University of São Paulo
"Luiz de Queiroz" College of Agriculture
Center for Nuclear Energy in Agriculture

Integrative taxonomy of the genus *Proechimys* (Rodentia: Echimyidae) from Western Amazon

Jeronymo Dalapicolla

Thesis presented to obtain the degree of Doctor in Science. Area: Applied Ecology

Piracicaba
2019

Jeronymo Dalapicolla
Bachelor in Biological Sciences

Integrative taxonomy of the genus *Proechimys* (Rodentia: Echimyidae) from Western Amazon

Advisor:
Prof. Dr. **ALEXANDRE REIS PERCEQUILLO**

Thesis presented to obtain the degree of Doctor in Science.
Area: Applied Ecology

Piracicaba
2019

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Dalapicolla, Jeronymo

Integrative taxonomy of the genus Proechimys (Rodentia: Echimyidae) from Western Amazon / Jeronymo Dalapicolla - - Piracicaba, 2019.

328 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”. Centro de Energia Nuclear na Agricultura.

1. Biogeografia 2. Delimitação de espécies 3. Eumysopinae 4. Filogeografia 5. Morfometria I. Título

In memory of Thiony Emanuel Simon, whose teachings and love for science I will always carry with me.

In memory of Christian J. Brown, who taught me so much about life in so little time.

To the 1,526 LGBT+ Brazilians who were murdered or committed suicide due to societal prejudice and hatred between 2015 and 2018 while I developed this thesis.

ACKNOWLEDGMENTS

This thesis was conceived in 52 months and each page was built with the help of many people, from the anonymous Brazilian citizens who paid their taxes and financed scientific research in Brazil, to my advisor who helped conceive the idea and gave logistical support for this work. Unfortunately, there are too many names to remember – if yours is not here, it is not due to ingratitude but rather a failing memory associated with the inherent lack of time and sleep at the end of a Ph.D. Thank you all!

To my family, especially my mother and sister, who, while not even fully understanding the Ph.D., have supported me in this decision. The distance and homesickness were tough, however we handled all the difficulties well, right?

To Alexandre Reis Percequillo, my advisor, who accepted me as a student even though he did not know me. Thank you for giving me the opportunity to work and study in your lab, to know the Amazon, and for believing in my potential. I hope I have not disappointed you. You have helped me become the professional I am today.

To Lacey Knowles that allowed me to spend a year in her lab at the University of Michigan, sharing her knowledge and especially challenging me to think about science in a different way. Thank you for opening the doors of your lab, and also your home, to another Brazilian. My life has changed completely after my stay in Ann Arbor.

To the Programa Pós-Graduação Interunidades em Ecologia Aplicada (PPGI-EA), my graduation program, to the Escola Superior de Agricultura “Luiz de Queiroz” (ESALQ), to the University of São Paulo (USP), to the Ecology and Evolutionary Biology Department (EEB), and to the University of Michigan (UM) and their employees and professors for providing me an environment conducive to my growth as a scientist and citizen. A special thank you to Mara from PPGI who always helped by replying to my emails about the procedures of the PPGI, or engaging in nice conversations about life. And to Maria Victoria “Vicky” Ballester, for always enthusiastically hosting the biodiversity projects in the PPGI-EA, for helping us think more broadly about them, and of course by agreeing to be responsible for my scholarship during Percequillo’s postdoc. I hope I have not given you too much work in that time.

To James “Jim” L. Patton, who has always been a resource to me since my undergraduate studies and who I have had the opportunity to meet personally. Thank you for providing data and photos for the project, and for your tips on the spiny rats and how to finish a Ph.D. without hurry. I think I should listened more closely to that last part.

To my Ph.D. Monitoring Committee formed by Giancarlo Xavier, Gilson Iack-Xinemes, Katia Ferraz, and Yuri Leite, and to my Qualification Exam Committee formed by Ana Paula Carmignotto and Reinaldo de Brito, for helping to delineate and improve the project. Undeniably, you were essential to the quality of this thesis. Thank you also to the professors and members of my Ph.D. Committee, for evaluating and helping to improve this four-years’ work, as well as my professional performance.

To museums and mammal collections from Brazil, USA and Europe and their curators for allowing me access to the specimens. This thesis would not exist without you! Thank you all so much! Namely: Ismael de Jesus, Juliana Gualda, Mario de Vivo, Luiz Fábio Silveira (MZUSP); Miguel T. Rodrigues, Sabrina Baroni (CIT); Yuri Leite, Leonora Costa, Monique Nascimento (UFES); Fernando Perini, Rafaela Missagia, Rodolfo Stumpp (UFMG); Claudia Costa (PUC-Minas); Pedro Cordeiro-Estrela, Alfredo Langguth (UFPB); João de Oliveira, Marcelo Weksler (MN); Alexandra Bezerra, José de Sousa e Silva Junior (MPEG); Priscilla Tucker, Cody Thompson (UMMZ); Eleanor Hoeger, Robert Voss, Svetlana Katanova (AMNH); Alfred Gardner, Darrin Lunde, Indrid Rochon, Louise Emmons (NMNH); Mark Omura (MCZ); Chris Conroy, Eileen Lacey, James Patton (MVZ); Lauren Smith, Bruce Patterson (FMNH); Burton Lim, Jacqueline Miller (ROM); Mark Wilkinson, Roberto Miguez (NHM); Kathy MacDonald, Caleb Phillips (TTU); Mariel Campbell, Jonathan Dunnum, Joseph Cook (MSB).

To the researchers who shared unpublished information on the spiny rats or samples that were important for this work: Mariana Machado, Camila Moreira, Ana Paula Carmignotto, Yuri Leite, Leonora Costa, James Patton, Miguel T. Rodrigues, and Pedro Peloso.

To friends who helped me find a place to stay during my visits to mammal collections in Brazil and the USA: Bárbara Costa (Tafinha), Dani Rossini, Carol Loss (São Paulo); Guilherme Garbino, Rodolfo Stumpp (Belo Horizonte); Pamella Brennand (João Pessoa); Alan Braz, Edú Guerra (Rio de Janeiro); Gabriela Gonçalves, Pablo Cerqueira, Silvia

Pavan, Pedro Peloso (Belém); Ana Pavan, Hilario Seibel, Ricardo Boechat, Veruska S. Boechat (New York); Silvia Pavan, Pedro Peloso (Washington, D.C.); Gustavo Libardi (Berkeley); Chris Nordhoff, Emilio Munoz, Jose Wellington Santos (Wly) (Chicago).

To my friends who, even with the distance, remained close and helped in various ways during these four years – friends in Vitória, Vila Velha, São Paulo, Piracicaba, Campinas, Ann Arbor, Chicago or Tuscon. If I wrote details about everyone and how you all helped me, it would add another volume to this thesis. However, I must give an honorable mention to those who were my roommates: Pamella Brennand (Pam), Cecilia Kruszynski (Xili), Lethicia Corniani (Máfia), Paulo Garcia (Abili), and Victor Mello in Piracicaba, and Andrea Thomaz (Déa), Joyce do Prado (100), Renata Pirani, and Fernanda Zaidan in Ann Arbor. Thank you for sharing great moments with me that I will treasure forever. To my friends who helped to build this thesis with ideas, or by teaching me the genomic, bioinformatics, statistics, and programming techniques: Bárbara (Tafinha), Dani, Carol, Ewerton (Tom), Aninha, Luciana, Joyce, Déa, Mariah, Cecília, Renata, Sarp, Giorgia, Sasha, Rachael, and Tao. A special thank you to Chuky, even when you did not use two hands for driving! To my Lab friends in Piracicaba: Ana Pavan (Aninha), Edson (Gaúcho), Paulo (Roth), Ricardo (Saldanha), Gustavo (Napister), Elisandra, Pam, Joyce, Edgar, Gerson, Denise, Rodrigo, Alexandre, and Prof. Jaime, who endured my daily optimism about life for four years. Life hurts! I will always take all you with me in my heart.

To all “useful idiots and imbeciles” from Brazil, whose teachings and examples inspired me. Perhaps someday we will get to know the molecular formula for water. To all #Elenão #NotHim community!

To the funding sources of the project, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and American Museum of Natural History (AMNH) for the initial scholarship and grant, and a special thanks to Fundação de Amparo à Pesquisa de São Paulo (FAPESP) for financing this thesis posteriorly through two projects: 2015/02853-6 and 2016/24464-4.

Thank you and Go Blue!

You know nothing, Jon Snow!

Ygritte, in “A Storm of Swords”,
A Song of Ice and Fire by G. R. R. Martin.

CONTENTS

RESUMO

Taxonomia Integrativa do gênero *Proechimys* (Rodentia: Echimyidae) da Amazônia Ocidental

Proechimys é um gênero da família Echimyidae com ampla distribuição na região Neotropical. Embora seja abundante e amplamente distribuído, este gênero é pouco estudado e tem sua taxonomia e sistemática pouco resolvida. Essa falta de conhecimento dificulta estudos em outras áreas, especialmente em ecologia, uma vez que as espécies são externamente semelhantes, tornando complexa a identificação de indivíduos no campo e nos museus. Existem poucos estudos publicados com variação morfológica e genética, e também sobre a sistemática de *Proechimys*. O objetivo desta tese foi propor uma filogenia baseada em dados genômicos e delimitar as espécies do gênero, utilizando também outros bancos de dados como o morfométrico, para entender melhor a diversificação e evolução de *Proechimys*, especialmente na Amazônia Ocidental. Durante o projeto de doutorado, eu identifiquei em morfotipos e fiz o georreferenciamento das localidades de 3.104 espécimes de *Proechimys* em 18 museus e coleções do Brasil, Estados Unidos da América e Inglaterra, e avaliei a variação morfológica de 22 caracteres quantitativos em 1.503 espécimes, e 58 caracteres qualitativos de crânio e pele em 315 espécimes. Nesta tese eu vou apresentar os resultados oriundos de uma parte dos dados morfométricos, de 479 espécimes adultos. Os demais dados morfométricos e morfológicos ainda estão sendo trabalhados para futuras publicações. Além disso, sequenciei parte do genoma de 278 espécimes usando a técnica ddRAD-seq para avaliar a variação genética. O Capítulo 1 diz respeito a uma introdução geral na qual eu apresentei a história taxonômica do gênero, o conhecimento atual sobre a história evolutiva de *Proechimys* e fiz um breve comentário sobre a evolução da paisagem da Amazônia Ocidental. Além disso, eu discuti sobre alguns conceitos de espécies e sobre a Taxonomia Integrativa, temas que foram abordados nessa tese. No Capítulo 2, propus uma filogenia para o gênero baseada em dados genômicos, identifiquei os clados e testei se eles poderiam ser considerados espécies diferentes com base no modelo coalescente multiespecífico para dados genéticos e no movimento Browniano para dados morfométricos. As relações filogenéticas recuperadas entre os indivíduos de *Proechimys* recuperaram cinco grandes clados dentro do gênero, com suporte estatístico para o reconhecimento a nível de espécie de pelo menos 28 linhagens. *Proechimys* não foi recuperado como monofilético e 12 das 28 linhagens não corresponderam a espécies válidas atualmente; algumas delas podem ser novos táxons, enquanto outras podem ser revalidações de táxons atualmente incluídas na sinonímia de espécies válidas. No Capítulo 3, eu datei os tempos de divergência entre os clados e testei modelos bayesianos de evolução da distribuição geográfica para estimar as áreas ancestrais dos clados. Com esses resultados, eu criei uma hipótese biogeográfica para a evolução do gênero. A origem do gênero foi estimada para o Mioceno, na Amazônia Ocidental e foi possível associar a história evolutiva do gênero com mudanças na paisagem da Amazônia. A diversificação dentro dos 5 principais clados do gênero ocorreu no Plioceno e no Pleistoceno. No Capítulo 4, eu avaliei o padrão filogeográfico de três espécies simpátricas de *Proechimys* da Amazônia Ocidental: *Proechimys brevicauda*, *Proechimys simonsi* e *Proechimys steerei*. O objetivo foi testar se essas espécies que compartilham o mesmo espaço geográfico, compartilhariam também os mesmos padrões de estruturação genética, ou se haveria uma segregação ao nível de micro-habitat que levaria a diferentes padrões filogeográficos. Para isso eu calculei a sobreposição dos hipervolumes ambiental e morfológicos e testei a importância de barreiras, da distância geográfica e ambiental entre as populações de cada espécie para explicar a estruturação genética. Cada uma das três espécies simpátricas mostrou pouca sobreposição do hipervolume morfológico, e grande sobreposição no ambiental. Cada espécie teve um padrão de estruturação genética diferente, mostrando que mesmo ocorrendo em simpatria e sendo espécies congêneras, elas respondem de

formas diferentes à evolução da paisagem e às mudanças ambientais. No Capítulo 5, eu apresento uma síntese com as principais implicações dessa tese para diferentes áreas relacionadas à Ecologia Aplicada e as perceptivas futuras sobre o estudo do gênero. Dessa forma, esta tese ampliou o conhecimento sobre os fatores que levam à estruturação genética de espécies de mamíferos da Amazônia, bem como sobre a história evolutiva do gênero *Proechimys* e de sua diversidade genética, morfológica e de espeċies. Estes resultados podem auxiliar trabalhos em ecologia, biologia da conservação e também nos levantamentos de fauna em grande parte da região Neotropical.

Palavras-chave: Biogeografia; Delimitação de espécies; Eumysopinae; Filogeografia; Morfometria

ABSTRACT

Integrative taxonomy of the genus *Proechimys* (Rodentia: Echimyidae) from Western Amazon

Proechimys is a genus of the family Echimyidae with wide distribution in the Neotropical region. Although it is abundant and widely distributed, this genus is little studied and has its taxonomy and systematics little solved. This lack of knowledge impairs studies in other areas, especially in ecology, since the species are externally similar, becoming the identification in the field and museums more difficult. There are few published studies with morphological and genetic variation, and on systematics of the genus *Proechimys*. The aim of this Thesis was to propose a phylogeny based on genomic data and to delimit the species of the genus, also using other dataset such as the morphometric, to better understand the diversification and evolution of *Proechimys*, especially in the Western Amazon. During the Ph.D. project, I identified in morphotypes and geo-referenced the localities of 3,104 *Proechimys* specimens in 18 museums and collections in Brazil, the USA and England, and evaluated the morphological variation from 22 quantitative characters in 1,503 specimens, and 58 qualitative characters of skull and skin in 315 specimens. In this thesis I will present the morphometric results of 479 adult specimens. The remaining morphometric and morphological data are still being studied for future publications. In addition, I sequenced part of the genome of 278 specimens using the ddRAD-seq technique to evaluate the genetic variation. Chapter 1 is a general introduction in which I presented the taxonomic history of the genus, the current knowledge on the evolutionary history of *Proechimys*, and remarks on the landscape evolution in Western Amazon. In addition, I discussed on species concepts and Integrative Taxonomy, themes that were addressed in this thesis. In Chapter 2, I proposed a genomic phylogeny based on genomic data, identified the clades and tested whether they could be considered different species based on the multispecies coalescent model for genetic data and Brownian motion for morphometric data. The phylogenetic relationships recovered among *Proechimys* individuals indicated five main clades within the genus, with statistical support for the recognition of at least 28 lineages at the species level. *Proechimys* was not recovered as monophyletic, and 12 of the 28 lineages did not correspond to currently valid species; some of them may be new taxa, while others may be revalidations of taxa currently included in the synonymy of valid species. In Chapter 3, I estimated divergence times between the clades and tested Bayesian models of geographic range evolution to indicate ancestral areas for the clades. With these results I created a biogeographic hypothesis for the evolution of *Proechimys*. The genus origin was estimated in the Miocene and in the Western Amazon. I was able to associate the biogeographic history to the landscape evolution of the Amazon. Diversification within the five main clades occurred in the Pliocene and Pleistocene. In Chapter 4, I evaluated the phylogeographic pattern of three sympatric species of *Proechimys* from the Western Amazon: *Proechimys brevicauda*, *Proechimys simonsi* and *Proechimys steerei*. The aim was to test whether these species sharing the same geographic space would also share the same patterns of genetic structure or whether there would be segregation at the microhabitat level that would lead to different phylogeographic patterns. For this I calculated the overlap and similarity among their environmental and morphological hypervolumes, and tested the importance of barriers, the geographic and environmental distance between populations in each species to explain the genetic structure. Each species showed little overlap of the morphological hypervolume, and great overlap in the environmental one. In addition, they presented different genetic structure patterns, showing that even though they are congeners species and occur in sympatry, they may respond differently to landscape evolution, and to environmental changes. In Chapter 5, I present a synthesis of the main conclusions and future perceptives about the study of genus and the implications of my

results for the conservation and studies on diversity patterns in the Amazon region. Thus, this Thesis increased the knowledge on factors that lead to the genetic structure of mammalian species in the Amazon, as well as on the evolutionary history of the genus *Proechimys* and its genetic, morphological and species diversity. These results may support future studies in ecology, conservation biology and also fauna surveys in the Neotropical region.

Keywords: Biogeography; Species delimitation; Eumysopinae; Phylogeography; Morfometrics

LIST OF ABBREVIATIONS AND ACRONYMS

2n	Diploid number (cytogenetics).
95% C.I.	95% confidence interval.
ABX	Abacaxis River (subclade).
AIC	Akaike information criterion.
AICwt	Akaike weights.
AMNH	American Museum of Natural History, New York, USA.
AMNH-AMCC	Ambrose Monell Cryo Collection, American Museum of Natural History, New York, USA.
AMOVA	Analysis of molecular variance.
AOE	Areas of Endemism.
ATH	Acre-Ta Huamanu rivers (subclade).
AUC	Area under the curve.
AUC_{SD}	Standard deviation of the area under the curve.
AUC_{TEST}	Area under the curve by test samples.
AUC_{TRAIN}	Area under the curve by training samples.
BaL	Basilar length of Hensel.
BAYAREALIKE	BayArea similar model for range evolution,
BBR	Boreal Brazilian dominion (area).
BEL	Belem (area).
Bio	Bioclimatic variable. It is followed by a number, *e.g.* Bio3.
BM	Brownian Motion.
bp	Base pair.
BSC	Biological concept of species.
BUH	Bullar depth.
BuL	Bullar length.

BUM	Upper second molar width.
CA	California.
CAM	Central America-Chocó (subclade).
CD	Cranial depth.
CDM1	Cranial depth at M1.
CER	Cerrado (subclade).
CHA	Chaco dominion (area).
CIL	Condyloincisive length.
CIT	Coleção de Tecidos Miguel Trefaut Rodrigues, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil.
CLJ	Central-Lower Juruá River (subclade).
CM	Coalescent Model.
CMIP5	Coupled Model Intercomparison Project phase 5.
COI	Cytochrome oxidase subunit I.
Cty b	Cytochrome b.
D	Diastema length.
D	Dispersal event. It is followed by a number, *e.g.* D3.
dbRDA	Distance-based redundancy analysis.
DBU	Distance between the bullae.
ddRAD-Seq	Double diggested restricted site-associated DNA sequencing.
DEC	Dispersion-Extinction-Cladogenesis model for range evolution.
DIVALIKE	Dispersal-Vicariance Analysis similar model for range evolution.
DNA	Deoxyribonucleic acid.
DOM	Morrone's biogeographic dominions for the Neotropical region.
dPSC	Diagnostic phylogenetic species concept.
EAM	Eastern Amazon (subclade).
Eco-RI	Restriction endonuclease enzyme.

ECU	Western Ecuador (subclade).
ENM	Ecological niche models.
ESC	Evolutionary species concept.
ESS	Effective sample size.
FC	Feature class.
FMNH	Field Museum of Natural History, Chicago, Illinois, USA.
G	Gamma distribution.
GJR	Galvez-Juruá rivers (subclade).
GLSC	General lineage species concept.
GSL	Greatest length of skull.
GTR	General time reversible model for nucleotide substitution.
GUS	Guiana Shield (subclade).
GUY	Guyana (area).
H	Hinge feature class.
HG	*Hoplomys gymnurus* (clade).
IBB	Isolation by barriers.
IBD	Isolation by distance.
IBE	Isolation by environment.
iBPP	Integrated Bayesian Phylogenetics and Phylogeography (software).
IFL	Length of incisive foramen.
IFW	Maximum width of incisive foramen.
IJM	Içá-Japurá-Madeira rivers (subclade).
IME	Imeri (area).
INA	Inambari (area).
IOC	Interorbital constriction.
IQT	Iquitos (subclade).

IUCN	International Union for Conservation of Nature.
J	Founder-event speciation model.
JAP	Japurá River (subclade).
JAU	Jau (area).
JMI	Juruá-Madeira Interfluve (suclade).
JUR	Juruá River (subclade).
L	Linear feature class.
LGM	Last glacial maximum.
LIG	Last interglacial.
LMD	Length of mandibular diastema.
LMR	Lower Madeira rivers (subclade).
LMUSP	Coleção do Laboratório de Mamíferos da Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, São Paulo, Brazil.
LnL	Likelihood.
LOR	Loreto, Peru (subclade).
LXR	Lower Xingu River (subclade).
MA	Massachusetts.
MAR	Magdalena River (subclade).
MaxB	Maxillary breadth at M2-M3.
MB	Greatest breadth across mastoid.
MCM	Multispecies Coalescent Model.
MCMC	Markov chain Monte Carlo.
MCN-M	Coleção de Mastozoologia do Museu de Ciências Naturais da Pontíficia Universidade Católica de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
MD	Mandibular length.
MES	Mesoamerican dominion (area).
MFW	Mesopterigoid fossa width.

