of probability of occupancy). As in other clades, models recovered important biogeographic events. (1) The basal split within clade D separates one east of Andes genus (Drymoromys) and one Andean genus (Eromomyines) from the other members of this clade. Their ancestor was hypothesized to have occurred in the PAR area east of the Andes, with a jump dispersal to the Andes (Andes) and to the PAC area (Dominion) of E. polius, about 2.2 mya. The common ancestor of the remaining clade D genera is inferred to have originated about 4 mya in eastern South America, in the CHA and PAR areas, and within this clade there was a dominance of east of Andes groups (with no important vicariant or dispersal events in Andes analysis), namely Lundayomys, Sooretamys, Cerradomys, Pseudoryzomys, and Holochilus. The generic diversification of all these lineages is inferred to have happened during a comparatively short time, from 4 to 3 mya, in eastern South America, with groups associated exclusively with Atlantic Forest (Lundayomys and Sooretamys), with Cerrado (Cerradomys and Pseudoryzomys), or with both (Holochilus).

Within this clade, the models recovered i) multiple invasions of the Andes by dispersal, from eastern (E. polius, in late Pliocene; N. apicalis in mid Pleistocene) and western (species of Tamuromys, Melanomys and Aegialomys, almost synchronically on late Pleistocene) stocks, and ii) one major vicariant event on the West of the Andes during the transition of Plio/Pleistocene. Given that the main events of the orogenesis of the Andean Cordillera were nearly complete by the late Pliocene/early Pleistocene (Hoorn et al., 2010), the inferred vicariance implies some other geological/climatological event must have occurred in this area. With an ancestor located on Central/Eastern South America, the major trends were independent movements towards west, ili-v) to the transitional areas between Amazon and Cerrado, by Pseudoryzomys and Cerradomys almost simultaneously in early Pleistocene and to western Amazon by species of Necromys in late Pleistocene; and to vi) northern South America, by the ancestor of the clade Necromys-Aegialomys likely through southwestern Amazon, Central Andes and then Northern Andes, in late Pliocene. This is a well-known dispersal route (Costa, 2003) and an area with closely related taxa as Drymoreomys and Eremoryzomys, and other taxa, as Sooretamys (Percequillo et al., 2017; 2011a, b). From northern South America, there was the vii) colonization of Central and North America by the ancestor of genus Orzyomys in mid Pleistocene, with subsequent re-invasion of South America by O. coesi; viii) the dispersal to the Galapagos archipelago by Nesoryzomys in mid Pleistocene (likely accompanied by Aegialomys galapagoensis, see Prado and Percequillo, 2018); and ix) the dispersal of Aegialomys ica to the semiarid areas and deserts of Southern Peru, in late Pleistocene.

4.4. Tempo and Mode: Orzyomynini and the South American biogeographic history

The several invasions of the Andes and of the west of the Andes were likely associated with the Huancabamba Depression, a region of deep valleys in northern Peru, known as an important barrier to Andean Flora and Fauna from north and south of the depression, but also as a region of dispersal across the Cordillera (Duellman, 1979; Weigend, 2002). It is probable that around the late Pliocene-early Pleistocene, the timing of several dispersal events for orzyomynines, the Depression was at lower altitudes than today, as it only attained its present altitude on the Pleistocene (Duellman, 1979).

The origin of the tribe in mid Miocene, likely east of the Andes in South America, strengthens the hypothesis of an early arrival of the ancestor of Sigmodontinae, as originally defended by Reig (1984) and Hershkovitz (1969, 1966), and recently advocated by Bacon et al. (2015) and Jaramillo et al. (2017). These last authors claimed that the biotic interchange between continents began on Oligocene-Miocene, before the emergence of the Isthmus of Panama; the origin of the Isthmus is also under dispute, with authors claiming early (Bacon et al., 2015; Montes et al., 2015) or late uplift (O’Dea et al., 2016). Undisputed is the fact that dispersals within the tribe Orzyomynini from South to Central and North America, and vice-versa, occurred after the formation of the Isthmus of the Panama, as traditionally established around 3–2.8 Mya (Raskin, 1978; Leigh et al., 2014; Lession, 2015, 2008; O’Dea et al., 2016; Patterson and Pascual, 1972; Simpson, 1983).