ML	Maximum likelihood.
MMD	Middle Miocene Disruption.
MMR	Madidi-Madre de Dios rivers (subclade).
MN	Museu Nacional da Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.
MPEG	Museu Paraense Emílio Goeldi, Belém, Pará, Brazil.
MPI-ESM-P	Earth system models of the Max-Planck-Institut für Meteorologie in paleo mode.
MRCA	Most recent common ancestor.
MSB	Museum of Southwestern Biology, Alburqueque, New Mexico, USA.
Mse-I	Restriction endonuclease enzyme.
mtDNA	Mitochondrial DNA.
MTR	Mato Grosso region (subclade).
MTRL	Alveolar length of upper molariforms.
MVZ	Museum of Vertebrate Zoology, Berkeley, California, USA.
MZUSP	Museu de Zoologia da Universidade de São Paulo, São Paulo, Brazil.
N	Uncalled bases.
n	Sample size.
NAP	Napo (area).
NAR	North Amazon River (subclade).
NF	Fundamental Number (cytogenetics).
NGS	Next-generation sequencing.
NL	Nasal length.
NMNH	National Museum of Natural History, Washington, D.C., USA.
NOR	North to the Amazon (area).
NRB	Negro River Basin (subclade).
NSM	North Solimões-Marañón rivers (subclade).
NSR	North Solimões River (subclade).

nuDNA	Nuclear DNA.
NVZ	Northwestern Venezuela (subclade).
OccW	Occipital condyle width.
OL	Orbit length.
OTU	Operational Taxonomic Units.
P	Product feature class.
PAB	Pando, Bolivia (subclade).
PAC	Pacific dominion (area).
PAR	Parana dominion (area).
PC	Principal component.
PCA	Principal component analysis.
PCNM	Principal coordinates of neighbourhood matrix.
PCR	Polymerase chain reaction.
PLa	Palatal length A.
PLb	Palatal length B.
PP	Posterior Probability.
PPL	Post-palatal length.
PRO	*Proechimys* specimens, except TEP individuals (clade).
PSC	Phylogenetic species concept.
PTC	Pantanal-Chaco (subclade).
PTI	Purus-Tapajós Interfluve (subclade).
Q	Quadratic feature class.
RAD	Restricted site-associated DNA Sequencing.
RB	Rostral breadth.
RD	Rostral depth.
RL	Rostral length.

RM	Regularization multiplier.
ROM	Royal Ontario Museum, Toronto, Ontario, Canada.
RON	Rondonia (area).
SAR	South Amazon River (subclade).
SBR	South Brazilian dominion (area).
SCE	South Cerrado (subclade).
SEA	South-eastern Amazonian dominion (area).
SNP	Single Nucleotide Polymorphism.
SOU	South to the Amazon basin (area).
SPB	South Peru-Bolivia (subclade).
SSM	South Solimões-Marañón rivers (subclade).
SSR	South Solimões River (subclade).
STR	Santiago River (subclade).
SVZ	Southeastern Venezuela (subclade).
T	Threshold feature class.
TAP	Tapajós (area).
TEP	Tepui (subclade).
TSS	True skill statistics.
TTU	Texas Tech University, Lubbock, Texas, USA.
TXI	Tapajós-Xingu Interfluve (subclade).
UFES-CTA	Coleção de Tecido Animal da Universidade Federal do Espírito Santo, Vitória, Espírito Santo, Brazil.
UFES-MAM	Coleção de Mamíferos da Universidade Federal do Espírito Santo, Vitória, Espírito Santo, Brazil.
UFMG	Coleção de Mamíferos, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.
UFPB	Universidade Federal da Paraíba, João Pessoa, Paraíba, Brazil.

UJR	Upper Juruá River (subclade).
UK	United Kingdom.
UMD	Upper Madre de Dios River (subclade).
UMMZ	University of Michigan Museum of Zoology, Ann Arbor, Michigan, USA.
UMR	Upper Madeira River (subclade).
USA	United States of America.
USC	Unified species concept.
USNM	National Museum of Natural History, Washington, D.C., USA.
V	Vicariance event. It is followed by a number, *e.g.* V3.
VIF	Variance inflation factor.
WAM	Western Amazon (subclade).
XAI	Xingu-Araguaia Interfluve (subclade).
XGU	Xingu (area).
ZB	Zygomatic arch breadth.
ΔAICc	Delta corrected Akaike information criterion.

LIST OF SYMBOLS

D	Jost's coefficient of genetic differentiation.
D	Geographic distance matrix.
E	Environmental distance matrix.
F_{IS}	Wright's inbreeing coefficient.
F_{ST}	F-statistics.
g	Grams.
G	Genetic distance matrix.
G''_{ST}	Hedrick's coefficient of genetic differentiation.
G_{ST}	Nei's coefficient of genetic differentiation.
H_{obs}	Observed heterozygosity.
Ka	Thousands years ago.
Km	Kilometers.
Km^2	Square kilometers.
m	Meters.
Ma	Millions years ago.
mm	Millimeters.
ºC	Degrees Celsius.
t'	Robustness of coefficient of association between two genetic maps.
t''	Robustness of coefficient of association between genetic and geographic maps.
t_0	Coefficient of association between genetic and geographic maps.
θ (theta)	Ancestral population sizes (simulations).
π (pi)	Nucleotide diversity.
τ (tau)	Divergence times (simulations).
Φ_{ST} (phi)	Meirmans' coefficient of genetic differentiation.

1. GENERAL INTRODUCTION

1.1. Integrative Taxonomy

Taxonomy is the science of delineating and classifying the biodiversity based on shared characteristics (Dayrat 2005), and it is a central issue to Biology since it defines the basic unit of its various fields of knowledge (Agapow et al. 2004; Sites and Marshall 2004). Taxonomy is often divided into three levels: alpha, beta and gamma, depending on the hierarchical level analyzed (Schlick-Steiner et al. 2010). Alpha taxonomy is restricted to taxonomic studies at the species level, beta taxonomy at higher categories (*e.g.*, genus, family, order), and gamma at the intraspecific level.

In the last decades, new technologies have emerged, especially in Molecular Biology and Computer Sciences, and new areas of knowledge related to Taxonomy were created within the Biological Sciences, such as Phylogeography (Avise 2009). These new areas and use of molecular markers, and posteriorly Next-Generation Sequencing (NGS) (Ronaghi et al. 1998) attracted more attention of younger researchers, and opened new funding opportunities to the detriment of the classical Taxonomy (Godfray 2002; Dalton 2003; Wheeler 2004). Thus, modern taxonomy has suffered and still suffers from its loss of importance to these new areas, what scientists call the a "Taxonomy Crisis" (Godfray 2002; Tautz et al. 2003; Agnarsson and Kuntner 2007).

Furthermore, the traditional proceedings with morphological data configured a slower and more subjective alpha taxonomy, while the new phylogenetic methods announced that they were faster and more objective (Blaxter 2004; Dayrat 2005). At the beginning of the 2000's the idea of using DNA to identify the biodiversity gained more space (Blaxter 2003; Mallet and Willmott 2003; Tautz et al. 2003), and the most famous of them was the DNA barcoding approach (Hebert and Gregory 2005). In this method the researcher used a genetic marker, the cytochrome oxidase subunit I (COI) for animals, to individual identification at the specific level through inference of phylogenetic trees and genetic distances (Hebert et al. 2003a; b).

However, many studies criticized the using of the DNA Barcodes for taxonomic purposes because gene trees may diverge from species trees for various reasons, among them,

by incomplete lineage sorting (Maddison 1997; DeSalle et al. 2005; Valdecasas et al. 2007). Furthermore, the subjectivity persisted when the researcher defines a species by one or another cladogenesis event from phylogenetic trees or by the percentage of genetic divergence (Johns and Avise 1998; Hebert et al. 2003b; Zachos 2018a). In the end, DNA barcoding approach for taxonomy was the replacement of a morphological method for a non-morphological method with the same problem: a typological, and single-datatype-system approach (Dayrat 2005; Will et al. 2005).

Rather than replacing one data type by another, studies started to integrate both data (Patton et al. 2000; Shaw and Allen 2000; Baker et al. 2003). Dayrat (2005) proposed the term "Integrative Taxonomy", for the new approach that uses multiple data types and additional perspectives, such as, phylogeography, morphology, population genetics, development and others, to delimit the units of life in a multidisciplinary approach (Padial et al. 2010; Schlick-Steiner et al. 2010). Several studies combined different databases for taxonomic purposes since the 1970's, especially on vertebrates (Patton and Gardner 1972; Hillis 1987) and microorganisms (Oren 2004). However, most of them did not integrate the data, they often used concordance of different databases for species delimitation (Yeates et al. 2011; Pante et al. 2015).

Integrative taxonomy has already been applied in the delimitation of species and diversity in plants (Barrett and Freudenstein 2011), invertebrates (Ross et al. 2010) and mammals (Lanzone et al. 2007; Chiquito et al. 2014; Prado and Percequillo 2018). Moreover, new methods of integration of databases data were developed, using simulations and coalescent theory (Yang and Rannala 2010; Fujita et al. 2012), Brownian motion for morphological evolution (Solís-Lemus et al. 2015), and ecological niche models and distributional data (Massatti and Knowles 2016). These new approaches may be useful for species delimitation and may help the Taxonomy to popularize again and leave the crisis in which it finds itself. I employed the assumptions of Integrative Taxonomy in this thesis, the integration of different types of data, and also some of these new statistical approaches to improve the species delimitation of a widely distributed rodent genus in South and Central America.

1.2. Species Concepts

In general, there is an agreement that the species is conceptually a fundamental and natural unit in biology (Sites and Marshall 2004), and is pragmatically delimited based on discontinuities (Ridley 2004). One of the great discussions on the theory of the species concept focuses on two points: (*i*) how to define a species in a broad way that encompasses all forms of life and (*ii*) what level of discontinuity is necessary for a group of individuals to be considered a fully formed species (González 2018).

These questions are expected because the species are the product of an evolutionary process. Evolution is a continuous process, and taxonomy is a discrete science (this group of individuals is a different species or not) (Zachos 2018b). Thus, taxonomy tries to simplify a continuous process in a discrete and binary classification, and consequently it will never match perfectly the ongoing process in nature (O'Hara 1993). To face with this impasse, several species concepts emerged as a response to multiple ways of thinking and defining the fundamental units of Taxonomy (Wilkins 2009).

The most famous of these concepts is the Biological Concept of Species (BSC) proposed by Mayr (1940, 1942). During his career, Mayr changed the BSC definition, sometimes giving emphasis to interbreeding (Mayr 1940), sometimes to reproductive isolation (Mayr 1942), sometimes focusing on the difference of species as a taxon or as a category (Mayr 1969). One of the last definitions for the BSC is that a species is a natural interbreeding community that is reproductively isolated from other communities (Mayr 1992; Beurton 2002). Regardless of the definition used, the central idea of BSC is that flow gene is an important criterion for species delimitation (De Queiroz 2005b). Over time, the BSC was more accepted by zoologists than by botanists, due to the frequent hybridization between plants, suggesting that the development of isolation mechanisms is not the main speciation factor in plants (Mallet 2008). Furthermore, the BSC do not cover all living beings because the criterion of interbreeding or reproductive isolation do not fit in asexual organisms (Ward 1998). Thus, BSC did not satisfy the premise of being broad enough to encompass all forms of life.

Simpson (1961) proposed the Evolutionary Species Concept (ESC), and later Wiley (1978) modified the ESC. An evolutionary species is a lineage (a sequence of ancestral-descendant populations) that evolves separately (maintains its identity) from other species and

has its own evolutionary tendencies and historical fate. ESC added a time dimension that was not present in the BSC (Hull 1997; Mayden 1997; Wilkins 2009). Some critiques have followed the ESC, especially on the definition of some terms as “identity”, “evolutionary tendencies” and “historical fate”. Therefore, many scientists have considered the ESC vague and impractical (Wiley 1978; Mayr 1982).

After the emergence of Phylogenetic Systematics (Hennig 1966), taxonomists and Hennig himself proposed that the classification of organisms should reflect the monophyly between groups, thus added phylogenetic relationships in the taxonomy (Baum 1992; Meier and Willmann 2000). Taxonomists elaborated species concepts that were compatible with phylogenetic systematics, named them as Phylogenetic Species Concept (PSC) (Baum and Donoghue 1995; Mishler and Theriot 2000a). Cracraft (1983) was the first to use the name PSC for species definition although earlier scholars have used the monophyly and other elements of phylogenetic systematics for species definition (Eldredge and Cracraft 1980; Nelson and Platnick 1981; Mishler and Donoghue 1982).

Nowadays there are numerous PSC versions and they fall into two main groups: character-based and history-based PSC’s (Baum and Donoghue 1995). PSC based on history defines species based on phylogenetic relationships, where individuals within the same species are mutually monophyletic (Mishler and Theriot 2000b). As monophyly can occur at any level of the hierarchical categories (above or below the species level), an additional criterion for phylogenetic species definition is usually required, the “ranking criteria” (Baum 1992). The character-based PSC considers that a set of individuals are from the same species when and only if they share a character or a combination of characters among them. This character must be absent from other species, whereas the origin of the character (ancestral or derived) is not important (Baum 1992; Baum and Donoghue 1995). The most applied character-based PSC is the diagnostic Phylogenetic Species Concept (dPSC), where a species is the smallest possible grouping of individual organisms defined by a character in which there is a pattern of ancestry and descendant, which together become basal diagnostic units (Cracraft 1983). dPSC maintains the idea of a species as a lineage but does not retain the phylogenetics aspects because the diagnostic characters may be primitive (plesiomorph or homoplasy) (Mishler and Theriot 2000a; Zachos 2018b).

The dPSC has been widely applied in the mammalian taxonomy, especially in large mammals with the raising of subspecies to species (Groves and Grubb 2011; Zachos 2018a). The "lumpers" (taxonomists who recognize that the similarities between populations are more important than differences, and in favor of synonymizing names) have called this strategy of taxonomic inflation, and the “splitters” (taxonomists who recognize that the differences between populations are more important) have accused the “lumpers” of taxonomic inertia (Frankham et al. 2012; Zachos et al. 2013; Heller et al. 2014). The dPSC is an attractive species concept because it is easily testable. The main criticism to it is that any isolated population, through genetic drift, reaches a level of divergence that makes it diagnosable, and could be a new species (Zachos 2018b). Therefore, although the diagnosis of the species is objective and testable in the dPSC, the level of discontinuity required and to consider individuals as population or species remains subjective.

Although the existing 34 species concepts (Zachos 2016a; Hill 2017; Shanker et al. 2017) exhibit distinct levels of disagreement among them, there is a common point in most species concepts: the general idea that a species is a lineage, in other words, a temporal/spatial sequence of ancestors and descendants populations, which are evolving separately (De Queiroz 1998). This general idea of species is similar on three species concepts: Evolutionary Species Concept (ESC, Wiley 1978), General Lineage Species Concept (GLSC, De Queiroz 1998) and Unified Species Concept (USC, De Queiroz 2005a). Mayden (1997) and De Queiroz (1998, 2007) name these concepts as primitive or ontological concepts, whereas other species concepts, such as BSC and PSC, would be criteria or operational concepts to separate populations in independent lineages (Sites and Marshall 2004). In this perspective, a species is a lineage that is evolving separately from others (ontological concept), and the operational concepts are useful to species delimitation. Thus, the ontological species concepts are broad enough to encompass all forms of lives, but still, the delimitation of a species through discontinuities remains questionable.

The discontinuities are present in all levels in the biodiversity, between species and also among individuals (Bolnick et al. 2011). Intraspecific variation may be due to sexual dimorphism (differences between sex), ontogenetic variation (differences between ages), and geographical variation (differences between localities) (Ridley 2004). Splitting this variation among individuals from a variation at the species level is not a simple task, and it becomes

more complex when the lineages are not completely separated, something quite common since evolution is a continuous process (Zachos 2016c).

When the species are in emergence or in process of speciation, the limits and the amount of discontinuities are still fuzzy, what some scholars call “grey area” (Lee 2003; Zachos 2016b). The largest discussions on the validity of a species are when these taxa are in this grey area (Zachos 2018a). In this context, alternative taxonomic arrangements are created and Taxonomy is considered imprecise (Gippoliti et al. 2018) but often the authors do not realize that imprecision is inherent to Taxonomy (Zachos 2018b). Even with the most sophisticated statistical approach to species delimitation, and with the increasing resolution of the Tree of Life, those grey areas will persist in Taxonomy (Sites and Marshall 2004; Sukumaran and Knowles 2017). With this view, it would be far more intuitive to test whether two lineages have left this grey area than to test whether they have some discontinuity to support them as diagnosable lineages.

Carstens et al. (2013) suggested that researchers should perform an extensive range of species-delimitation analyses with their datasets to test if different lineages are different species. In addition, taxonomic decisions and rearrangements should be conservative, delimiting species only when the results are congruent in the different approaches. If the species boundaries are the same for different methods, and there are no ambiguous results, probably those lineages are not in the grey area. Thus, for these authors, it is better to fail to delimit species than to delimit them when they do not represent separated evolutionary lineages. This proposal is in line with the Integrative Taxonomy, which, in addition to several approaches, encourages the using of several different datasets for the delimitation of taxonomic units (Dayrat 2005).

Here I will employ the ontological concept of species from Simpson (1961) and modified by Wiley (1978), the Evolutionary Species Concept (ESC). Therefore, I agree that species are lineages that maintains their identity (evolving separately from other species), and have their own historical fate (extinction or speciation). I did not choose one of the more modern ontological concepts as GLSC or USC for considering them as modifications from ESC. In addition, I will use the diagnosability, one of the corollaries of the Phylogenetic Species Concept (dPSC), as an operational criterion to identify lineages under the ontological Evolutionary Species Concept; and to establish the lineages and their independence I will use

species delimitation techniques, and with genetic and morphological dataset. To avoid confusing the tokogenetic relationships among populations with the phylogenetic relationships among species, I will use two approaches: (*i*) the Dayat's suggestion about the Integrative Taxonomy to search discontinuities in different datasets; and (*ii*) Carstens et al. (2013) proposal, to use different approaches for species delimitation analysis and integrate the different databases in order to consider whether the lineages have left the grey area, and can be recognized as a taxonomic unit.

1.3. The Genus *Proechimys*

The Order Rodentia BOWDICH is the most diverse order among living mammals with 42% of the known diversity of mammals (Wilson and Reeder 2005). This diversity was organized into five suborders based on morphological data sensu Musser and Carleton (2005): Anomaluromorpha BUGGE; Castorimorpha WOOD; Sciuromorpha BRANDT; Myomorpha BRANDT; and Hystricomorpha BRANDT, and only the first one does not have representatives in the Neotropical region. Molecular data showed a different organization, with the diversity of rodents being organized into three large clades (Blanga-Kanfi et al. 2009; Churakov et al. 2010; Fabre et al. 2012). In this approach, rodents are structured into (*i*) "mouse-related" clade (including the Musser and Carleton suborders: Anomaluromorpha, Castorimorpha and Myomorpha); (*ii*) "squirrel-related" clade (Sciuromorpha sensu Musser and Carleton); and (*iii*) Ctenohystrica HUCHON, CATZEFLIS AND DOUZERY (Hystricomorpha sensu Musser and Carleton) (Huchon et al. 2000; Montgelard et al. 2008; Wu et al. 2012).

Rodents have originated in Asia and several hypotheses have been elaborated on their arrival in South America and later diversification (Luckett and Hartenberger 1985, 1993; Meng et al. 1994; Marivaux et al. 2002) because South America was an isolated continent since its separation from Africa [between 115-120 Ma] (Moulin et al. 2010; Heine et al. 2013; Will and Frimmel 2018) until the connection with North America by the formation of the Panama land bridge [between 7-15 Ma] (Hoorn and Flantua 2015; Montes et al. 2015).

Some rodents from Subfamily Murinae ILLIGER were introduced in South America during the European colonization of the continent (Long 2003; Puckett et al. 2016). Considering natural invasions, some groups colonized the continent by the north before

[Subfamily Sigmodontinae WAGNER (Leite et al. 2014)] or after [Tribe Sciurini G. FISCHER (Pečnerová and Martínková 2012)] the appearance of the Isthmus of Panama. The Ctenohystrica representatives from South America are known as Caviomorpha WOOD and their arrival in the continent is an older event, the oldest caviomorph fossil record is for Middle Eocene (41 Ma) from Peru (Antoine et al. 2012). Molecular data have suggested the monophyly of Caviomorpha and that its sister group would be African ctenohystrics (Huchon and Douzery 2001; Rowe et al. 2010; Voloch et al. 2013; Upham and Patterson 2015). Rowe et al. (2010), Voloch et al. (2013), and Upham and Patterson (2015) indicated that caviomorphs separated from phiomorphs between 55 and 41 Ma in agreement with the fossil record. The most controversial point on the caviomorphs arrival in the continent is the route (Huchon and Douzery 2001). Caviomorphs arrived in the Neotropics when Africa and South America were completely separated, which would become difficult a direct dispersion, raising alternative hypotheses of arrival routes across North America (Woods and Patterson 1959; Hoffstetter 1972; Woods 1980) or Antarctica (Houle 1999; Poux et al. 2006). Nowadays, the most accepted hypothesis is a single event of dispersion from Africa directly to South America by marine currents or "stepping stone" islands (Lavocat 1969; Flynn and Wyss 1998; Huchon and Douzery 2001; Poux et al. 2006; Rowe et al. 2010).

Once on the continent, the caviomorphs have experienced a huge diversification, reaching 244 living species (Upham and Patterson 2015) and currently are classified in four superfamilies: Cavioidea FISHER DE WALDHEIM; Chinchilloidea BENNETT; Erethizontoidea BONAPARTE; and Octodontoidea WATERHOUSE. The most diverse superfamily is Octodontoidea (spiny rats, degus, and their allies) with 70% of caviomorph genera and 75% of living species (Upham and Patterson 2015). Within this superfamily, the family of spiny rats, Echimyidae GRAY, is the most diverse among caviomorph families (Fabre et al. 2016). Molecular-based taxonomy (Fabre et al. 2017; Courcelle et al. 2019) has altered the classification and composition of this family when compared to morphological arrangements (Woods and Kilpatrick 2005). Here, I will follow the more recent view and adopt a composition of Family Echimyidae including the spiny rats, hutias and the coypu, in accordance to Courcelle et al. (2019).

In this view, the family Echimyidae is composed of four subfamilies: Capromyinae SMITH; Carterodontinae COURCELLE ET AL.; Echimyinae GRAY; and Euryzygomatomyinae FABRE ET AL. The Subfamily Echimyinae represents the spiny rats and it is divided into two

tribes: Echimyini FABRE ET AL., with all representatives being arboreal spiny rats; and the Myocastorini FABRE ET AL., represented by five genera with different habits use patterns: *Hoplomys* J. A. ALLEN, *Proechimys* J. A. ALLEN, and *Thichomys* TROUESSART are terrestrial, *Myocastor* KERR is semi-aquatic, and *Callistomys* EMMONS & VUCETICH is arboreal (Fabre et al. 2016, 2017; Courcelle et al. 2019).

Proechimys is the most diverse genus within in the Family Echimyidae, with 22 valid species and more than 60 nominal taxa available for the species group (Patton and Leite 2015). Its sister group is another terrestrial genus *Hoplomys*, and the separation between these genera was around 10 Ma (Upham and Patterson 2015; Álvarez et al. 2017). *Proechimys* has a wide distribution in the Neotropics, that occurs from Central America to the Brazilian Cerrado, covering the entire Amazon region (Woods and Kilpatrick 2005). In addition, *Proechimys* is very abundant in wildlife survey studies in the Amazon region and the record of sympatric species is common in the genus (Malcolm 1992; Patton et al. 2000; Steiner et al. 2000). Despite being abundant and widely distributed, this genus is little studied and has its taxonomy and systematics quite unresolved (Da Silva 1998). This lack of knowledge affects studies in other areas, especially in ecology because the species are externally similar, making it difficult to identify them in the field. The confusing taxonomy and systematics of the group is a consequence of (*i*) poorly and not adequately delimited species, and (*ii*) nomenclatural problems. Regarding species delimitation, this occurs, among other reasons, by the complex morphological variation. The analysis of this variation is difficult to perform in the genus *Proechimys* and throughout family Echimyidae because most morphological characters has not changed during the evolution of echimyids (Lara et al. 2002). Thus, plesiomorphic states of the characters were retained throughout the evolutionary history of the group, especially in terrestrial genera as *Proechimys* (Fabre et al. 2013). Regarding nomenclatural problems, the confusion in the taxonomy of *Proechimys* is associated to the large number of nominal taxa for the species group of the genus, with 65 available names in total (Fig. 1), with holotypes spread across museums in Europe, United States of America (USA) and Brazil (Table 1). This makes it difficult naming correctly the specimens analyzed in studies, which increases the chance of nomenclatural errors. However, it is important to notice that a comprehensive study on the morphological variation never was properly conducted for the genus, through the assessment of large series from its entire distribution, with subsequent proper name attribution based on the comparative study of specimens and type specimens.

In the most comprehensive study so far, Patton (1987) contributed to the understanding of morphological variation of the genus species when separated *Proechimys* into nine groups based on cranial and bacular characters. More recently, Patton and Leite (2015) proposed a new rearrangement these species groups totaling 10 groups (Fig. 1). The authors reviewed each of the 22 species, with comments on morphology, cytogenetics and range distribution. In addition, they presented an identification key for species using morphological characters, mostly cranial. However, this classification considered mainly the morphological similarities because data on the phylogenetic relationships among the species are scarce. Levels of morphological variation and genetics within and between *Proechimys* species remain largely unknown, making it difficult to identify them in studies of fauna survey and ecological studies that are not based on testimony specimens.

Before the Patton and Leite (2015) study (that, although important, is not a proper revision) the latest and broadest revision of genus *Proechimys* was conducted by Moojen (1948). Studies published between these revisions had been proposed to resolve the relationships between groups of species and delimitation of some groups, or to present a compilation of valid species (Cabrera 1961; Patton and Gardner 1972; Gardner and Emmons 1984; Woods and Kilpatrick 2005). Few studies have been published on the morphological variation of *Proechimys* (Patton and Rogers 1983; Gardner and Emmons 1984; Patton 1987; Patton et al. 2000; Corti et al. 2001; Voss et al. 2001). Corti et al. (2001) focused on the variation of six species of *Proechimys* of the Orinoco basin from Colombia and Venezuela, while Gardner and Emmons (1984) grouped the *Proechimys* species in species groups, based on the bulla structure and cytogenetic data. Patton and Rogers (1983) analyzed the intraspecific variation of a population of *Proechimys* brevicauda in Peru. Patton et al. (2000) and Voss et al. (2001) analyzed the morphological variation intra and interspecific in *Proechimys* occurring in the basin of the Juruá River in the state of Amazonas, Brazil and in certain localities of French Guiana, respectively.

Genetic studies using DNA sequences are also uncommon (Da Silva and Patton 1998; Patton et al. 2000; Steiner et al. 2000; Van Vuuren et al. 2004; Silva et al. 2018) and deal with the phylogeography of few species (Da Silva and Patton 1998; Silva et al. 2018), or the molecular differentiation of two species of French Guiana (Steiner et al. 2000; Van Vuuren et al. 2004), or even with the species definition that occur in the basin of the Juruá River (Patton et al. 2000). There are four published studies about phylogeography of

Proechimys species in the Amazon Forest. Three of them used the mitochondrial marker cytochrome b (Cyt b) sequences to study the sympatric species *Proechimys guyannensis* and *Proechimys cuvieri* from French Guiana (Steiner et al. 2000; Van Vuuren et al. 2004) or from a larger area in northeastern Amazon (Silva et al. 2018). The studies from French Guiana populations did not show a pattern of genetic structure according to geography in the target species, although *P. cuvieri* present greater genetic diversity than *P. guyannensis* (Van Vuuren et al. 2004). However, Silva et al. (2018), in a larger geographic area for the same two species, found no evidence that the rivers acted as barriers, and indicated that the climatic changes of the Pleistocene and the geological changes in the region would be more likely to act as diversification drivers, besides *P. guyannensis* present greater genetic distances when compared to the *P. cuvieri*. On the other hand, the fourth study that also used Cyt b sequences of the species group *goeldii*, showed that Solimões-Amazonas River system and that the endemism areas proposed by Cracraft (1985) are important to explain the genetic structure of the goeldii species group (Da Silva and Patton 1998).

Matocq et al. (2000) published one of the few studies on population genetics in the genus *Proechimys*. They analyzed the variation of a mitochondrial marker in two sympatric species, *P. steerei* and *P. simonsi*, occurring in the Western Amazon. In *P. steerei*, occupying seasonal floodplain forest (*várzea* forests), they found high gene flow between populations and the river does not seem to be a barrier to dispersal of the species. In *P. simonsi*, which occupies more stable habitats (*terra-firme* forests), they found little gene flow among populations on the same side of the river as between the opposite river bank. The authors stated that more data is needed to be collected to explain the mechanisms that keep the little gene flow in *P. simonsi*. However, most of the genetic studies is focused on the cytogenetics (Kasahara and Yonenaga-Yassuda 1984; Aguilera et al. 1995; Bonvicino et al. 2005; Machado et al. 2005; Amaral et al. 2013; Rodrigues da Costa et al. 2016), some of them without the identification of specimens at the specific level. Machado (2017) performed a great review of the published karyotypes for the genus *Proechimys*, totaling 39 distinct karyotypes, associated with 21 of 22 valid species for the genus currently (only *P. hoplomyoides* has no karyotypic description). The diploid number in *Proechimys* ranges from 14 to 62 and a fundamental number from 16 to 80. In addition, cytogenetics is informative for species identification, since only the species *P. gardneri* and *P. pattoni* share the same cytotype with 2n = 40 and FN = 56 (Machado 2017).

Only two unpublished studies proposed phylogenetic hypothesis based on genetic data for a large number of species in the genus. Schetino (2008) used data from two mitochondrial markers, cytochrome b (Cyt b) and cytochrome C oxidase subunit I (COI) to study the barcodes efficiency in the genus *Proechimys*. Phylogenetic analyses recovered a basal polytomy, also observed within some clades, being uninformative for the relationships among species and species groups (Fig. 2). Leite et al. (2015) proposed a phylogeny for the genus with five markers (one mitochondrial and four nuclear). In this study it is possible to establish the phylogenetic relationships among some species groups, even with low interspecific resolution within each group (Fig. 3). It is important to note that even with five markers the phylogeny proposed still presents polytomies that preclude further analysis of the evolution of *Proechimys*. The authors concluded that further studies with new approaches would be needed to improve the phylogenetic hypothesis for this genus.

Other studies involving the genus *Proechimys* refer to ecology (Everard and Tikasingh 1973; Alho 1980; Emmons 1982; Linardi et al. 1991; Adler and Beatty 1997), the description of new species (Da Silva 1998), comments on the systematics and taxonomy of some species (Weksler et al. 2001), new records and expansion of the distribution of some species (Sánchez-Vendizú et al. 2018), and position of *Proechimys* within the Family Echimyidae radiation (Lara et al. 1996; Leite and Patton 2002; Galewski et al. 2005; Fabre et al. 2013, 2017; Upham and Patterson 2015; Courcelle et al. 2019).

In summary, it is evident that there are few studies on genus *Proechimys*, more precisely on the species limits definition, based on morphologic and genetic variation patterns. A study focusing in an integrated taxonomic approach, using genetic and morphological data could achieve results that would allow advancements on the knowledge of the *Proechimys* evolution. Being a dominant species in small mammal assemblages in Amazon, the comprehension of *Proechimys* evolutionary history also will bring information on the biogeographic history of this biome.

1.4. Taxonomic History

Proechimys J. A. ALLEN and its nominal taxa have a complex taxonomic history since the first nominal taxon has been described 216 years ago. The first nominal taxa for

what is currently considered as *Proechimys* was described by Étienne Geoffroy Saint-Hillaire (1803), as "*le rat de la Guyane*" (*Mus guyannensis*). His work was considered as not available for a long time because É. Geoffroy did not publish the final volumes, and his work was released among friends, with few copies (Hershkovitz 1955). Thus, his descriptions were disregarded to the taxonomic history by some scientists, which led to several re-descriptions of the É. Geoffroy's species, including *Mus guyannensis*. For decades there has been debate among scientists whether Étienne Geoffroy Saint-Hillaire's Catalogue should be considered published or not, and hence discussions on the validity of nominal taxa too. Lastly, an ICZN decision considered the work published and the nominal taxa proposed by Étienne Geoffroy Saint-Hillaire were considered valid and available (Grubb 2001).

A few years later, Cuvier (1809) published the name *Echimys* to designate South American spiny rats in a study on the characteristics of mammalian teeth. In fact, the genus *Echimys* was named by Etienne Geoffroy Saint-Hilaire in a manuscript written in 1808 or 1809 and never published (Allen 1899), so the authorship of the genus is by Cuvier. The genus was described in more details later by Cuvier (1812), adding two species already described to the genus *Echimys*, '*le rat epineux*' [*Echimys spinosus*, currently *Euryzogomatomys spinosus* (G. FISCHER)] and '*le rat a queue doree*' [*Echimys cristatus*, currently *Echimys chrysurus* (ZIMMERMANN)].

Desmarest (1817) described the species *Echimys cayennensis* based on *Mus guyannensis* from E. Geoffroy Saint-Hillaire. For decades there was discussion about which name should be used to designate the taxon. After ICZN decision, considering the Geoffroy's catalogue as a valid study and published (Grubb 2001), the correct name to this taxon would be *Mus guyannensis*, today *Proechimys guyannensis*, and *Echimys cayennensis* is its junior synonym.

Rengger described *Echimys longicaudatus* based in one individual from northern Paraguay near the 21st parallel (Rengger 1830) and close to the border with Brazil (Tate 1935). In his first review of spiny rats, I. Geoffroy Saint-Hilaire (1838) allocated *E. longicaudatus* to the genus *Loncheres* ILLIGER, using the combination *Loncheres longicaudatus*, also the author used the name *Echimys cayennensis* for the taxon described by his father, following Desmarest (1817). In another review of spiny rats, I. Geoffroy Saint-Hilaire (1840) considered *E. longicaudatus* as a synonym of *Loncheres myosurus* (a

Trinomys) (Table 2). Pictet (1841) considered all three species (*E. guyannensis*, *E. cayennensis*, and *E. longicaudatus*) as a single species under the name of *Echimys cayennensis* (Table 2), as he considered that the morphological variation was due to age variation (Tate, 1935).

Tomes (1860) described *Echimys semispinosus* and *Echimys brevicauda* was described by Günther (1876). Trouessart (1880) proposed the division of *Echimys* into two subgenera (*Echimys* and *Thrichomys*) in his study on living and fossil mammals. For this author, *E. brevicauda* would belong to the subgenus *Thrichomys*, and *E. semispinosus* and *E. cayennensis* to the subgenus *Echimys*, with *E. longicaudatus* being a synonym for the later species; he did not mention *E. guyannensis* (Table 2).

Only in 1899 the genus *Proechimys* was created by Allen (1899). He rearranged the genus *Echimys*, and the terrestrial forms were allocated in the new genus *Proechimys*, while *Echimys* was restricted to arboreal spiny rats (Table 2). Between 1899 and 1926, 34 of the currently 65 nominal taxa for the genus *Proechimys* were described (Table 1), especially due to the efforts of two researchers: J. A. Allen at the American Museum of Natural History, in New York, and O. Thomas at the British Museum of Natural History, in London. The large number of taxa described in this period may be explained by the influx of specimens from previously unsampled areas in Central and South America; the scarce series of specimens available at the museums for comparisons; and by the philosophy of taxonomy, more typological/morphological and "spliter" in that time, where geographic variations were described as subspecies or even species (Sokal 1962; Wilkins 2009). Furthermore, some taxa have been described based on subadult individuals (*e.g., Proechimys ochraceus*, *Proechimys poliopus*, *Proechimys leucomystax*), making it difficult to compare species since ontogenetic variation can be confused with interspecific variation (Ridley 2004).

Thomas (1921) proposed a subdivision for *Proechimys* into two subgenera based on the number of "laminae" (crests) present in the cheekteeth: species with four laminae should be allocated to the subgenus *Proechimys* and with three laminae to the newly described subgenus *Trinomys*. Other additional cranial characters, such as shorter rostrum for *Trinomys*, and opisthodont incisors and temporal and parietal ridges more developed for *Proechimys* were also employed by Thomas (op.cit) as diagnostic traits on the description of this new subgenus. In this classification, *Proechimys iheringi* THOMAS, and *Proechimys dimidiatus*

GÜNTHER, currently in genus *Trinomys* (sensu Pessôa et al. 2015), were allocated within the subgenus *Proechimys* (Table 2).

Tate (1935) published a study about the taxonomy of the genera of Neotropical "Hystricoid" rodents, and he summarized the taxonomic history of *Proechimys* and also presented a compilation for the species described for the genus. Ellerman (1940) reviewed 38 of 50 described nominal taxa of *Proechimys* and divided them into two subgenera: subgenus *Trinomys* with three nominal taxa and the subgenus *Proechimys* with 35. The subgenus *Proechimys* was divided into two groups: 34 species attributed to the *cayennensis* group and one nominal taxa to the *iheringi* group. The 12 nominal taxa not analyzed by Ellerman were left without groups under the subgenus *Proechimys* (Table 2). Accordingly to Ellerman, most of the characteristics indicated by Thomas and others as diagnostic for species of *Proechimys* were mere interspecific differences, representing extreme variations connected by intermediate forms, for instance, on the variation of body size; other trait variation, as the development of parietal ridges, could be related to age variation, with older specimens exhibiting more robust ridges than younger specimens.

The next revision for the genus was performed by Hershkovitz (1948). He used the name *guyannensis* rather than *cayennensis*, and he analyzed the 12 taxa that Ellerman did not examine. His definition of subgenus *Trinomys* was the same as Thomas (1921), and he considered the nominal taxa macrourus as form of *Echimys armatus* (=*Makalata* HUSSON). For him, *Proechimys* consist of five species: *iheringi*, *canicollis*, *hendeei*, *quadruplicatus* and the polytipic *P. guyannensis* (Table 2).

Moojen (1948) reviewed the spiny rats from Brazil, arranging them in two subgenera: *Trinomys* (4 species, 3 polytypic) and *Proechimys* (4 species, 4 polytypic) and 1 *incertae sedis*. He described 14 taxa for the genus, nine in the subgenus *Proechimys*. He used the geographical distribution and the main fold in the molariform teeth to differentiated the subgenera: in *Trinomys* the main fold crosses the crown of the tooth and in *Proechimys* it extends halfway across. He pointed the possibility of *Hoplomys* to represent another subgenus of *Proechimys*, and also considered *Proechimys hoplomyoides* as allied to *Hoplomys* rather than to *Proechimys* (Table 2). Moojen also drew attention to some characters that apparently have clinal variation according to the humidity, with specimens from moist areas exhibiting

longer tails, darker color, smaller incisive foramen, and greater number of counterfolds in molariforms (Moojen 1948).

Cabrera (1961) recognized seven species in the subgenus *Proechimys*, six polytypic (Table 2), and he followed Moojen (1948), allocating hoplomyoides as a subspecies of *Hoplomys gymnurus*. The next proposal of taxonomic rearrangement was made 20 years after Cabrera's catalog, for Central America region: in this contribution six subspecies of Proechimys semispinosus were recognized by Hall (1981): *burrus*, *centralis*, *goldmani*, *ignotus*, *panamensis*, and *rubellus*. Gardner and Emmons (1984) divided the species of *Proechimys* into four groups of species according to the karyotype and the septal patterns of the bullae: *brevicauda*, *guairae*, *semispinosus*, and *Trinomys* (Table 2). Patton (1987) divided the nominal taxa for the species group of *Proechimys* into nine groups based on the baculum and on cranial characters (Table 2).

Lara et al. (1996) suggested the elevation of the *Trinomys* to the genus level, based on molecular data. Lara and Patton (2000) with more molecular data and Carvalho and Salles (2004) with morphological data corroborated this hypothesis, and in the catalog of mammals of the world, Woods and Kilpatrick (2005) also considered *Trinomys* as a genus. The last species described for *Proechimys* were: *Proechimys echinothrix*, *Proechimys gardneri*, *Proechimys kulinae*, and *Proechimys pattoni* (Da Silva 1998). Weksler et al. (2001) used cytogenetic, morphological and morphometric data to associate the taxon *P. roberti* to the *guyannensis* group, a species that sometimes was considered as a member of *longicaudatus* group (Moojen 1948; Cabrera 1961), and they affirmed that *P. oris* would be the junior synonym of *P. roberti*.

In 2001 another taxonomic problem that affected *Proechimys*, besides the genera *Holochilus* BRANDT and *Trinomys*, was solved (ICZN 2001). Brandt (1835) described the subgenus *Holochilus* based on two specimens from Langsdorff's expedition in Brazil: one individual was a new species *Mus (Holochilus) leucogaster* BRANDT and other one was identified as *Mus (Holochilus) anguya* DESMAREST (a misspelling of *Mus angouya* =*Sooretamys angouya*). Later, Brandt (1855) realized that these specimens were hystricomorphous rodents and to correct the error he classified *Holochilus*, now as a genus, within the suborder Hystrichomorphi BRANDT, containing the species *H. leucogaster* (currently a *Trinomys*) and the specimen *Mus (Holochilus) anguya* that was described as a

new species *Holochilus langsdorffi* (currently a *Proechimys*). For the myomorphous rodents species already described as *Holochilus* (*e.g.*, *Mus brasiliensis* DESMAREST, *Mus vulpinus* BRANDT, and *Holochilus sciureus* WAGNER) he proposed the name *Holochilomys* BRANDT. Unfortunately, his taxonomic rearrangement went unnoticed by the other scholars and the name *Holochilus* continued to be used to designate a Muridae genus. Miller and Rehn (1902) fixed the type species of *Holochilus* as being *Mus (Holochilus) leucogaster*, a *Trinomys*. As a taxonomic consequence, the names *Trinomys* and *Proechimys* would be junior synonyms of *Holochilus*. Voss and Abramson (1999) drew attention to this fact, and they proposed to change the type species of the genus *Holochilus* to *Holochilus sciureus* to maintain the stability of the three generic names, which was approved by the ICZN (2001). Thus, currently *Holochilus sciureus* is the type species of the genus *Holochilus*, the nominal taxon *leucogaster* was allocated to the genus *Trinomys* and the *langsdorffi* to the genus *Proechimys*. Abramov and Baranova (2008) identified *Holochilus langsdorffi* as a member of *longicaudatus* group, but, meanwhile, Patton and Leite (2015) said that the type should be analyzed to verify if it can be considered a senior synonym of some current species.