Although there is some level of overlap on the geographic distribution of genera and species between the major clades (see Fig. 3; see also Prado and Percequillo, 2013), their histories are distinct in time and space: with the ancestors of clades B and D originating during a similar time period, while the ancestor of clade C is almost 1 mya younger (Table 1). They also originated in different areas/biomes/habitats of the continent, with vicariant and dispersal events occurring in different temporal moments and in distinct directions. It is noteworthy that the clades B and D originated earlier and are more diverse and more widespread geographically and ecologically than the more recent clade C.

It is likely that when the ancestors of clades B and D colonized the Amazon and Cerrado/Atlantic Forest, respectively, they dispersed and speciated allopatrically within those areas, benefiting from the ecological opportunity of new available niches aint these areas (Schenk and Steppan, 2018). In fact, species of clade B are much more diverse in Amazon (where only species of the clade D genus Necromys occur; most other species of this clade are distributed peripherally to the Amazon; see Prado and Percequillo, 2013), while species of clade D are more diverse in Atlantic Forest and Cerrado, and also in Andes and west of Andes habitats (areas in which the clade B is not very speciose).

The ancestor of clade C is more recent and most of the habitats were already occupied by other lineages of Oryzomyini when it dispersed to Andes and east of Andes areas, so that competitive exclusion limited its diversification (Schenk and Steppan, 2018; Schluter, 2000). Schenk and Steppan (2018: 416) predicted that “species within regions would be more closely related to each other… than to species in other regions”, and at least for the early diversification of the tribe, this seems to be what happened, with early species of clades B, C and D, occupying definite and distinct areas within South America.

Within Orzyomynini, clade B exhibits less diversity at the genus level and within diversity at the species level, in comparison with lineages of clade D, in which there are several monotypic genera (clade C is less diverse at both ranks; see Appendix A, Supplementary Material). One hypothesis to explain this pattern is extinction events. With the retraction of forests in South America in the Holocene (Vivo and Carmignotto, 2004), sister taxa to these lineages that inhabited the forests of central portions of South America (presently dominated by open landscapes) may have become extinct. Alternatively, the species of these clades may have experienced distinctive rates of diversification and experienced an
adaptive radiation as they “colonized” the new habitats, reaching more ecological and phenotypic diversity (Steppan and Schenk, 2017; for an opposite view, see Maestri et al., 2017).

The evolutionary history of the tribe Oryzomyini is deeply associated with the east of Andes portion of the South American continent, more precisely within the Amazon region, and we believe that this is the AOD for the tribe. Our models recovered domicinations on this area with high probability, and that genera have inferred ancestrial distributions in different areas of the continent, depending upon the clade to which they belong agrees with studies based on patterns of distribution (e.g., Prado and Percequillo, 2013). Ancestral range was ambiguous in Leite et al. (2014) and Maestri et al. (2019) also obtained the Amazon (an area equivalent to Morrone’s SBRA) as a likely candidate as the ancestral area of diversification of Oryzomyini. Schenkel and Steppan (2018) demonstrated that Oryzomyini is the only sigmodontine tribe whose ancestral area is the Amazon region (similar to our results); all other tribes diversified on other regions of South America, but eventually some of them reached Amazon, as in Thomasomyini (Rhipidomys) and Akodontini (Oxymycterus, Akodon, Necromys). In our hypothesis, the ancestral range of the clade containing Oryzomyi is in South America, west of the Andes, with a subsequent invasion of Central and North America by the ancestor of genus Oryzomys, while Schenkel and Steppan (2018) inferred the ancestor of this clade was in Central and North America, with independent invasions of South America by species of the genera Aegialomys, Melanomys and Sigmodontomys. These differences might be a consequence of the taxonomic sampling (wider in the dataset of Schenkel and Steppan, 2018) and/or the methods employed (BioGeoBEARS in the present approach, S-DIVA, in Schenkel and Steppan, 2018) and remain unresolved.

Understanding which factors explain these differences in diversity and diversification will depend on a yet more comprehensive phylogenetic hypothesis, including more taxonomic diversity, coupled with simulation approaches on geographic distribution and rates of diversification, as well as studies on their energetic balance, ecophysiology and reproductive strategies.