The last compilation of mammal species of the world considered *Proechimys* with 25 species (Woods and Kilpatrick 2005), and in a book series on mammals of South America performed by Patton and Leite (2015), the genus assembles 22 valid species, and 10 species groups (Table 2). Fabre et al. (2016) followed the classification of Patton and Leite (2015), with the main changes being related to the subspecies recognition in *Proechimys semispinosus* and *Proechimys guairae*, and the synonymy between *Proechimys roberti roberti* and *P. roberti oris*, the latter a taxon that Patton and Leite (2015) considered as valid subspecies (Table 2). Fabre et al. (2016) did not recognized monotypic species, such as *Proechimys canicollis*, *P. decumanus*, *P. echinothrix* and *P. simonsi*, in unique and monotypic species groups. In addition, they did not include *cayennensis* in their list of synonymies, and attributed the nominal taxa *myosurus* and *myorurus* as synonyms of *Proechimys longicaudatus* without justification, a major controversy, as historically these taxa names had been associated to the terrestrial spiny rats from the Atlantic forest in the genus *Trinomys*. Here, I will follow the more complete arrangement (sensu Patton and Leite), with 65 nominal taxa for the species group of genus *Proechimys*.

1.5. Western Amazon

The Amazon region is one of the world's most biodiverse areas (Myers et al. 2000; ter Steege et al. 2003), and several hypotheses were proposed to explain its biological richness (Bonvicino and Weksler 2012; Leite and Rogers 2013). Some of them are based on geological events such as structural paleoarchs (Da Silva and Patton 1998; Lougheed et al. 1999; Patton et al. 2000) or the formation of Amazonian rivers (Wallace 1854; Ayres and Clutton-Brock 1992; Hoorn et al. 2010a), and other hypotheses emphasize climatic or environmental variations, such as Pleistocene refugia (Haffer 1969) or environmental gradients (Patton and Smith 1992; Funk et al. 2007). Despite the apparent disagreement among hypotheses, all have one thing in common: Amazon is not geologically and environmentally homogeneous, and regional differences within the region are important to explain its biodiversity (Bonvicino and Weksler 2012; Leite and Rogers 2013).

A classic regional division of the Amazon region is between the Western and Eastern Amazon (Aleixo and Rossetti 2007; Leite and Rogers 2013). Each of them was affected differently by the geological phases that shaped the Amazon region. Eastern Amazon encompasses areas of the Brazilian and Guyana Shields and it was related to the Cratons formation, with very old and ultra-stable basement, which had suffered little influence of tectonic activity after the separation of South America and Africa (Kroonenberg and Roever 2010). Thus, this region was more geologically stable, even in relation to the deposition of sediments that was limited after the Cretaceous (Rossetti et al. 2005; Aleixo and Rossetti 2007). On the other hand, regarding the climate and the vegetation of this area, this notion of stability can be challenged, as during the Pleistocene glacial cycles, the drier and colder climate could have caused fragmentation or at least retraction of the forests in all Amazonian region or at least in the edge regions and Eastern Amazon (Haffer 1969; Hammen and Hooghiemstra 2000). Even if the reduction of forested areas did not occur, a change in the floristic composition of Amazonian forests is plausible (Colinvaux et al. 2001; Lessa et al. 2003). In addition, most of the Amazonian savanna areas are retained in the Eastern Amazon (Resende-Moreira et al. 2019), suggesting a possible change in the vegetation, with the expansion of open areas during glacial cycles in the region.

Western Amazon usually comprises the interfluve area between the Negro and Madeira rivers (Leite and Rogers 2013), between the Purus Arch to the east, Andes

Mountains to the west, Vaupés Arch or Swell to the north, and Fitzcarrald Arch to the south. Its biodiversity is greater than the Eastern Amazon (Finer et al. 2008; Wade 2015), even though it is less studied (Mendes-Oliveira et al. 2015). Unlikely Eastern Amazon, Western Amazon is more humid (Costa and Foley 1998) and the recent geological activity presented more influence in its currently landscape and climate than the glacial cycles in Pleistocene due to the uplift of the Andes, especially in the last 30 Ma (Hoorn et al. 2010a). In the Western Amazon, the climate has already stabilized, with similar rainfall regimes than present-day, since the Miocene (Kaandorp et al. 2005), and even during Pleistocene climatic oscillations there was little change in precipitation regime (Cheng et al. 2013), and no reductions in the forested areas in the region (Häggi et al. 2017). Thus, the geological instability and landscape evolution of the last 30 Ma seem to be the key to understanding the Western Amazon biodiversity.

The western border of South America had already affected by tectonic activity since 100 Ma but in the last 30 Ma the activity intensified to reach the current relief pattern (Mora et al. 2010). The turning point for tectonic activity was the Farallon plate's rupture into Nazca and Cocos plates around 23.4 Ma, because after the break up there was an increase of the subduction of the plates in the region (Cobbold et al. 2007).

Western Amazon had a different landscape during the Miocene and Pliocene when compared to the present-day. A major Amazon River crossing South America from west to east did not exist (Hoorn et al. 1995), and a structural arch called Purus Arch was a barrier to the water flows in west-east direction, separating the Amazon and the Solimões rivers basins (Albert et al. 2018). During the Miocene, Amazon basin shape was different, called Pan-Amazon basin, the region encompassed the basins of the Orinoco, Magdalena, Maracaibo, Northern Paraná rivers (Lundberg et al. 1998). In this basin, Western Amazon was a marshy and swampy region that had the waters drained to the north into the Caribbean Sea, through a paleo-basin of the Orinoco River (Hoorn et al. 2010b), and also it was exposed to marine incursions events (Hoorn 1993). In short, the region was composed of a mega-wetlands area with marine influence called Pebas System or Lake Pebas (Wesselingh et al. 2001). The forests that existed at the time were fragmented and bordered the rivers and lakes in that system (Hoorn 1993; Pons and De Franceschi 2007). In this environment, aquatic biota had connections to disperse from Western Amazon to Guyana Shield, Northern South America and Tropical Andes but these connections were not available to the terrestrial biota

(Wesselingh 2006). Pebas system had similar or greater biodiversity than today (Hooghiemstra and Hammen 1998; Hammen and Hooghiemstra 2000), which could indicate pluviometric and temperatures indexes similar to the present (Mora et al. 2010). Pebas system persisted until around 10 Ma, right after it stopped receiving marine influence and became a fluvial-tidal mega-wetland called the Acre system (Albert et al. 2018).

One of the possible causes for this change is the emergence of the Vaupés Arch or Swell in southern Colombia by tectonic activity (Jaramillo et al. 2010). This arch became a barrier to most rivers that flowed from Western Amazon into Caribbean Sea, isolating the Orinoco and Western Amazon basins (Hoorn et al. 2010a). Purus Arch, about the same time, was no longer a barrier to the water flow from west to east due to the millions of years of Andean sediment deposition in the Western Amazon, forming the present sedimentary basins, and silting the area (Hoorn et al. 2010b; Mora et al. 2010; Albert et al. 2018). First indications of a transcontinental system were around 9.4 Ma, and around 7 Ma the current known river bed was already established (Hoorn et al. 2017). However, the largest sediment increase at the mouth of the Amazon River was in the Quaternary, after 3 Ma (Figueiredo et al. 2009; Hoorn et al. 2017). This can be explained by the Iquitos Arch, which worked as a sediment trap in the region up to Quaternary, creating the Nauta/Içá Formation (van Soelen et al. 2017). Nevertheless, Iquitos Arch did not prevent the rivers water from Western Amazon flowing to the Solimões and Amazonas basins, only the most part of the sediments (Albert et al. 2018).

Acre system began to be drained with the establishment of the transcontinental Amazon River, reducing the wetlands area from 10% to 2% of the modern Amazon basin (Albert et al. 2018), being replaced in the first moment by grasses and later by terra-firme forests around 7 – 5 Ma (Hoorn et al. 2010a; Albert et al. 2018). Thus, the connection among different Amazonian regions for terrestrial biota increased to the detriment of the connections between aquatic one. One of the most recent changes caused by tectonic activity in the Western Amazon was the rise of the Fitzcarrald Arch around 4 Ma years ago in southern Peru, which divided the basins of the Ucayali and Madre de Dios rivers (Espurt et al. 2010). Between Late Pliocene and the present-day, the Iquitos Arch lost its role as a barrier to the sediments, and due to some river captures, the headwaters of some Amazonian rivers extended to the Northern Andes (Albert et al. 2018).

In short, the tectonic activity and the uplift of the Andes had more influence on the current Western Amazon landscape than the climatic oscillations. The past geological history had direct implications on the formation of sedimentary basins, and edaphic characteristics in Western Amazon, which consequently influenced vegetation and biodiversity in the region. Thus, to study the evolutionary history of taxa inhabiting the region, it is necessary to understand, in addition to the climatic change, the landscape evolution in the Western Amazon caused by the geological events.

1.6. Objectives and Hypotheses

The genus *Proechimys* is a diverse group of terrestrial rats that inhabit both Eastern and Western Amazon, with large number of sympatric and allopatric species on both areas. As a diverse and widespread species, with cases on allopatry and sympatry, the genus is very interesting model to study in order to establish hypothesis on the evolution of Amazonian biota. However, to achieve this, a consistent taxonomy is imperative, but most of these species remained poorly defined, with uncertain geographic distribution and unknow phylogenetic relationships. Therefore, I aim to conduct a thorough and consistent analyses of morphological and genetic variation of hundreds of samples, under an integrative approach and modern methods and concepts of species delimitation, to establish the diversity of the genus and present its history in Amazon. These data may help in wildlife surveys in the region, the identification of specimens in museums and in the field and in the conservation of species of the genus because these data are important for diversity assessments, management plans and conservation policies.

My thesis is divided into three chapters organized as scientific manuscripts, along with this initial chapter on the general introduction, and a last chapter about the conclusions, implications of my results, and future perspectives.

In Chapter 2, I tested the hypothesis that it was possible to infer the phylogenetic relationships for a group with deep divergence time, such as *Proechimys* through the NGS technique used mainly for taxa with shallow divergence times. For that, I inferred phylogenetic trees, analyzed their statistical support, and performed simulations with morphometric data for the identification of isolated lineages that could be considered putative

species. At the end of this chapter, I was able to determine the main clades of the genus, propose some taxonomic implications, and identify isolated lineages that may be considered as distinct taxa.

In Chapter 3, I tested whether the evolutionary history of the genus *Proechimys* would reflect the landscape evolution of the Amazon region, identifying the importance of climate and geological changes in the region for the diversification of the genus. For that I estimated the divergence times between clades, and I proposed a biogeographic hypothesis for *Proechimys* evolution, indicating the dispersal and vicariance events along to the genus phylogeny.

In Chapter 4, I tested whether three sympatric species occupying the same geographical region, such as the Western Amazon, would also share (*i*) the same phylogeographic patterns and (*ii*) the same environmental and morphological hypervolumes. For that, I accessed the phylogeographic patterns and analyzed if they could be explained by the isolation by distance (IBD), isolation by barriers (IBB), or isolation by environment (IBE) models. In addition, I identified regions of historical habitat stability through ecological niche models, as well as I calculated the overlap and similarity of environmental and morphological hypervolumes in the sympatric species to understand how they use and share the habitat. At the end of this chapter, I indicated some implication for conservation.

References

ABRAMOV, A. V, AND G. I. BARANOVA. 2008. The Langsdorff's expedition to Brazil and its mammal collection kept in the Zoological Institute of St. Petersburg, Russia. Russian Journal of Theriology 7:41–50.

ADLER, G. H., AND R. P. BEATTY. 1997. Changing reproductive rates in a Neotropical forest rodent, *Proechimys semispinosus*. The Journal of Animal Ecology 66:472.

AGAPOW, P. ET AL. 2004. The impact of species concept on biodiversity studies. The Quarterly Review of Biology 79:161–179.

AGNARSSON, I., AND M. KUNTNER. 2007. Taxonomy in a changing world: seeking solutions for a science in crisis. Systematic Biology 56:531–539.

AGUILERA, M., O. A. ALDO REIG, AND A. PEREZ-ZAPATA. 1995. G-and C-banding karyotypes of spiny rats (*Proechimys*) of Venezuela. Revista Chilena de Historia Natural 68:185–196.

ALBERT, J. S., P. VAL, C. HOORN, J. S. ALBERT, P. VAL, AND C. HOORN. 2018. The changing course of the Amazon River in the Neogene: center stage for Neotropical diversification. Neotropical Ichthyology 16:e180033.

ALEIXO, A., AND D. DE F. ROSSETTI. 2007. Avian gene trees, landscape evolution, and geology: towards a modern synthesis of Amazonian historical biogeography? Journal of Ornithology 148:443–453.

ALHO, C. J. R. 1980. Homing ability in the wild rodent *Proechimys roberti* determined by radiotelemetry. Revista Brasileira de Biologia 40:91–94.

ALLEN, J. A. 1899. The generic names *Echimys* and *Loncheres*. Bulletin of the American Museum of Natural History 12:257–264.

ÁLVAREZ, A., R. L. M. ARÉVALO, AND D. H. VERZI. 2017. Diversification patterns and size evolution in caviomorph rodents. Biological Journal of the Linnean Society 121:907–922.

AMARAL, P. J. ET AL. 2013. *Proechimys* (Rodentia, Echimyidae): characterization and taxonomic considerations of a form with a very low diploid number and a multiple sex chromosome system. BMC Genetics 14:21.

ANTOINE, P.-O. ET AL. 2012. Middle Eocene rodents from Peruvian Amazonia reveal the pattern and timing of caviomorph origins and biogeography. Proceedings of the Royal Society B: Biological Sciences 279:1319–26.

AVISE, J. C. 2009. Phylogeography: retrospect and prospect. Journal of Biogeography 36:3–15.

AYRES, J. M., AND T. H. CLUTTON-BROCK. 1992. River boundaries and species range size in Amazonian primates. The American Naturalist 140:531–7.

BAKER, A. M., C. BARTLETT, S. E. BUNN, K. GOUDKAMP, F. SHELDON, AND J. M. HUGHES. 2003. Cryptic species and morphological plasticity in long-lived bivalves (Unionoida: Hyriidae) from inland Australia. Molecular Ecology 12:2707–2717.

BARRETT, C. F., AND J. V. FREUDENSTEIN. 2011. An integrative approach to delimiting species in a rare but widespread mycoheterotrophic orchid. Molecular Ecology 20:2771–2786.

BAUM, D. 1992. Phylogenetic species concepts. Trends in Ecology & Evolution 7:1–2.

BAUM, D. A., AND M. J. DONOGHUE. 1995. Choosing among alternative "phylogenetic"species concepts. Systematic Botany 20:560–573.

BEURTON, P. J. 2002. Ernst Mayr through Time on the Biological Species Concept - a Conceptual Analysis. Theory in Biosciences 121:81–98.

BLANGA-KANFI, S., H. MIRANDA, O. PENN, T. PUPKO, R. W. DEBRY, AND D. HUCHON. 2009. Rodent phylogeny revised: analysis of six nuclear genes from all major rodent clades. BMC Evol Biol 9:71.

BLAXTER, M. 2003. Counting angels with DNA. Nature 421:122–123.

BLAXTER, M. L. 2004. The promise of a DNA taxonomy. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences 359:669–79.

BOLNICK, D. I. ET AL. 2011. Why intraspecific trait variation matters in community ecology. Trends in Ecology & Evolution 26:183–192.

BONVICINO, C. R., I. B. OTAZÚ, AND J. F. VILELA. 2005. Karyologic and molecular analysis of Proechimys Allen, 1899 (Rodentia, Echimyidae) from Amazonian region. Arquivos do Museu Nacional 63:191–200.

BONVICINO, C. R., AND M. WEKSLER. 2012. Speciation in Amazonia: patterns and predictions of a network of hypotheses. Pp. 259–282 in Bones, clones, and biomes: the history and geography of recent Neotropical mammals (B. D. Patterson & L. P. Costa, eds.). The University of Chicago Press, Chicago.

BRANDT, J. F. 1835. Mammalium rodentium exoticorum novorum vel minus rite cognitorum Musei Academici Zoologici descriptiones et icones. Memoires de L' Académic Impériale des Sciences de Saint- Pétersbourg Sciences N:358–442.

BRANDT, J. F. 1855. Untersuchungen ueber die craniologischen Entwickelungsstufen und die davon herzuleilenden Verwandtschaften und Classificationen der Nager der Jetztwelt, mit besonderer Beziehung auf die Gattung *Castor*. Memoires de L' Académic Impériale des Sciences de Saint- Pétersbourg Sciences N:125–336.

CABRERA, A. 1961. Catalogo de los Mamiferos de America del Sur. Revista del Museo Argentino de Ciencias Naturales Bernardino Rivadavia IV:518–531.

CARSTENS, B. C., T. A. PELLETIER, N. M. REID, AND J. D. SATLER. 2013. How to fail at species delimitation. Molecular Ecology 22:4369–4383.

CARVALHO, G. A. S., AND L. O. SALLES. 2004. Relationships among extant and fossil echimyids (Rodentia: Hystricognathi). Zoological Journal of the Linnean Society 142:445–477.

CHENG, H. ET AL. 2013. Climate change patterns in Amazonia and biodiversity. Nature Communications 4:1411.

CHIQUITO, E. DE A., G. D'ELÍA, AND A. R. PERCEQUILLO. 2014. Taxonomic review of genus *Sooretamys* Weksler, Percequillo & Voss (Rodentia: Cricetidae: Sigmodontinae): an integrative approach. Zoological Journal of the Linnean Society 171:842–877.

CHURAKOV, G., M. K. SADASIVUNI, K. R. ROSENBLOOM, D. HUCHON, J. BROSIUS, AND J. SCHMITZ. 2010. Rodent Evolution: Back to the Root. Molecular Biology and Evolution 27:1315–1326.

COBBOLD, P. R., E. A. ROSSELLO, P. ROPERCH, C. ARRIAGADA, L. A. GÓMEZ, AND C. LIMA. 2007. Distribution, timing, and causes of Andean deformation across South America. Geological Society, London, Special Publications 272:321–343.

COLINVAUX, P. A., G. IRION, M. E. RÄSÄNEN, M. B. BUSH, AND J. N. DE MELLO. 2001. A paradigm to be discarded: geological and paleoecological data falsify the Haffer & Prance refuge hypothesis of Amazonian speciation. Amazoniana 16:609–646.

CORTI, M., M. AGUILERA, AND E. CAPANNA. 2001. Size and shape changes in the skull accompanying speciation of South American spiny rats (Rodentia: *Proechimys* spp.). Journal of Zoology 253:S0952836901000498.

COSTA, M. H., AND J. A. FOLEY. 1998. A comparison of precipitation datasets for the Amazon Basin. Geophysical Research Letters 25:155–158.

COURCELLE, M., M. TILAK, Y. L. R. LEITE, E. J. P. DOUZERY, AND P.-H. FABRE. 2019. Digging for the spiny rat and hutia phylogeny using a gene capture approach, with the description of a new mammal subfamily. Molecular Phylogenetics and Evolution 136:241–253.

CRACRAFT, J. 1983. Species concepts and speciation analysis. Pp. 159–187 in Current Ornithology (R. F. Johnston, ed.). Plenum Press, Boston.

CRACRAFT, J. 1985. Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. Ornithological Monographs:49–84.

CUVIER, G. F. 1809. Extrait des premiers mémoires de M. F. Cuvier, sur les dents des mammifères considérées comme caractères génériques. Nouveau Bulletin des Sciences, par la Société Philomathique 24:394–395.

CUVIER, G. F. 1812. Essai sur de nouveaux caractères pour les genres des Mammifères. Annales du Muséum d'Histoire Naturelle 19:268–295.

DALTON, R. 2003. Natural history collections in crisis as funding is slashed. Nature 423:575–575.

DAYRAT, B. 2005. Towards integrative taxonomy. Biological Journal of the Linnean Society 85:407–415.

DESALLE, R., M. G. EGAN, AND M. SIDDALL. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences 360:1905–1916.

DESMAREST, A. G. 1817. *Echimys*. Pp. 54–59 in Nouveau dic- tionaire d'histoire naturelle, appliquée aus arts, à l'agrruiculture, à l'économie rurale et domestique, à la medicine, etc. Par une société de naturalists et d'agriculteurs. 10:1– 6 (u. Nouv. éd. Paris, Deterville.

ELDREDGE, N., AND J. CRACRAFT. 1980. Phylogenetic patterns and the evolutionary process. Columbia University Press, New York.

ELLERMAN, J. R. 1940. The families and genera of living rodents. Vol. 1. Rodents other than Muridae. Trustees of the British Museum (Natural History), London.

EMMONS, L. H. 1982. Ecology of *Proechimys* (Rodentia, Echimyidae) in southeastern Peru. Tropical Ecology 23:280–290.

ESPURT, N., P. BABY, S. BRUSSET, M. RODDAZ, W. HERMOZA, AND J. BARBARAND. 2010. The Nazca Ridge and uplift of the Fitzcarrald Arch: implications for regional geology in northern South America. Pp. 89–100 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

EVERARD, C. O. R., AND E. S. TIKASINGH. 1973. Ecology of the Rodents, *Proechimys guyannensis trinitatis* and *Oryzomys capito velutinus*, on Trinidad. Journal of Mammalogy 54:875–886.

FABRE, P.-H. ET AL. 2017. Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. Molecular Biology and Evolution 34:613–633.

FABRE, P.-H., L. HAUTIER, D. DIMITROV, AND E. J. P DOUZERY. 2012. A glimpse on the pattern of rodent diversification: a phylogenetic approach. BMC Evolutionary Biology 12:88.

FABRE, P. H., T. GALEWSKI, M. K. TILAK, AND E. J. P. DOUZERY. 2013. Diversification of South American spiny rats (Echimyidae): A multigene phylogenetic approach. Zoologica Scripta 42:117–134.

FABRE, P. H., J. L. PATTON, AND Y. L. R. LEITE. 2016. Family Echimyidae (hutias, South American spiny-rats and coypu). Pp. 552–641 in Handbook of the Mammals of the World. Vol 6. Lagomorphs and Rodents I (D. E. Wilson, T. E. J. Lacher & R. A. Mittermeier, eds.). Lynx Edicions, Barcelona.

FIGUEIREDO, J., C. HOORN, P. VAN DER VEN, AND E. SOARES. 2009. Late Miocene onset of the Amazon River and the Amazon deep-sea fan: evidence from the Foz do Amazonas Basin. Geology 37:619–622.

FINER, M., C. N. JENKINS, S. L. PIMM, B. KEANE, AND C. ROSS. 2008. Oil and gas projects in the Western Amazon: threats to wilderness, biodiversity, and indigenous peoples. PLoS ONE 3:e2932.

FLYNN, J. J., AND A. R. WYSS. 1998. Recent advances in South American mammalian paleontology. Trends in Ecology & Evolution 13:449–454.

FRANKHAM, R. ET AL. 2012. Implications of different species concepts for conserving biodiversity. Biological Conservation 153:25–31.

FUJITA, M. K., A. D. LEACHÉ, F. T. BURBRINK, J. A. MCGUIRE, AND C. MORITZ. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends in Ecology & Evolution 27:480–488.

FUNK, W. C., J. P. CALDWELL, C. E. PEDEN, J. M. PADIAL, I. DE LA RIVA, AND D. C. CANNATELLA. 2007. Tests of biogeographic hypotheses for diversification in the Amazonian forest frog, *Physalaemus petersi*. Molecular Phylogenetics and Evolution 44:825–837.

GALEWSKI, T., J. F. MAUFFREY, Y. L. LEITE, J. L. PATTON, AND E. J. DOUZERY. 2005. Ecomorphological diversification among South American spiny rats (Rodentia; Echimyidae): a phylogenetic and chronological approach. Molecular Phylogenetics and Evolution 34:601–615.

GARDNER, A. L., AND L. H. EMMONS. 1984. Species groups in *Proechimys* (Rodentia, Echimyidae) as indicated by karyology and bullar morphology. Journal of Mammalogy 65:10–25.

GEOFFROY SAINT- HILAIRE, I. 1838. Notice sur les rongeurs épineux désignés par les auteurs sous les noms d'Echimys, Loncheres, Heteromys et Nelomys. Comptes rendus hebdomadaires des séances de l'Académie des sciences 6:884–888.

GEOFFROY SAINT- HILAIRE, I. 1840. Notice dur les rongeurs épineux, désignés par les auteurs sous les noms d'*Echimys*, *Loncheres*, *Heteromys* et *Nelomys*. Magasin de Zoologie, d'Anatomie Comparee et de Palaeontologie 12:1–57.

GEOFFROY SAINT-HILAIRE, É. 1803. Catalogue des Mammifères du Muséum National d'histoire naturelle. Muséum National d'Histoire Naturelle, Paris.

GIPPOLITI, S., F. P. D. COTTERILL, D. ZINNER, AND C. P. GROVES. 2018. Impacts of taxonomic inertia for the conservation of African ungulate diversity: an overview. Biological Reviews 93:115–130.

GODFRAY, H. C. J. 2002. Challenges for taxonomy. Nature 417:17–19.

GONZÁLEZ, F. 2018. El problema de la especie a 150 años de El origen. Pp. 122–162 in Filosofía de la Biología (P. C. Abrantes, ed.). 2nd edition. Editora do PPGFIL-UFRRJ, Seropedica.

GROVES, C., AND P. GRUBB. 2011. Ungulate Taxonomy. The Johns Hopkins University Press, Baltimore.

GRUBB, P. 2001. Case 3022. Catalogue des Mammiferes du Muséum National d'Histoire Naturelle By Étienne Geoffroy Saint-Hilaire (1803): proposed placement on the Official List of Works Approved as Available for Zoological Nomenclature. Bulletin of Zoological Nomenclature 58:41–52.

GÜNTHER, A. 1876. Report on some of the additions to the collection of Mammalia in the British Museum. 4. on some new mammals from tropical America. Proceedings of the general meetings for scientific business of the zoological society of London:743–751.

HAFFER, J. 1969. Speciation in Amazonian forest birds. Science 165:131–137.

HÄGGI, C. ET AL. 2017. Response of the Amazon rainforest to late Pleistocene climate variability. Earth and Planetary Science Letters 479:50–59.

HALL, E. R. 1981. The mammals of North America. 2nd edition. Wiley and Sons, New York.

HAMMEN, T. VAN DER, AND H. HOOGHIEMSTRA. 2000. Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. Quaternary Science Reviews 19:725–742.

HEBERT, P. D. N., A. CYWINSKA, S. L. BALL, AND J. R. DEWAARD. 2003a. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London. Series B, Biological Science 270:313–21.

HEBERT, P. D. N., AND T. R. GREGORY. 2005. The promise of DNA barcoding for taxonomy. Systematic Biology 54:852–859.

HEBERT, P. D. N., S. RATNASINGHAM, AND J. R. DE WAARD. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society of London. Series B, Biological Science 270:S96-99.

HEINE, C., J. ZOETHOUT, AND R. D. MÜLLER. 2013. Kinematics of the South Atlantic rift. Solid Earth 4:215–253.

HELLER, R., P. FRANDSEN, E. D. LORENZEN, AND H. R. SIEGISMUND. 2014. Is diagnosability an indicator of speciation? Response to "Why one century of phenetics is enough." Systematic Biology 63:833–837.

HENNIG, W. 1966. Phylogenetic systematics. University of Illinois Press, Urbana.

HERSHKOVITZ, P. 1948. Mammals of northern Colombia. Preliminary report no. 2: spiny rats (Echimyidae), with supplemental notes on related forms. Proceedings of the United States National Museum 97:125–140.

HERSHKOVITZ, P. 1955. Status of the generic name *Zorilla* (Mammalia): nomenclature by rule or by caprice. Proceedings of The Biological Society of Washington 68:185–191.

HILL, G. E. 2017. The mitonuclear compatibility species concept. The Auk 134:393–409.

HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. Annual Review of Ecology and Systematics 18:23–42.

HOFFSTETTER, R. 1972. Relationships, origins, and history of the ceboid monkeys and caviomorph rodents: a modern reinterpretation. Pp. 323–348 in Evolutionary Biology: Volume 6 (T. Dobzhansky, M. K. Hecht & W. C. Steere, eds.). Appleton-Century-Crofts, New York.

HOOGHIEMSTRA, H., AND T. VAN DER HAMMEN. 1998. Neogene and Quaternary development of the neotropical rain forest the forest refugia hypothesis, and a literature overview. Earth-Science Reviews 44:147–183.

HOORN, C. 1993. Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia: results of a palynostratigraphic study. Palaeogeography, Palaeoclimatology, Palaeoecology 105:267–309.

HOORN, C. ET AL. 2010a. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330:927–931.

HOORN, C. ET AL. 2017. The Amazon at sea: Onset and stages of the Amazon River from a marine record, with special reference to Neogene plant turnover in the drainage basin. Global and Planetary Change 153:51–65.

HOORN, C., AND S. FLANTUA. 2015. An early start for the Panama land bridge. Science 348.

HOORN, C., J. GUERRERO, G. A. SARMIENTO, AND M. A. LORENTE. 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. Geology 23:237.

HOORN, C., F. P. WESSELINGH, J. HOVIKOSKI, AND J. GUERRERO. 2010b. The development of the Amazonian mega-wetland (Miocene; Brazil, Colombia, Peru, Bolivia). Pp. 123–142 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

HOULE, A. 1999. The origin of platyrrhines: An evaluation of the Antarctic scenario and the floating island model. American Journal of Physical Anthropology 109:541–559.

HUCHON, D., F. M. CATZEFLIS, AND E. J. DOUZERY. 2000. Variance of molecular datings, evolution of rodents and the phylogenetic affinities between Ctenodactylidae and Hystricognathi. Proceedings of the Royal Society B: Biological Sciences 267:393–402.

HUCHON, D., AND E. J. DOUZERY. 2001. From the Old World to the New World: a molecular chronicle of the phylogeny and biogeography of hystricognath rodents. Molecular Phylogenetics and Evolution 20:238–251.

HULL, D. L. 1997. The ideal species concept - and why we can't get it. Pp. 357–380 in Species: the units of diversity (M. F. Claridge, H. A. Dawah & M. R. Wilson, eds.). Chapman & Hall, London.

ICZN. 2001. Opinion 1984 (Case 3121). *Holochilus* Brandt, 1835, *Proechimys* J.A. Allen, 1899 and *Trinomys* Thomas, 1921 (Mammalia, Rodentia): conserved by the designation of *H. sciureus* Wagner, 1842 as the type species of *Holochilus*. Bulletin of Zoological Nomenclature 58:245–246.

JARAMILLO, C. ET AL. 2010. The origin of the modern Amazon rainforest: implications of the palynological and palaeobotanical record. Pp. 317–334 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

JOHNS, G. C., AND J. C. AVISE. 1998. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochome b gene. Molecular Biology and Evolution 15:1481–1490.

KAANDORP, R. J. G., H. B. VONHOF, F. P. WESSELINGH, L. R. PITTMAN, D. KROON, AND J. E. VAN HINTE. 2005. Seasonal Amazonian rainfall variation in the Miocene Climate Optimum. Palaeogeography, Palaeoclimatology, Palaeoecology 221:1–6.

KASAHARA, S., AND Y. YONENAGA-YASSUDA. 1984. A progress report of cytogenetic data on Brazilian rodents. Brazilian Journal of Genetics 7:509–533.

KROONENBERG, S. B., AND E. W. F. DE ROEVER. 2010. Geological evolution of the Amazonian Craton. Pp. 9–28 in Amazonia: Landscape and Species Evolution (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

LANZONE, C., R. A. OJEDA, AND M. H. GALLARDO. 2007. Integrative taxonomy, systematics and distribution of the genus *Eligmodontia* (Rodentia, Cricetidae, Sigmodontinae) in the temperate Monte Desert of Argentina. Mammalian Biology 72:299–312.

LARA, M. C., J. L. PATTON, AND M. N. F. DA SILVA. 1996. The simultaneous diversification of South American echimyid rodents (Hystricognathi) based on complete citochrome b sequences. Molecular Phylogenetics and Evolution 5:403–413.

LARA, M., AND J. L. PATTON. 2000. Evolutionary diversification of spiny rats (genus *Trinomys*, Rodentia: Echimyidae) in the Atlantic Forest of Brazil. Zoological Journal of the Linnean Society 130:661–686.

LARA, M., J. L. PATTON, AND E. HINGST-ZAHER. 2002. *Trinomys mirapitanga*, a new species of spiny rat (Rodentia: Echimyidae) from the Brazilian Atlantic Forest. Mammalian Biology 67:233–242.

LAVOCAT, R. 1969. La systematique des rongeurs hystricomorphes et la derive des continents. Comptes Rendus de l'Académie des Sciences - Series D - (Sciences Naturelles) 269:1496–1497.

LEE, M. S. Y. 2003. Species concepts and species reality: salvaging a Linnaean rank. Journal of Evolutionary Biology 16:179–188.

LEITE, R. N., S.-O. KOLOKOTRONIS, F. C. ALMEIDA, F. P. WERNECK, D. S. ROGERS, AND M. WEKSLER. 2014. In the wake of invasion: tracing the historical biogeography of the South American cricetid radiation (Rodentia, Sigmodontinae). PLoS ONE 9:e100687.

LEITE, R. N., AND D. S. ROGERS. 2013. Revisiting Amazonian phylogeography: insights into diversification hypotheses and novel perspectives. Organisms Diversity & Evolution 13:639–664.

LEITE, R. N., M. N. F. DA SILVA, AND D. S. ROGERS. 2015. Delimitação das espécies de ratos-de-espinho neotropicais: em busca de uma taxonomia integrativa para o gênero *Proechimys*. P. 378 in Livro de Resumos do 8º Congresso Brasileiro de Mastozoologia. Sociedade Brasileira de Mastozoologia, João Pessoa, Paraíba.

LEITE, Y. L. ., AND J. L. PATTON. 2002. Evolution of South American spiny rats (Rodentia, Echimyidae): the star-phylogeny hypothesis revisited. Molecular Phylogenetics and Evolution 25:455–464.

LESSA, E. P., J. A. COOK, AND J. L. PATTON. 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. Proceedings of the National Academy of Sciences of the United States of America 100:10331–4.

LINARDI, P. M., J. R. BOTELHO, J. A. RAFAEL, C. M. C. CALLE, A. CUNHA, AND P. A. R. MACHADO. 1991. Ectoparasites of small mammals in the Ilha de Maracá, Roraima, Brazil. I. Ectoparasite fauna and geographical and host records. Acta Amazonica 21:131–140.

LONG, J. L. 2003. Rodentia. Pp. 133–236 in Introduced mammals of the world: their history, distribution and influence (J. L. Long, ed.). CABI Publishing, Wallingford.

LOUGHEED, S. C., C. GASCON, D. A. JONES, J. P. BOGART, AND P. T. BOAG. 1999. Ridges and rivers: a test of competing hypotheses of Amazonian diversification using a dart-poison frog (*Epipedobates femoralis*). Proceedings of the Royal Society of London. Series B: Biological Sciences 266:1829–1835.

LUCKETT, W. P., AND J.-L. HARTENBERGER. 1985. Evolutionary relationships among rodents: comments and conclusions. Pp. 685–712 in Evolutionary relationships among rodents: A Multidisciplinary Analysis (W. P. Luckett & J.-L. Hartenberger, eds.). Springer, Boston.

LUCKETT, W. P., AND J.-L. HARTENBERGER. 1993. Monophyly or polyphyly of the order Rodentia: Possible conflict between morphological and molecular interpretations. Journal of Mammalian Evolution 1:127–147.

LUNDBERG, J. G., L. G. MARSHALL, J. GUERRERO, B. K. HORTON, M. C. S. L. MALABARBA, AND F. WESSELINGH. 1998. The stage for neotropical fish diversification: A history of tropical South American rivers. Pp. 13–48 in Phylogeny and Classification of Neotropical Fishes (L. R. Malabarba, R. E. Reis, R. P. Vari, Z. M. Lucena & C. A. S. Lucena, eds.). Edipucrs, Porto Alegre.

MACHADO, M. X. 2017. Cariologia e filogenia molecular em *Trinomys* e *Proechimys* (Echimyidae, Rodentia). Universidade Federal do Espírito Santo.

MACHADO, T., M. J. DE J. SILVA, E. R. LEAL-MESQUITA, A. P. CARMIGNOTTO, AND Y. YONENAGA-YASSUDA. 2005. Nine karyomorphs for spiny rats of the genus *Proechimys* (Echimyidae, Rodentia) from North and Central Brazil. Genetics and Molecular Biology 28:682–692.

MADDISON, W. P. 1997. Gene trees in species trees. Systematic Biology 46:523–536.

MALCOLM, J. R. 1992. Use of tooth impressions to identify and age live *Proechimys guyannensis* and *P. cuvieri* (Rodentia: Echimyidae). Journal of Zoology 227:537–546.

MALLET, J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Philos Trans R Soc Lond B Biol Sci 363:2971–2986.

MALLET, J., AND K. WILLMOTT. 2003. Taxonomy: renaissance or Tower of Babel? Trends in Ecology & Evolution 18:57–59.

MARIVAUX, L., M. VIANEY-LIAUD, J. L. WELSOMME, AND J. J. JAEGER. 2002. The role of Asia in the origin and diversification of hystricognathous rodents. Zoologica Scripta 31:225–239.

MASSATTI, R., AND L. L. KNOWLES. 2016. Contrasting support for alternative models of genomic variation based on microhabitat preference: species-specific effects of climate change in alpine sedges. Molecular Ecology 25:3974–3986.

MATOCQ, M. D., J. L. PATTON, AND M. N. F. DA SILVA. 2000. Population genetic structure of two ecologically distinct amazonian spiny rats: separating history and current ecology. Evolution 54:1423–1432.

MAYDEN, R. L. 1997. A hierarchy of species concepts: the denouement in the saga of the species problem. Pp. 381–431 in Species: the units of diversity (M. F. Claridge, H. A. Dawah & M. R. Wilson, eds.). Chapman & Hall, London.

MAYR, E. 1940. Speciation Phenomena in Birds. The American Naturalist 74:249–278.

MAYR, E. 1942. Systematics and origin of species. Columbia University Press, New York.

MAYR, E. 1969. Principles of systematic zoology. Principles of systematic zoology. New York (& London): McGraw Hill, New York.

MAYR, E. 1982. The growth of biological thought: diversity, evolution, and inheritance. Belknap Press of Harvard University Press, Cambridge.

MAYR, E. 1992. A local flora and the biological species concept. American Journal of Botany 79:222–238.

MEIER, R., AND R. WILLMANN. 2000. The Hennigian Species Concept. Pp. 30–43 in Species concepts and phylogenetic theory: a debate (Q. D. Wheeler & R. Meier, eds.). Columbia University Press, New York.

MENDES-OLIVEIRA, A. C., P. E. D. BOBROWIEC, C. R. SILVA, L. C. PINTO, AND L. C. TREVELIN. 2015. Estudos sobre pequenos mamíferos não-voadores da Amazônia brasileira: amostragens e lacunas de conhecimento. Pp. 21–50 in Pequenos mamíferos não-voadores da Amazônia brasileira, Série Livros: 2 (A. C. Mendes-Oliveira & C. L. Miranda, eds.). Sociedade Brasileira de Mastozoologia, Rio de Janeiro.

MENG, J., A. R. WYSS, M. R. DAWSON, AND R. ZHAI. 1994. Primitive fossil rodent from Inner Mongolia and its implications for mammalian phylogeny. Nature 370:134–136.

MILLER, G. S., AND J. A. REHN. 1902. Systematic results of the study of North American land mammals to the close of the year 1900. Proceedings of the Boston Society of Natural History 30:1–352.

MISHLER, B. D., AND M. J. DONOGHUE. 1982. Species concepts: a case for pluralism. Systematic Zoology 31:491–503.

MISHLER, B. D., AND E. C. THERIOT. 2000a. A critique from the Mishler and Theriot phylogenetic species concept perspective: monophyly, apomorphy, and phylogenetic species concepts. Pp. 119–132 in Species concepts and phylogenetic theory: a debate (Q. D. Wheeler & R. Meier, eds.). Columbia University Press, New York.

MISHLER, B. D., AND E. C. THERIOT. 2000b. The phylogenetic species concept (sensu Mishler and Theriot): monophyly, apomorphy, and phylogenetic species concepts. Pp. 44–54 in Species concepts and phylogenetic theory: a debate (Q. D. Wheeler & R. Meier, eds.). Columbia University Press, New York.

MONTES, C. ET AL. 2015. Middle Miocene closure of the Central American Seaway. Science 348:226–9.

MONTGELARD, C., E. FORTY, V. ARNAL, AND C. A. MATTHEE. 2008. Suprafamilial relationships among Rodentia and the phylogenetic effect of removing fast-evolving nucleotides in mitochondrial, exon and intron fragments. BMC Evolutionary Biology 8:321.

MOOJEN, J. 1948. Speciation in the Brazilian spiny rats (genus *Proechimys*, Family Echimyidae). University of Kansas Publications, Museum Natural History 1:303–401.

MORA, A. ET AL. 2010. Tectonic history of the Andes and sub-Andean zones: Implications for the development of the Amazon drainage basin. Pp. 38–60 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

MOULIN, M., D. ASLANIAN, AND P. UNTERNEHR. 2010. A new starting point for the South and Equatorial Atlantic Ocean. Earth-Science Reviews 98:1–37.

MUSSER, G. G., AND M. D. CARLETON. 2005. Order Rodentia. Pp. 745–752 in Mammal Species of the World: A Taxonomic and Geographic Reference - volume 2 (D. E. Wilson & D. M. Reeder, eds.). The Johns Hopkins University Press, Baltimore.