5. Conclusions

Most previous phylogenetic hypotheses for the tribe Oryzomyini were obtained with concatenated matrices of few genes (mainly Cyt b and IRBP) employing Bayesian and ML as optimality criteria. In this contribution, we presented the first genomic dataset generated for a comprehensive diversity of the genus group for the tribe Oryzomyini, employing Next Generation Sequencing. Based on 400 loci and 598,786 sites, we employed concatenated (RAxML) and coalescent (Astral) methods to establish a phylogenomic hypothesis for this tribe, as well as date estimates and historical biogeographic analysis. These phylogenomic analyses corroborated the tribe and the major three clades (B, C and D) as monophyletic, and also revealed novel relationships within Oryzomyini, especially in clades B and D. Although morphology is valuable for species and genus delimitation, its role on the establishment of phylogenetic relationships is not very effective and needs further appraisals, when compared to the genomic data employed here. The origin of the tribe is inferred to be in the Miocene, with genus level cladogenesis in mid and late Pliocene and early Pleistocene and speciation events taking place in the mid and late Pleistocene. The east of Andes region is estimated as the geographic origin for the ancestor of ootyomyines, and the Amazon basin is hypothesized as the original area of diversification of tribe. The tribe Oryzomyini is autochthonous radiation of South America, that subsequently invaded Central and North America multiple times, as well as the Galapagos Archipelago, and dispersal is the main driver of diversification.

Credit authorship contribution statement

Alexandre Reis Percequillo: Conceptualization, Methodology, Investigation, Resources, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. Joyce Rodrigues do Prado: Methodology, Validation, Formal analysis, Data curation, Writing - review & editing, Visualization. Edson Fiedler Abreu: Methodology, Validation, Formal analysis, Data curation, Writing - review & editing, Visualization. Jeronimo Dalapiccola: Methodology, Validation, Formal analysis, Data curation, Writing - review & editing, Visualization. Ana Carolina Pavan: Methodology, Validation, Formal analysis, Data curation, Writing - review & editing, Visualization. Elisandra de Almeida Chiquito: Methodology, Validation, Formal analysis, Data curation, Writing - review & editing, Visualization. Scott J. Steppan: Investigation, Writing - review & editing, Supervision. Alan R. Lemmon: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - review & editing. Emily Moriyarty Lemmon: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - review & editing. Mark Wilkinson: Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We are deeply indebted to curators and support of all collections that loaned tissue samples of most species analyzed in this study: without their generous support this study would not be completed. Several samples were also obtained during field work in Atlantic Forest (Bananal, São Paulo and Murici, Alagoas) and Amazon Forest (Rio Japurá and Rio Xingu, Amazônas; Rio Jufari, Roraima; and Rio Madeira, Amazônas and Rondônia) conducted by AR Percequillo and his team; we are grateful to all field assistants and also to the local people that provided essential assistance and warm hospitality during long weeks away from home. We also would like to thanks to Felipe Grazziotin for his valuable assistance on earlier stages of this project, with important conceptual insights and also with logistic contribution. AR Percequillo would like to acknowledge Mario de Vivo, for his encouragement and support; he also is deeply indebted to the Natural History Museum staff, Roberto Portela Miguez, Louise Tomsett, Richard Sabin and Paula Jenkins for all their kindness and help during his one-year term in the mammal section, in London. We also would like to acknowledge two anonymous reviewers for their contribution to earlier versions of this contribution: their dedication and time greatly improved the manuscript.

Funding

This work, both in the field and lab, was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, grant numbers 09/16009-1, 16/20055-1) and Conselho Nacional de Pesquisa e Desenvolvimento Científico (CNPq, grant numbers 477842/2004-6 and 304156/2019-1).

Appendix A

The tribe currently harbors 30 genera and 147 extant species (Burgin et al., 2018). Genera Scelomys and Zygodontomys presents two species. Considering all taxa recognized as valid (see Patton et al., 2015; Burgin et al., 2018), clade B currently harbors 8 genera and 55 species, the most speciose radiation within the tribe. Based on current data (Patton et al., 2015; Burgin et al., 2018; Schenkel and Steppan, 2018), this is the smaller clade within the tribe in number of genera, 5 (including Microakodontomys), and 36 species. Clade D exhibits the largest diversity on
generic forms, 15, and is almost as diverse as clad B, with 52 species, considering the number of currently recognized taxa (Patton et al., 2015; Burgin et al., 2018).

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2021.107120.

References


Molecular Phylogenetics and Evolution 159 (2021) 107120

A.R. Percequillo et al.