MYERS, N., R. A. MITTERMEIER, C. G. MITTERMEIER, G. A. B. DA FONSECA, AND J. KENT. 2000. Biodiversity hotspots for conservation priorities. Nature 403:853–858.

NELSON, G. J., AND N. I. PLATNICK. 1981. Systematics and biogeography: cladistics and vicariance. Columbia University Press, New York.

O'HARA, R. J. 1993. Systematic generalization, historical fate, and species problem. Systematic Biology 42:231–246.

OREN, A. 2004. Prokaryote diversity and taxonomy: current status and future challenges. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences 359:623–38.

PADIAL, J. M., A. MIRALLES, I. DE LA RIVA, AND M. VENCES. 2010. The integrative future of taxonomy. Frontiers in Zoology 7:16.

PANTE, E., C. SCHOELINCK, AND N. PUILLANDRE. 2015. From integrative taxonomy to species description: one step beyond. Systematic Biology 64:152–160.

PATTON, J. L. 1987. Species groups of spiny rats, genus Proechimys (Rodentia : Echimyidae). Fieldiana Zoology 39:305–346.

PATTON, J. L., AND A. L. GARDNER. 1972. Notes on the systematics of *Proechimys* (Rodentia, Echimyidae), with emphasis on peruvian forms. Occasional Papers 13:1–30.

PATTON, J. L., AND M. A. ROGERS. 1983. Systematic implications of non-geographic variation in the spiny rat genus *Proechimys* (Echimyidae). Zeitschrift für Säugetierkunde 48:363–370.

PATTON, J. L., M. N. DA SILVA, AND J. R. MALCOLM. 2000. Mammals of the rio Juruá and the evolutionary and ecological diversification of Amazonia. Bulletin American Museum of Natural History 244:1–306.

PATTON, J. L., AND M. F. SMITH. 1992. mtDNA Phylogeny of Andean mice: a test of diversification across ecological gradients. Evolution 46:174–183.

PATTON, J. P., AND R. N. LEITE. 2015. Genus *Proechimys* J. A. Allen, 1889. Pp. 950–989 in Mammals of South America, Volume 2: Rodents (J. L. Patton, U. F. J. Pardiñas & G. D'Elía, eds.). University Of Chicago Press, Chicago.

PEČNEROVÁ, P., AND N. MARTÍNKOVÁ. 2012. Evolutionary history of tree squirrels (Rodentia, Sciurini) based on multilocus phylogeny reconstruction. Zoologica Scripta 41:211–219.

PESSÔA, L. M., W. C. TAVARES, J. A. DE OLIVEIRA, AND J. L. PATTON. 2015. Genus *Trinomys* Thomas, 1921. Pp. 999–1019 in Mammals of South America, Volume 2: Rodents (J. L. Patton, U. F. J. Pardiñas & G. D'Elía, eds.). The University of Chicago Press, Chicago and London.

PICTET, F. J. 1841. Premiere notice sur les animaux nouveaux ou peu connus du Musée de Genève. Observations sur quelques rongeurs épineux du Brésil. Memoires de la Société de physique et d'histoire naturelle de Genève 9:143–160.

PONS, D., AND D. DE FRANCESCHI. 2007. Neogene woods from western Peruvian Amazon and palaeoenvironmental interpretation. Bulletin of Geosciences 82:343–354.

POUX, C., P. CHEVRET, D. HUCHON, W. W. DE JONG, AND E. J. DOUZERY. 2006. Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. Systematic Biology 55:228–244.

PRADO, J. R. DO, AND A. R. PERCEQUILLO. 2018. Systematic studies of the genus *Aegialomys* Weksler et al., 2006 (Rodentia: Cricetidae: Sigmodontinae): geographic variation, species delimitation, and biogeography. Journal of Mammalian Evolution 25:71–118.

PUCKETT, E. E. ET AL. 2016. Global population divergence and admixture of the brown rat (*Rattus norvegicus*). Proceedings of the Royal Society of London B: Biological Sciences 283.

DE QUEIROZ, K. 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. Pp. 57–75 in Endless forms: species and speciation (D. J. Howard & S. H. Berlocher, eds.). Oxford University Press, New York and Oxford.

DE QUEIROZ, K. 2005a. A Unified Concept of Species and Its Consequences for the Future of Taxonomy. Proceedings of the California Academy of Sciences 4:196–215.

DE QUEIROZ, K. 2005b. Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences of the United States of America 102 Suppl:6600–6607.

DE QUEIROZ, K. 2007. Species concepts and species delimitation. Systematic Biology 56:879–886.

RENGGER, J. R. 1830. Naturgeschichte der Säugethiere von Paraguay. Schweighauserschen Buchhandlumg, Basel.

RESENDE-MOREIRA, L. C. ET AL. 2019. Evolving in isolation: Genetic tests reject recent connections of Amazonian savannas with the central Cerrado. Journal of Biogeography 46:196–211.

RIDLEY, M. 2004. Species concepts and intraspecific variation. Pp. 345–380 in Evolution. 3rd edition. Blackwell Science Ltd, Malden.

RODRIGUES DA COSTA, M. J. ET AL. 2016. Cryptic species in Proechimys goeldii (Rodentia, Echimyidae)? A case of molecular and chromosomal differentiation in allopatric Populations. Cytogenetic and genome research 148:199–210.

RONAGHI, M., M. UHLÉN, AND P. NYRÉN. 1998. A sequencing method based on real-time pyrophosphate. Science 281:363–365.

ROSS, K. G., D. GOTZEK, M. S. ASCUNCE, AND D. D. SHOEMAKER. 2010. Species delimitation: a case study in a problematic ant taxon. Systematic Biology 59:162–184.

ROSSETTI, D. DE F., P. MANN DE TOLEDO, AND A. M. GÓES. 2005. New geological framework for Western Amazonia (Brazil) and implications for biogeography and evolution. Quaternary Research 63:78–89.

ROWE, D. L., K. A. DUNN, R. M. ADKINS, AND R. L. HONEYCUTT. 2010. Molecular clocks keep dispersal hypotheses afloat: evidence for trans-Atlantic rafting by rodents. Journal of Biogeography 37:305–324.

SÁNCHEZ-VENDIZÚ, P., J. A. COOK, J. WOOD, AND J. SALAZAR-BRAVO. 2018. First record of *Proechimys pattoni* da Silva, 1998 (Rodentia, Echimyidae) in northwestern Bolivia. Check List 14:1115–1121.

SCHETINO, M. A. A. 2008. Código de barras de DNA em espécies de *Proechimys* (Rodentia: Echimyidae) da Amazônia. Instituto Nacional de Pesquisas da Amazônia.

SCHLICK-STEINER, B. C., F. M. STEINER, B. SEIFERT, C. STAUFFER, E. CHRISTIAN, AND R. H. CROZIER. 2010. Integrative taxonomy: a multisource approach to exploring biodiversity. Annual Review of Entomology 55:421–438.

SHANKER, K., S. P. VIJAYAKUMAR, AND K. N. GANESHAIAH. 2017. Unpacking the species conundrum: philosophy, practice and a way forward. Journal of Genetics 96:413–430.

SHAW, A. J., AND B. ALLEN. 2000. Phylogenetic relationships, morphological incongruence, and geographic speciation in the Fontinalaceae (Bryophyta). Molecular Phylogenetics and Evolution 16:225–237.

SILVA, C. R. ET AL. 2018. The role of Pleistocene climate change in the genetic variability, distribution and demography of Proechimys cuvieri and *P. guyannensis* (Rodentia: Echimyidae) in northeastern Amazonia. PLoS One 13:e0206660.

DA SILVA, M. N., AND J. L. PATTON. 1998. Molecular phylogeography and the evolution and conservation of amazonian mammals. Mol Ecol 7:475–486.

DA SILVA, M. N. F. 1998. Four new species of spiny rats of the genus *Proechimys* (Rodentia: Echimyidae) from the western Amazon of Brazil. Proceedings of the Biological Society of Washington 111:436–471.

SIMPSON, G. G. 1961. Principles of animal taxonomy. Columbia University Press, New York.

SITES, J. W., AND J. C. MARSHALL. 2004. Operational criteria for delimiting species. Annual Review of Ecology, Evolution, and Systematics 35:199–227.

VAN SOELEN, E. E. ET AL. 2017. A 30 Ma history of the Amazon River inferred from terrigenous sediments and organic matter on the Ceará Rise. Earth and Planetary Science Letters 474:40–48.

SOKAL, R. R. 1962. Typology and empiricism in taxonomy. Journal of Theoretical Biology 3:230–267.

SOLÍS-LEMUS, C., L. L. KNOWLES, AND C. ANÉ. 2015. Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution 69:492–507.

TER STEEGE, H. ET AL. 2003. A spatial model of tree α-diversity and tree density for the Amazon. Biodiversity and Conservation 12:2255–2277.

STEINER, C., P. SOURROUILLE, AND F. CATZEFLIS. 2000. Molecular characterization and mitochondrial sequence variation in two sympatric species of *Proechimys* (Rodentia: Echimyidae) in French Guiana. Biochemical Systematics and Ecology 28:963–973.

SUKUMARAN, J., AND L. L. KNOWLES. 2017. Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences of the United States of America 114:1607–1612.

TATE, G. H. H. 1935. The taxonomy of the genera of neotropical hystricoid rodents. Bulletin of the American Museum of Natural History 68:295–447.

TAUTZ, D., P. ARCTANDER, A. MINELLI, R. H. THOMAS, AND A. P. VOGLER. 2003. A plea for DNA taxonomy. Trends in Ecology & Evolution 18:70–74.

THOMAS, O. 1921. On spiny rats of the Proechimys group from southeastern Brazil. The Annals and Magazine of Natural History; Zoology, Botany, and Geology 9:140–143.

TOMES, R. F. 1860. Notes on a third collection of Mammalia made by Mr. Fraser in the Republic of Ecuador. Proceedings of the Zoological Society of London:206–268.

TROUESSART, E. L. 1880. Catalogue des mammifères vivants et fossils. Ordre des rongeurs. Bulletin de la Société d'Études scientifiques d'Angers 10:58–212.

UPHAM, N. S., AND B. D. PATTERSON. 2015. Evolution of caviomorph rodents: a complete phylogeny and timetree for living genera. Pp. 63–120 in Biology of caviomorph rodents: diversity and evolution (A. I. Vassallo & D. Antenucci, eds.). SAREM Series A, Buenos Aires.

VALDECASAS, A. G., D. WILLIAMS, AND Q. D. WHEELER. 2007. 'Integrative taxonomy' then and now: a response to Dayrat (2005). Biological Journal of the Linnean Society 93:211–216.

VOLOCH, C. M., J. F. VILELA, L. LOSS-OLIVEIRA, AND C. G. SCHRAGO. 2013. Phylogeny and chronology of the major lineages of New World hystricognath rodents: insights on the biogeography of the Eocene/Oligocene arrival of mammals in South America. BMC Research Notes 6:160.

VOSS, R. S., AND N. I. ABRAMSON. 1999. Case 3121. *Holochilus* Brandt, 1835, *Proechimys* J.A. Allen, 1899 and *Trinomys* Thomas, 1921 (Mammalia, Rodentia): proposed conservation by the designation of *H. sciureus* Wagner, 1842 as the type species of *Holochilus*. Bulletin of Zoological Nomenclature 56:255–261.

VOSS, R. S., D. P. LUNDE, AND N. B. SIMMONS. 2001. The mammals of Paracou, French Guiana: a neotropical lowland rainforest fauna part 2. Nonvolant species. Bulletin of the American Museum of Natural History 263:3–236.

VAN VUUREN, B. J., S. KINET, J. CHOPELET, AND F. CATZEFLIS. 2004. Geographic patterns of genetic variation in four Neotropical rodents: conservation implications for small game mammals in French Guiana. Biological Journal of the Linnean Society 81:203–218.

WADE, L. 2015. Feature: How the Amazon became a crucible of life. Science News:DOI: 10.1126/science.aad4781.

WALLACE, A. R. 1854. On the Monkeys of the Amazon. Annals and Magazine of Natural History 14:451–454.

WARD, D. M. 1998. A natural species concept for prokaryotes. Current Opinion in Microbiology 1:271–277.

WEKSLER, M., C. R. BONVICINO, I. B. OTAZU, AND J. S. SILVA. 2001. Status of *Proechimys roberti* and *P. oris* (Rodentia: Echimyidae) from eastern Amazonia and Central Brazil. Journal of Mammalogy 82:109–122.

WESSELINGH, F. P. ET AL. 2001. Lake Pebas: a palaeoecological reconstruction of a Miocene, long-lived lake complex in western Amazonia. Cainozoic Research 1:35–81.

WESSELINGH, F. P. 2006. Miocene long-lived lake Pebas as a stage of mollusc radiations, with implications for landscape evolution in western Amazonia. Scripta Geologica 133:1–17.

WHEELER, Q. D. 2004. Taxonomic triage and the poverty of phylogeny. Philosophical Transactions of the Royal Society of London. Series B, Biological sciences 359:571–583.

WILEY, E. O. 1978. The evolutionary species concept reconsidered. Systematic Zoology 27:17.

WILKINS, J. S. 2009. Species: a history of the idea. University of California Press, Berkeley and Los Angeles.

WILL, K. W., B. D. MISHLER, AND Q. D. WHEELER. 2005. The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology 54:844–851.

WILL, T. M., AND H. E. FRIMMEL. 2018. Where does a continent prefer to break up? Some lessons from the South Atlantic margins. Gondwana Research 53:9–19.

WILSON, D. E., AND D. M. REEDER. 2005. Mammal species of the world: a taxonomic and geographic reference. The Johns Hopkins University Press, Baltimore.

WOODS, A. E. 1980. The origin of the caviomorph rodents from a source in middle America: a clue to the area of origin of the platyrrhine primates. Pp. 79–92 in Evolutionary Biology of the New World Monkeys and Continental Drift (Advances in Primatology) (R. L. Ciochon & A. B. Chiarelli, eds.). Plenum Press, New York and London.

WOODS, A. E., AND B. PATTERSON. 1959. The rodents in the Deseaden Oligocene of Patagonia and the beginnings of South American rodent evolution. Bulletin of the Museum of Comparative Zoology 120:279–428.

WOODS, AND KILPATRICK. 2005. Infraorder Hystricognathi Brandt, 1855. Pp. 1538–1600 in Mammal Species of the World. A Taxonomic and Geographic Reference (D. E. Wilson & D. M. Reeder, eds.). Volume 2. Johns Hopkins University Press, Baltimore.

WU, S. ET AL. 2012. Molecular and paleontological evidence for a post-Cretaceous origin of rodents. PLoS ONE 7:e46445.

YANG, Z., AND B. RANNALA. 2010. Bayesian species delimitation using multilocus sequence data. Proceedings of the National Academy of Sciences of the United States of America 107:9264–9269.

YEATES, D. K., A. SEAGO, L. NELSON, S. L. CAMERON, L. JOSEPH, AND J. W. H. TRUEMAN. 2011. Integrative taxonomy, or iterative taxonomy? Systematic Entomology 36:209–217.

ZACHOS, F. E. ET AL. 2013. Species inflation and taxonomic artefacts—A critical comment on recent trends in mammalian classification. Mammalian Biology 78:1–6.

ZACHOS, F. E. 2016a. An annotated list of species concepts. Pp. 77–96 in Species concepts in biology: historical development, theoretical foundations and practical relevance. Springer International Publishing Switzerland, Cham.

ZACHOS, F. E. 2016b. Species delimitation: discrete names in a continuous world with fuzzy boundaries. Pp. 143–162 in Species concepts in biology: historical development, theoretical foundations and practical relevance. Springer International Publishing Switzerland, Cham.

ZACHOS, F. E. 2016c. Tree thinking and species delimitation: Guidelines for taxonomy and phylogenetic terminology. Mammalian Biology 81:185–188.

ZACHOS, F. E. 2018a. Mammals and meaningful taxonomic units: the debate about species concepts and conservation. Mammal Review 48:153–159.

ZACHOS, F. E. 2018b. (New) Species concepts, species delimitation and the inherent limitations of taxonomy. Journal of Genetics 97:811–815.

Tables

Table 1: List of 65 nominal taxa for the species group of the genus *Proechimys*, organized in alphabetical order with information about the author, year and museum where the types are housed. There is no information on where holotype for *longicaudatus* is housed. **AMNH**: American Museum of Natural History, New York, New York, USA; **FMNH**: Field Museum of Natural History, Chicago, Illinois, USA; **INPA**: Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil; **MCZ**: Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, USA; **MHN**: Musee National d'Histoire Naturalle, Paris, France; **MN:** Coleção de Mamíferos do Museu Nacional da Universidade Federal do Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil; **NHM**: Natural History Museum, London, UK; **UMMZ**: University of Michigan Museum of Zoology, Ann Arbor, Michigan, USA; **USNM**: National Museum of Natural History, Smithsonian Institution, Washington, District of Columbia, USA; **ZINRAS:** Zoological Institute, Russian Academy of Sciences, Saint Petersburg, Russia. UK: United Kingdom; USA: United States of America.

Nominal taxa	Authors	Year	Museum
amphichoricus	Moojen	1948	AMNH, New York, USA
arabupu	Moojen	1948	AMNH, New York, USA
arescens	Osgood	1944	FMNH, Chicago, USA
boimensis	Allen	1916	AMNH, New York, USA
bolivianus	Thomas	1912	NHM, London, UK
brevicauda	Gunther	1876	NHM, London, UK
burrus	Bangs	1901	MCZ, Cambridge, USA
calidior	Thomas	1911	NHM, London, UK
canicollis	Allen	1899	AMNH, New York, USA
cayennensis	Desmerast	1817	MHN, Paris, France
centralis	Thomas	1896	NHM, London, UK
cherriei	Thomas	1899	NHM, London, UK
chiriquinus	Thomas	1900	NHM, London, UK
chrysaeolus	Thomas	1898	NHM, London, UK
colombianus	Thomas	1914	NHM, London, UK
cuvieri	Petter	1978	MHN, Paris, France
decumanus	Thomas	1899	NHM, London, UK
echinothrix	Da Silva	1998	INPA, Manaus, Brazil
elassopus	Osgood	1944	FMNH, Chicago, USA
gardneri	Da Silva	1998	INPA, Manaus, Brazil
goeldii	Thomas	1905	NHM, London, UK
goldmani	Bangs	1937	UMMZ, Ann Arbor, USA
gorgonae	Bangs	1905	MCZ, Cambridge, USA
guairae	Thomas	1901	USNM, Washington, USA

Table 1: Continuation.

Nominal taxa	Authors	Year	Museum
gularis	Thomas	1911	NHM, London, UK
guyannensis	Geoffroy	1803	MHN, Paris, France
hendeei	Thomas	1926	NHM, London, UK
hilda	Thomas	1924	NHM, London, UK
hoplomyoides	Tate	1939	AMNH, New York, USA
hyleae	Moojen	1948	MCZ, Cambridge, USA
ignotus	Kellogg	1946	USNM, Washington, USA
kermiti	Allen	1915	AMNH, New York, USA
kulinae	Da Silva	1998	INPA, Manaus, Brazil
langsdorffii	Brandt	1855	ZINRAS, St. Petersburg, Russia
leioprimna	Moojen	1948	FMNH, Chicago, USA
leucomystax	Ribeiro	1914	MN, Rio de Janeiro, Brazil
liminalis	Moojen	1948	MN, Rio de Janeiro, Brazil
longicaudatus	Rengger	1830	Unknown
magdalenae	Hershkovitz	1948	USNM, Washington, USA
mincae	Allen	1899	AMNH, New York, USA
nesiotes	Moojen	1948	FMNH, Chicago, USA
nigrofulvus	Osgood	1944	FMNH, Chicago, USA
ochraceus	Osgood	1912	FMNH, Chicago, USA
oconnelli	Allen	1913	AMNH, New York, USA
oris	Thomas	1912	NHM, London, UK
pachita	Thomas	1923	NHM, London, UK
panamensis	Thomas	1900	NHM, London, UK
pattoni	Da Silva	1998	INPA, Manaus, Brazil
poliopus	Osgood	1914	FMNH, Chicago, USA
quadruplicatus	Hershkovitz	1948	UMMZ, Ann Arbor, USA
rattinus	Thomas	1926	NHM, London, UK
ribeiroi	Moojen	1948	MN, Rio de Janeiro, Brazil
riparum	Moojen	1948	AMNH, New York, USA
roberti	Thomas	1901	NHM, London, UK
rosa	Thomas	1900	NHM, London, UK
rubellus	Hollister	1914	USNM, Washington, USA
securus	Thomas	1902	NHM, London, UK
semispinosus	Tomes	1860	NHM, London, UK
simonsi	Thomas	1900	NHM, London, UK
steerei	Goldman	1911	USNM, Washington, USA
trinitatis	Allen and Chapman	1893	AMNH, New York, USA
urichi	Allen	1899	AMNH, New York, USA

Table 1: Continuation.

Nominal taxa	Authors	Year	Museum
vacillator	Thomas	1903	NHM, London, UK
villicauda	Moojen	1948	MN, Rio de Janeiro, Brazil
warreni	Thomas	1905	NHM, London, UK

Table 2: Taxonomic arrangements proposed by different authors for the genus *Proechimys*. Genera and subgenera names are in **bold**, species and subspecies names are in *italics,* underlined taxa do not belong to *Proechimys,* according to the latest taxonomic classification, and names in parentheses indicates the genus that nominal taxa belong currently. Nominal taxa left-aligned correspond to species (single name) or subspecies (double name) and the center-aligned names are synonyms of the name above.

GEOFFROY (1838) [1]	GEOFFROY (1840)	PICTET (1841)	TROUESSART (1880) [1]	ALLEN & CHAPMAN (1899) [1]	THOMAS (1921) [2]
Echimys	**Echimys**	**Echimys**	**Echimys (Echimys)**	**Proechimys**	**Proechimys (Proechimys)**
cayennensis	*cayennensis*	*cayennensis*	*cayennensis*	*albispinus (=Trinomys)*	*dimidiatus (=Trinomys)*
	guyannensis	*guyannensis*	*longicaudatus*	*canicollis*	*iheringi (=Trinomys)*
		longicaudatus	*semispinosus*	*cayennensis*	*roberti*
				centralis	
				cherriei	
Loncheres	**Loncheres**		**Echimys (Thrichomys)**	*chrysaeolus*	**Proechimys (Trinomys)**
longicaudatus	*myosuros (=Trinomys)*		*brevicauda*	*decumanus*	*albispinus albispinus (=Trinomys)*
	longicaudatus			*dimidiatus (=Trinomys)*	*albispinus sertonius (=Trinomys)*
				ferrugineus (=Mesomys)	*setosus (=Trinomys)*
				gymnurus (=Hoplomys)	*cinnamomeus (=Trinomys)*
				hispidus (=Mesomys)	*elegans (=Trinomys)*
				mincae	*fuliginosus (=Trinomys)*
				semispinosus	*leptosoma (=Trinomys)*
				setosus (=Trinomys)	*myosuros (=Trinomys)*
				trinitatis	
				urichi	

[1] Did not mention the nominal taxa *guyannensis*.

[2] Only included the *Proechimys* forms from Southeastern Brazil, all others forms from South and Central America were included in subgenus *Proechimys.*

Table 2: Continuation.

ELLERMAN (1940) [3]	HERSHKOVITZ (1948) [4]	MOOJEN (1948) [5]	CABRERA (1961)
P. (Proechimys)	**P. (Proechimys)**	**P. (Proechimys)**	**P. (Proechimys)**
cayennensis GROUP	*canicollis*	*canicollis*	*canicollis canicollis*
cayennensis bolivianus	*iheringi (=Trinomys)*	*goeldii goeldii*	*canicollis vacillator*
cayennensis brevicauda	*hendeei*	*goeldii steerei*	*goeldii goeldii*
cayennensis burrus	*guyannensis boimensis*	*guyannensis arabupu*	*goeldii steerei*
cayennensis calidior	*guyannensis bolivianus*	*guyannensis arescens*	*guyannensis arabupu*
cayennensis cayennensis	*guyannensis brevicauda*	*guyannensis bolivianus*	*guyannensis arescens*
cayennensis centralis	*guyannensis burrus*	*guyannensis cherriei*	*guyannensis bolivianus*
cayennensis cherriei	*guyannensis calidior*	*guyannensis chrysaeolus*	*guyannensis chrysaeolus*
cayennensis chrysaeolus	*guyannensis centralis*	*guyannensis guairae*	*guyannensis colombianus*
cayennensis colombianus	*guyannensis cherriei*	*guyannensis guyannensis*	*guyannensis decumanus*
cayennensis decumanus	*guyannensis chrysaeolus*	*guyannensis hyleae*	*guyannensis gorgonae*
cayennensis goeldii	*guyannensis colombianus*	*guyannensis leioprimna*	*guyannensis guairae*
cayennensis gorgonae	*guyannensis decumanus*	*guyannensis mincae*	*guyannensis gularis*
cayennensis guairae	*guyannensis goeldii*	*guyannensis nesiotes*	*guyannensis guyannensis*
cayennensis gularis	*guyannensis gorgonae*	*guyannensis ochraceus*	*guyannensis hyleae*
cayennensis hilda	*guyannensis guairae*	*guyannensis oconnelli*	*guyannensis leioprimna*
cayennensis longicaudatus	*guyannensis gularis*	*guyannensis oris*	*guyannensis magdalenae*
cayennensis mincae	*guyannensis guyannensis*	*guyannensis poliopus*	*guyannensis mincae*
cayennensis oris	*cayennensis*	*guyannensis ribeiroi*	*guyannensis nesiotes*
cayennensis pachita	*guyannensis hilda*	*guyannensis riparium*	*guyannensis ochraceus*
cayennensis panamensis	*guyannensis kermiti*	*guyannensis trinitatis*	*guyannensis oconnelli*
cayennensis roberti	*guyannensis leucomystax*	*guyannensis urichi*	*guyannensis oris*
cayennensis rosa	*guyannensis longicaudatus*	*guyannensis villicauda*	*guyannensis poliopus*
cayennensis rubellus	*guyannensis magdalenae*	*guyannensis warreni*	*guyannensis rattinus*
cayennensis securus	*guyannensis mincae*	*longicaudatus boimensis*	*guyannensis riparium*
cayennensis semispinosus	*guyannensis ochraceus*	*longicaudatus brevicauda*	*guyannensis urichi*
cayennensis simonsi	*guyannensis oconnelli*	*longicaudatus elapossus*	*guyannensis villicauda*
cayennensis trinitatis	*guyannensis oris*	*longicaudatus hendeei*	*hendeei hendeei*
cayennensis urichi	*guyannensis pachita*	*longicaudatus leucomystax*	*hendeei nigrofulvus*
cayennensis warreni	*guyannensis panamensis*	*longicaudatus longicaudatus*	*longicaudatus boimensis*
vacillator	*guyannensis poliopus*	*longicaudatus nigrofulvus*	*longicaudatus brevicauda*
hendeei	*guyannensis rattinus*	*longicaudatus pachita*	*longicaudatus elapossus*
rattinus	*guyannensis roberti*	*longicaudatus roberti*	*longicaudatus leucomystax*
canicollis	*guyannensis rosa*	*longicaudatus securus*	*longicaudatus longicaudatus*
dimidiatus (=Trinomys)	*guyannensis rubellus*	*longicaudatus simonsi*	*longicaudatus roberti*
	guyannensis securus	*semispinosus amphichoricus*	*longicaudatus securus*
iheringi GROUP	*guyannensis semispinosus*	*semispinosus chiquirinus*	*longicaudatus simonsi*
iheringi (=Trinomys)	*guyannensis simonsi*	*semispinosus gorgonae*	*quadruplicatus*

Table 2: Continuation.

ELLERMAN (1940) [3]	HERSHKOVITZ (1948) [4]	MOOJEN (1948) [5]	CABRERA (1961)
Proechimys (Trinomys)	*guyannensis steerei*	*semispinosus gularis*	*semispinosus amphichoricus*
albispinus albispinus (=Trinomys)	*guyannensis trinitatis*	*semispinosus ignotus*	*semispinosus calidior*
albispinus sertonius (=Trinomys)	*guyannensis urichi*	*semispinosus kermiti*	*semispinosus hilda*
setosus (=Trinomys)	*guyannensis warreni*	*semispinosus liminalis*	*semispinosus kermiti*
		semispinosus panamensis	*semispinosus liminalis*
	quadruplicatus GROUP	*semispinosus semispinosus*	*semispinosus rosa*
	quadruplicatus		*semispinosus semispinosus*
	ignotus	**Proechimys (Trinomys)**	
		albispinus albispinus (=Trinomys)	**Proechimys (Trinomys)**
	Proechimys (Trinomys)	*albispinus sertonius (=Trinomys)*	*albispinus albispinus (=Trinomys)*
	albispinus albispinus (=Trinomys)	*dimidiatus (=Trinomys)*	*albispinus sertonius (=Trinomys)*
	albispinus sertonius (=Trinomys)	*iheringi bonafidei (=Trinomys)*	*dimidiatus (=Trinomys)*
	setosus (=Trinomys)	*iheringi denigratus (=Trinomys)*	*iheringi bonafidei (=Trinomys)*
	cinnamomeus (=Trinomys)	*iheringi gratiosus (=Trinomys)*	*iheringi denigratus (=Trinomys)*
	elegans (=Trinomys)	*iheringi iheringi (=Trinomys)*	*iheringi gratiosus (=Trinomys)*
	fuliginosus (=Trinomys)	*iheringi panema (=Trinomys)*	*iheringi iheringi (=Trinomys)*
	leptosoma (=Trinomys)	*iheringi paratus (=Trinomys)*	*iheringi panema (=Trinomys)*
	myosuros (=Trinomys)	*setosus elegans (=Trinomys)*	*iheringi paratus (=Trinomys)*
		setosus setosus (=Trinomys)	*myosuros (=Trinomys)*
	INCERTA SEDIS		*setosus elegans (=Trinomys)*
	dimidiatus (=Trinomys)	**INCERTA SEDIS**	*setosus setosus (=Trinomys)*
	vacillator	*myosuros (=Trinomys)*	

[3] He did not analized 12 nominal taxa.
[4] He did not mention the nominal taxa *hoplomyoides*.
[5] He only included the *Proechimys* forms from Brazil, others taxonomic status from South and Central America were removed in text during his comparisons.

Table 2: Continuation.

EMMONS & GARDNER (1984) [6]	PATTON (1987) [7]	WOODS & KILPATRICK (2005)	PATTON & LEITE (2015)	FABRE *et al.* (2016) [8]
brevicauda GROUP	*canicollis GROUP*	*brevicauda*	*canicollis GROUP*	*canicollis*
amphichoricus	*canicollis*	*bolivianus*	*canicollis*	
bolivianus		*elassopus*		*decumanus*
brevicauda	*decumanus GROUP*	*gularis*	*decumanus GROUP*	
canicollis	*decumanus*	*securus*	*decumanus*	*echinothrix*
cuvieri		*canicollis*		
decumanus	*simonsi GROUP*	*chrysaeolus*	*echinothrix GROUP*	*gardneri GROUP*
goeldii	*simonsi*	*cuvieri*	*echinothrix*	*gardneri*
gularis	*hendeei*	*decumanus*		*kulinae*
guyannensis	*nigrofulvus*	*echinothrix*	*gardneri GROUP*	*pattoni*
longicaudatus		*gardneri*	*gardneri*	
magdalenae	*guyannensis GROUP*	*goeldii*	*kulinae*	*goeldii GROUP*
oris	*guyannensis*	*hyleae*	*pattoni*	*goeldii*
quadruplicatus	*roberti*	*leioprymna*		*hyleae*
roberti	*arabupu*	*nesiotes*	*goeldii GROUP*	*leioprimna*
simonsi	*arescens*	*guairae*	*goeldii*	*nesiotes*
steerei	*boimensis*	*ochraceus*	*hyleae*	*quadruplicatus*
	cherriei	*guyannensis guyannensis*	*leioprimna*	*amphichoricus*
guairae GROUP	*oris*	*cayennensis*	*nesiotes*	*steerei*
guairae	*riparum*	*warreni*	*quadruplicatus*	*hilda*
poliopus	*vacillator*	*guyannensis arabupu*	*amphichoricus*	*kermiti*
hoplomyoides	*warreni*	*guyannensis cherriei*	*steerei*	*liminalis*
mincae		*guyannensis riparum*	*hilda*	*pachita*
trinitatis	*goeldii GROUP*	*guyannensis vacillator*	*kermiti*	*rattinus*
urichi	*goeldii*	*hoplomyoides*	*liminalis*	
	steerei	*kulinae*	*pachita*	*guyannensis GROUP*
semispinosus GROUP	*amphichoricus*	*longicaudatus*	*rattinus*	*guyannensis*
chrysaeolus	*hilda*	*leucomystax*		*arabupu*
oconnelli	*hyleae*	*ribeiroi*	*guyannensis GROUP*	*cherriei*
semispinosus	*kermiti*	*villacauda*	*guyannensis*	*riparum*
	leioprimna	*magdalenae*	*arabupu*	*vacillator*
Trinomys GROUP	*liminalis*	*mincae*	*cayennensis*	*warreni*
albispinus	*nesiotes*	*oconnelli*	*cherriei*	*roberti*
iheringi	*pachita*	*pattoni*	*riparum*	*arescens*
myosuros	*quadruplicatus*	*poliopus*	*vacillator*	*boimensis*
setosus	*rattinus*	*quadruplicatus*	*warreni*	*oris*
		amphichoricus	*roberti roberti*	

Table 2: Continuation.

PATTON (1987) [7]	WOODS & KILPATRICK (2005)	PATTON & LEITE (2015)	FABRE *et al.* (2016) [8]
longicaudatus GROUP	*roberti*	*roberti oris*	*longicaudatus GROUP*
brevicauda	*arescens*	*arescens*	*brevicauda*
longicaudatus	*boimensis*	*boimensis*	*bolivianus*
boliviensis	*oris*		*elassopus*
elassopus	*semispinosus semispinosus*	*longicaudatus GROUP*	*gularis*
gularis	*semispinosus burrus*	*brevicauda*	*securus*
leucomystax	*semispinosus calidior*	*bolivianus*	*cuvieri*
ribeiroi	*semispinosus centralis*	*elassopus*	*longicaudatus*
securus	*semispinosus colombianus*	*gularis*	*myosurus*
villacauda	*gorgonae*	*securus*	*myosruru*
	semispinosus rosa	*cuvieri*	*leucomystax*
cuvieri GROUP	*semispinosus rubellus*	*longicaudatus*	*ribeiroi*
cuvieri	*simonsi*	*leucomystax*	*villicauda*
	hendeei	*ribeiroi*	
semispinosus GROUP	*nigrofulvus*	*villicauda*	*semispinosus GROUP*
oconnelli	*steerei*		*oconnelli*
semispinosus	*hilda*	*semispinosus GROUP*	*semispinosus semispinosus*
burrus	*kermiti*	*oconnelli*	*calidior*
calidior	*liminalis*	*semispinosus*	*semispinosus burrus*
centralis	*pachita*	*burrus*	*semispinosus centralis*
chiriquinus	*rattinus*	*calidior*	*semispinosus colombianus*
colombianus	*trinitatis*	*centralis*	*semispinosus goldmani*
goldmani	*urichi*	*chiriquinus*	*semispinosus gorgonae*
gorgonae		*colombianus*	*semispinosus ignotus*
ignotus		*goldmani*	*semispinosus panamensis*
panamensis		*gorgonae*	*chiriquinus*
rosa		*ignotus*	*semispinosus rosa*
rubellus		*panamensis*	*semispinosus rubellus*
		rosa	
trinitatis GROUP		*rubellus*	*simonsi*
chrysaeolus			*hendeei*
guairae		*simonsi GROUP*	*nigrofulvus*
hoplomyoides		*simonsi*	
trinitatis		*hendeei*	*trinitatis GROUP*
magdalenae		*nigrofulvus*	*chrysaeolus*
mincae			*magdalenae*
ochraceous		*trinitatis GROUP*	*guairae guairae*
poliopus		*chrysaeolus*	*guairae ochraceus*
urichi		*magdalenae*	*guairae poliopus*

Table 2: Continuation.

EMMONS & GARDNER (1984) [6]	PATTON (1987) [7]	WOODS & KILPATRICK (2005)	PATTON & LEITE (2015)	FABRE *et al.* (2016) [8]
			guairae	*hoplomyoides*
			ochraceus	*mincae*
			poliopus	*trinitatis*
			hoplomyoides	*urichi*
			mincae	
			trinitatis	
			urichi	

[6] They did not included all nominal taxa.
[7] He did not included the subgenus *Trinomys* in this study. Also, he did not defined the synonyms by species, he only delimited the groups and the nominal taxa in each group. Later, he cited the nominal taxa which could be considered a species. For the others nominal taxa he affirmed it is necessary a taxonomic revision.
[8] They did not mention the nominal taxa *cayennensis*.

Figures

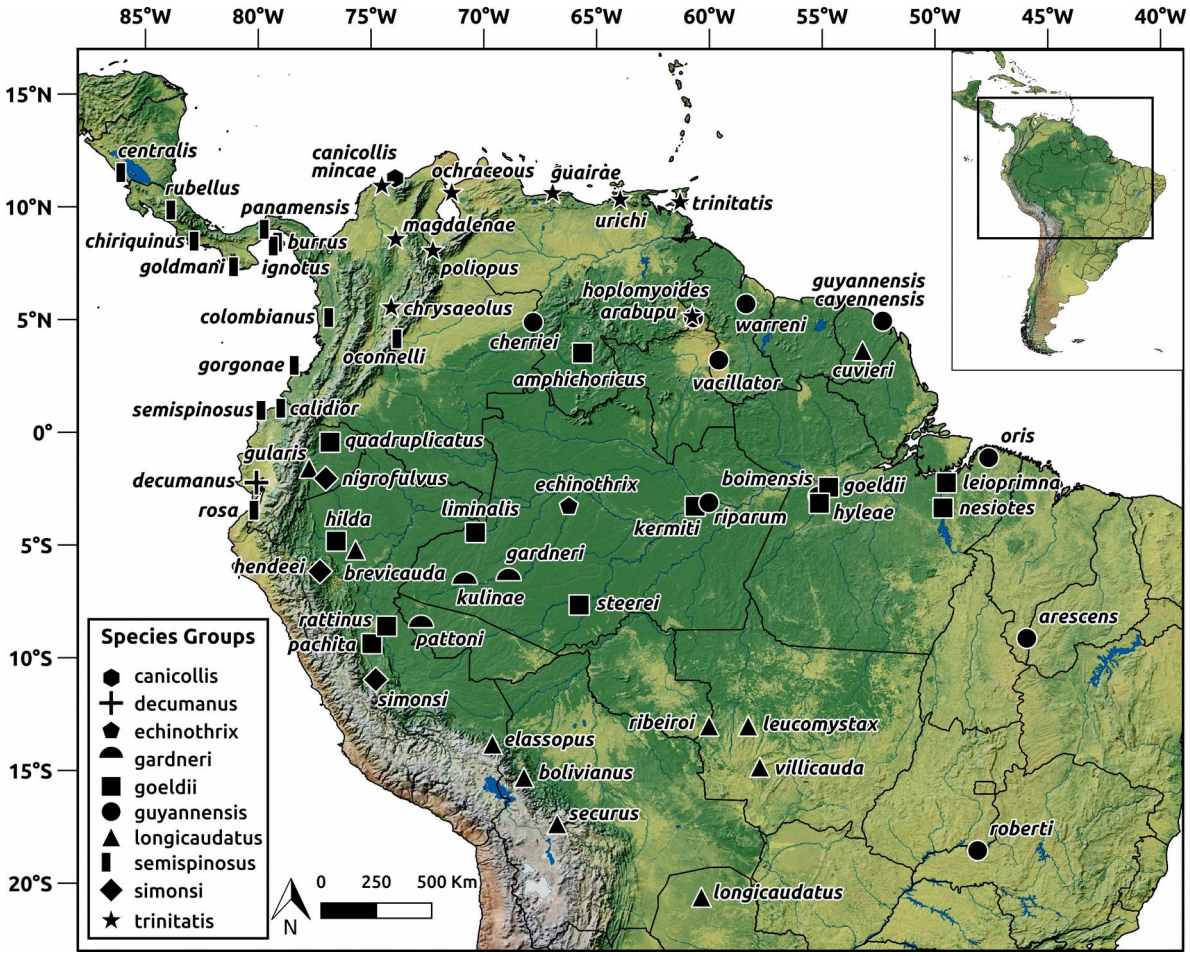


Figure 1: The 64 type localities of nominal taxa for the species groups of genus *Proechimys*. The different symbols represent the 10 species groups defined for Patton and Leite (2015). The nominal taxon *langsdorffii* is not represented in this map because there is no information about its type locality.

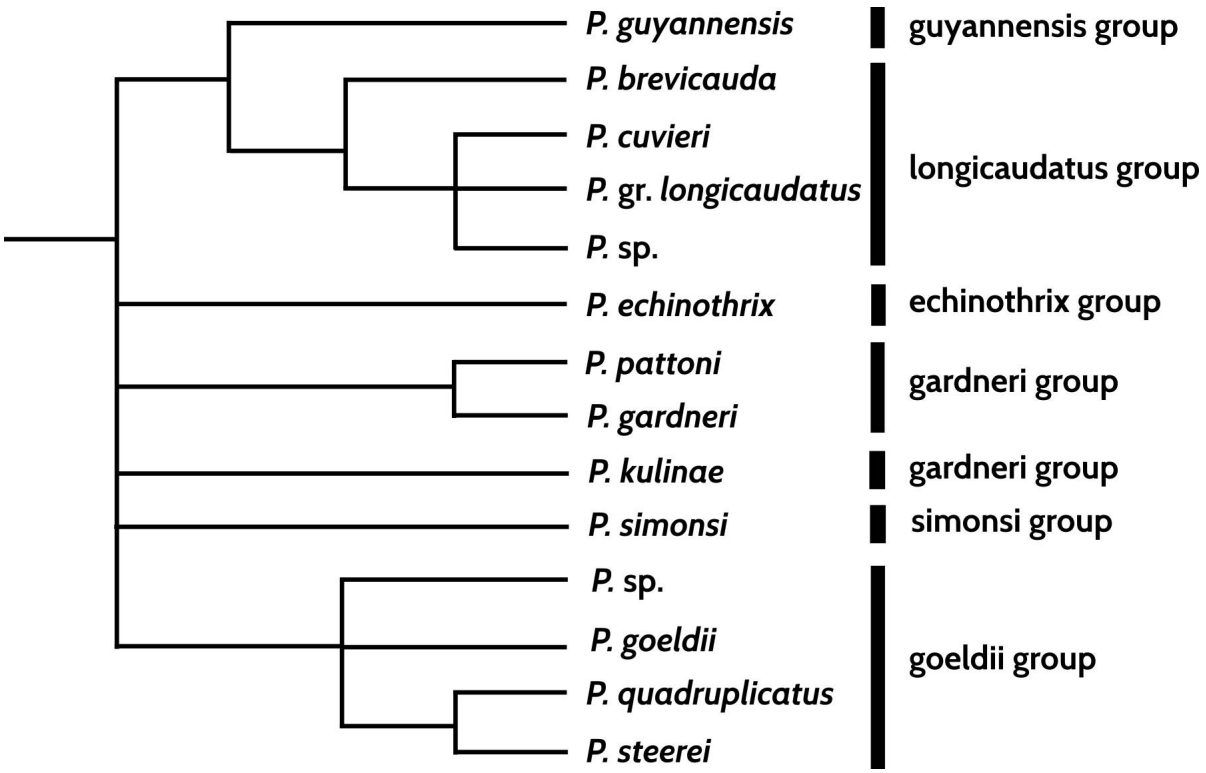


Figure 2: Cladogram representing the phylogeny proposed for the genus *Proechimys* (Schetino 2008). The black bars represent the groups of species according Patton and Leite (2015). Independent lineages were named as *sp.* according to the proposal of the author.

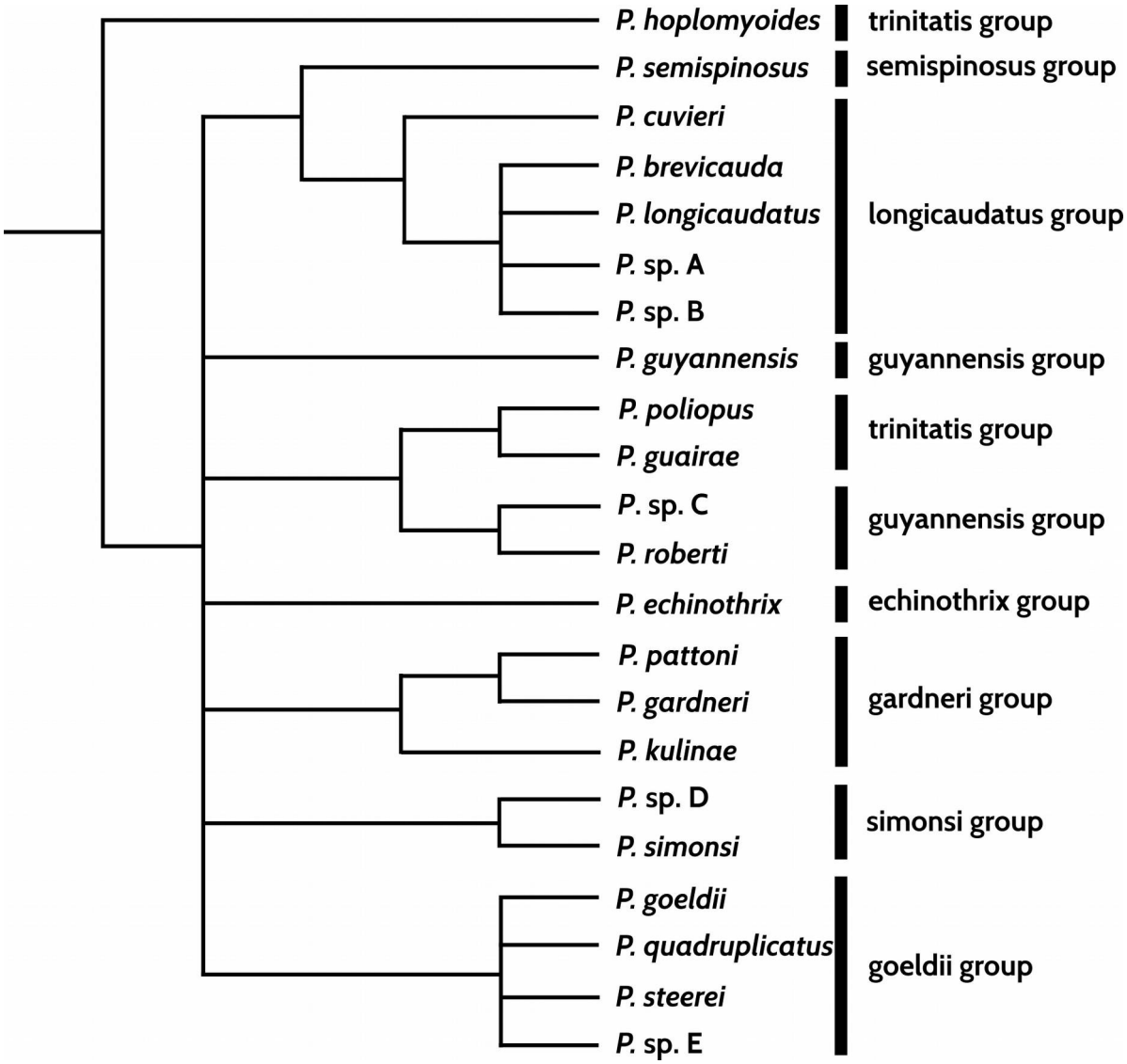


Figure 3: Cladogram representing the phylogeny proposed for the genus *Proechimys* (Leite et al. 2015). The black bars represent the groups of species according Patton and Leite (2015). Independent lineages were named as sp. A, sp. B, sp. C, sp. D, and sp. E according to the proposal of the authors.

2. LET THE RATS OUT OF THE BAG: RAD-SEQUENCING REVEALS THE EVOLUTIONARY HISTORY OF ONE OF THE MOST ABUNDANT AND LITTLE-STUDIED SMALL MAMMALS IN THE AMAZON

Abstract

RAD (Restriction site associated DNA) is one of the most common Next-Generation sequencing techniques widely used in non-model species. It is a versatile sequencing method that can be used for phylogeographic, phylogenetic or population genetics analyses. Here, I used this technique to infer the phylogeny, and delimit species for the spiny rat of the genus *Proechimys*, one of the most common and least studied small mammals in the Amazon. *Proechimys* is the most diverse genus of family Echimyidae, with 22 valid species. Phylogenetic hypotheses presented previously, using mitochondrial data and about 10 of the 22 species, contained a basal polytomy and a phylogeny that includes nuclear DNA has not yet been published. I sequenced 222 specimens, representing most of the geographic distribution of the genus, generating around 90,000 loci. I tested whether inferred lineages in the phylogenetic trees could be considered evolutionarily independent from other ones based on the genomic dataset, in addition to morphometric data of 479 specimens. *Proechimys* was not recovered as a monophyletic genus, as individuals from the Tepuis in the Guyanan Shield, currently known as *Proechimys hoplomyoides*, formed an independent lineage from the other specimens of *Proechimys* and from the representatives of the sister genus *Hoplomys*. There are five main clades, and 28 lineages with statistical support to be considered independent in the analyses of species delimitation. Most cases of sympatry in *Proechimys* occurs among lineages from different clades. I could not associate an available nominal taxon to 12 of the 28 lineages, some of these may represent putative new species, while others may result in revalidations of taxa currently included in the synonymy of valid species.

Keywords: Echimyidae; Morphometry; NGS; *Proechimys;* Species delimitation

2.1. Introduction

Genomic data provided by Next-Generation Sequencing (NGS) technologies have propitiated revolutions in various fields of science (Koboldt et al., 2013; Roukos, 2012), including in the Phylogenetic Systematics (McCormack et al., 2013). Restricted site-associated DNA Sequencing (RAD-Seq or only RAD) is one of the most common techniques used for obtaining such genomic data, especially for non-model taxa, and groups with recent diversification (Peterson et al., 2012). One of the advantages of using the RAD technique is

that the data can be used for both phylogenetic (Eaton, 2014; Rubin et al., 2012) and population genetics analyses (McCormack et al., 2013; Reitzel et al., 2013). Phylogeographic and phylogenetic studies with RAD data have already been published with insects (J. P. Huang and Knowles, 2016), marine invertebrates (Reitzel et al., 2013), mammals (Prado et al., 2019; Puckett et al., 2016), freshwater fishes (Thomaz et al., 2019) and plants (Eaton and Ree, 2013; Resende-Moreira et al., 2019), showing the effectiveness of the data to answering these questions.

One of the criticisms for the use of RAD in phylogenetics is that, in deep divergences, such data does not present good resolution, and others methods as hybrid enrichment approaches would be recommended for groups with older divergence times (Lemmon and Lemmon, 2013). These techniques have been used to infer the evolutionary history of higher taxonomic categories, as tribes (Percequillo et al., *In Prep*) and families (Courcelle et al., 2019). Family Echimyidae, known as spiny rats and commonly distributed in the Neotropical region, is one of Neotropical rodents with more taxonomic studies based on genomic data (Courcelle et al., 2019; Fabre et al., 2017, 2013).

Although the evolutionary history of the Family Echimyidae presented important advances in recent years (Courcelle et al., 2019; Fabre et al., 2017, 2013a; Galewski et al., 2005; Upham et al., 2013; Upham and Patterson, 2015), most of the its genera remain little studied. Among these genera is *Proechimys*, which consists of terrestrial individuals of spiny rats with white belly, elongated ears, long rostra, and narrow and long hindfeet with the smaller species not exceeding 180 mm of body length and 200 g of body mass, and larger ones exceeding 300 mm in length and 500 g in body mass (Da Silva, 1998; Patton and Leite, 2015). *Proechimys* is the most diverse genus within the family, with 22 valid species and more than 60 available nominal taxa for the species group (Patton, 1987; Patton and Leite, 2015). It has a wide distribution in the Neotropics, from Central America to the Brazilian Cerrado, covering the entire Amazon region (Fabre et al., 2016; Woods and Kilpatrick, 2005). In addition, *Proechimys* is very abundant in wildlife surveys, considered the commonest non-volant small mammal in the Amazonian forests (Patton and Leite, 2015). Records of up to five sympatric species are common, especially in the Western Amazon (Malcolm, 1992; Patton et al., 2000; Steiner et al., 2000), and these sympatric species occupy different habitats and micro-habitat in the same locality (Patton et al., 2000; Voss et al., 2001). Some species preferentially occupy *varzea* forests (seasonally floodplain forests), while others inhabit only

terra-firme forests (non-flooded forests) or even both habitats (Matocq et al., 2000; Patton et al., 2000). Also, there is a large difference in body size among sympatric species, which could allow their co-existence, but segregated into different micro-habitats.

Proechimys presented a deep divergence time in studies on the origin and diversification of the family Echimyidae, with date estimates for its origin during the Miocene, in a time window ranging from 12 to 6 Ma, depending on the datasets employed (Álvarez et al. 2017; Fabre et al. 2017; Upham and Patterson 2015). In addition, the origin of the echimyid rodents is considered an event of rapid diversification or a star phylogeny (Lara et al. 1996; Leite and Patton, 2002) that makes difficult the resolution of phylogenetic relationships among and inside genera (Courcelle et al. 2019).

Phylogenetic hypotheses presented previously for the genus, using mitochondrial data and about 10 of the 22 species, showed a basal polytomy, which did not allow the inference of the phylogenetic relationships among the species (Amaral et al., 2013; Da Silva and Patton, 1998; Da Silva, 1998; Patton et al., 2000; Rodrigues da Costa et al., 2016); a phylogeny of the species of the genus including nuclear DNA has not yet been published. This scanty information on the evolutionary history favors the lack of clear boundaries among species in *Proechimys*, especially due to the great geographical variation of morphological characters (Da Silva, 1998; Patton et al., 2000; Patton and Leite, 2015). Thus, it is more common the *Proechimys* individuals to be assigned into the 10 species groups, defined initially by Patton (1987), based on cranial, dental, and bacular characters rather than to the species (Fabre et al., 2016; Patton, 1987; Patton and Leite, 2015).

Here, my aims are (*i*) to recover the phylogenetic relationships of a taxon with deep diversification time, using a RAD technique; and (*ii*) to employ genomic and mophometric data in an integrated way to delimit putative species. I expect the RAD data will be useful to solving the most recent phylogenetic relationships and in the species delimitation analyses but the resolution will decrease in deep divergence. However, even for deep divergences, RAD will be adequate to suggest a diversification hypothesis for one of the poorly studied Neotropical mammal genus.

2.2. Material and Methods

2.2.1. Libraries preparation, sequencing, and reads processing

I built three genomic libraries with 270 samples of *Proechimys* (two with 96 samples, and 86 in the last one), representing the known distribution for the genus (Patton and Leite, 2015; Supplementary Material: Fig. S1; Table S1). In addition, I included 8 individuals as outgroup, corresponding to five genera of the family Echimyidae: *Clyomys laticeps* (n = 1), *Euryzygomatomys spinosus* (n = 1), *Hoplomys gymnurus* (n = 3) the sister-genus of *Proechimys, Myocastor coypus* (n = 1), *Thichomys pachyurus* (n = 1), and *Trinomys dimidiatus* (n = 1). Genomic DNA was extracted with DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA). All libraries were prepared following the Peterson et al. (2012) protocol for ddRAD-Seq, with some modifications (see more details on library preparation in Appendix A in the Supplementary Material), and sequenced at the Hospital for Sick Children (Toronto, Ontario, Canada), in three lanes in the HiSeq2500 (Illumina, San Diego, CA, USA) to generate 150 pb single-end reads.

The raw reads were processed with *ipyrad* pipeline (http://ipyrad.readthedocs.io/). No mismatches in barcodes during the demultiplex step was allowed. I also eliminated reads: i) with more than 5 bp with quality < Q20, ii) with Illumina Phred Q score below 33, iii) without barcodes, and iv) with less than 110 bp of length. The reads were considered homologous if they had similarity equal or greater than 90% and the sequences were aligned by *de novo* method (*i.e.*, without a reference genome). The outputs allowing only alleles with up to 5 N (uncalled bases), up to 6 indels per locus, and up to 50% of heterozygous sites per locus (more details about the reads processing in Appendix B in the Supplementary Material).

2.2.2. Phylogenetic inference

I created a conservative dataset with all individuals with over 100,000 sequenced reads and with loci that were present in at least 15% of individuals (85% of missing data). Phylogenetic trees with individuals as operational taxonomic units (OTU) were constructed

using two methods: (*i*) Maximum Likelihood (ML) in RaxML 8.2.10 (Stamatakis, 2014) with the general time-reversible (GTRGAMMA) substitution model and concatenated data, and by (*ii*) quartet-based distance under coalescent model (CM), using one SNP per loci in SVDquartets implemented in PAUP* version 4.0a (Chifman and Kubatko, 2015). Both methods were performed with 100 bootstrap replications.

I identified and named clades in the conservative trees. These names did not represent taxa, rather the geographical region of the samples, especially the interfluves of Amazonian rivers. After, I built a strict dataset with 73 individuals using two individuals with the largest number of reads per named clades, and with a minimum of 500,000 reads per individual. In addition, I created other RAD matrices from the strict dataset with different amount of missing data (30%, 40%, 50%, 60%, 70%, 80% and 90%). I inferred phylogenetic trees under ML and CM using all strict datasets, as was done for the conservative dataset, to test how the tree topology and statistical support change due the missing data values. For this, I measured the mean of bootstrap values, and its standard deviation considering all nodes in a tree, the number of clades with bootstrap value below 50%, and the topology distance among the trees under Penny and Hendy (1985) model with *dist.topo* function from the "ape" package version 5.2 (Paradis et al. 2004) in R 3.4.4 (R-Development CoreTeam, 2018).

2.2.3. Species delimitation

I employed molecular and morphometric data to perform the species delimitation analyses using iBPP (Solís-Lemus et al., 2015), with three datasets combinations (*i*) only morphometric data, (*ii*) only molecular data, and (*iii*) integrating both datasets. I informed which individuals belong to each putative species along with a guide tree, representing the phylogenetic relationships that I would like to test. The program created alternative hypotheses about the putative species, and tested how many putative species were supported by the data. iBPP collapsed one or more nodes from the guide tree and calculated the posterior probability (PP) for each alternative hypothesis. From the molecular data, gene trees for each locus were estimated independently, using the multispecies coalescent model (MCM) (Rannala and Yang, 2003). For each morphometric variable, a parameter λ modeled the ratio of variance between and within the putative species. Variables were considered independent

and governed by Brownian motion (BM) along the guide tree with individuals having a normal distribution around the mean of the putative species. If putative species were isolated lineages it was expected, by natural selection, low variance among individuals of the same species. Variance values calculated from real data were compared to the expected values in a scenario without selection with Brownian motion from simulations (Solís-Lemus et al., 2015).

I performed three analyses with iBPP: (1) I tested the isolation of Venezuelan Tepui individual (TEP) in relation to the other *Proechimys* (namely afterhere as PRO) and *Hoplomys gymnurus* (HG) specimens; (2) I tested the isolation of main clades recovered by the strict datasets, selecting the four topologies that had bootstrap support equal to or greater than 50% for all clades as guides trees; (3) I tested whether the named clades recovered and identified in the conservative dataset, for each one in the main *Proechimys* clades, should be considered isolated lineages from the others. For Analyses 2 and 3, I randomly selected 500 loci present in all putative species and clades evaluated. For Analysis 1, I was able to recover 250 common loci to all three clades because there was only one individual representing the clade TEP.

Morphometric data was represented by 479 specimens, adult of age classes 8, 9 and 10 (following Patton and Rogers, 1983), representing the same localities or the same individuals for all clades included in the conservative dataset (Table S2). I possessed more morphometric than genomic data, and due to sympatry cases, only geography could not be used to organize morphometric data into genomic clusters. Thus, I performed an extensive morphological analysis and separated specimens with genomic and morphometric data and also those with only morphometric data into morphogroups (Table S2). I relied on the cranial, external and dental characters, following Patton and Leite (2015). Initially, 29 cranial measurements were proposed to be taken with a digital caliper with an accuracy of 0.01 millimeters (Fig. S2), and I eliminated variables with low repeatability; for this test, 40 *Proechimys* skulls were measured. I measured the 29 cranial variables five times, each one in a different day, and I calculated the intraclass correlation coefficient (ICC) and the confidence interval using *Nest* and *ICCest* functions from "ICC" package (Wolak et al., 2012) in R 3.4.4 (R-Development CoreTeam, 2018). After I discarded the variables that presented the confidence interval below 0.8 for the ICC, according to the criteria established by Wolak et al. (2012).

I performed exploratory analyses for eliminating outliers and identifying correlative variables following Zuur et al. (2010). I used only log-transformed variables without missing data, and with normal distribution accessed through Shapiro test (*shapiro.test*) and QQ-plots (*qqnorm* and *qqline*). In addition, I removed the size effect from variables; for this, I performed a Principal Components Analysis (PCA) with *dudi.pca* function in "ade4" package (Dray and Dufour, 2007), using the PC1 to represent the size. I accomplished a linear regression (*lm* function) with PC1 and each variable and used their residuals to create a free size effect dataset. As previous studies did not show the existence of sexual dimorphism in *Proechimys* (Patton and Rogers, 1983) and in other genera of the family Echimyidae (Dalapicolla and Leite, 2015), I did not remove the sex effect from the variables. Finally, I performed a correlation test among the variables and I only used in the iBPP those with correlations below 0.5. All scripts for R analyses are available in my github page (https://github.com/jdalapicolla/Dalapicolla2019).

The MCMC chains for all analyses ran with 500,000 generations, sampling parameters every 5 generations and with a burnin of 20,000, and I verified that all runs reached ESS > 200. I carried out four demographic scenarios for all analyses, used two different θ (ancestral population sizes) and τ (divergence times) values, and specifying a gamma distribution (G) for the priors: (i) θ = G (1, 10) and τ = G (1, 10): small ancestral population sizes with relatively shallow divergence times; (ii) θ = G (1, 10) and τ = G (2, 2000): small ancestral population sizes with relatively deep divergence times; (iii) θ = G (2, 2000) and τ = G (1, 10): large ancestral population sizes with shallow divergence times; (iv) θ = G (2, 2000) and τ = G (2, 2000): large ancestral population sizes with relatively deep divergence times. I set the finetune of 1, with algorithm 0 for rjMCMC searches, and the other parameters were kept in default. In total, I tested 108 models: 12 for Analysis 1 and 48 for each Analyses 2 and 3.

If two lineages are not independent I would expect low PP in their split at the guide tree. I considered two lineages as independent when their split presented PP = 1 at least in three of the four demographic scenarios, and at least in two of three datasets used (only morphometric, only molecular, or integrated data). This conservative approach, considering only PP = 1 as independence evidence, is due to some studies that showed that MCM, used in iBPP, may indicate populational genetic structure and not species differentiation using only molecular data (Sukumaran and Knowles, 2017).

2.2.4. Cytogenetic, mtDNA and nuDNA data

I gathered published data or available data in theses and dissertations, on karyotypes, mitochondrial DNA (mtDNA) and nuclear DNA (nuDNA) of individuals of the genus *Proechimys*. Some of these studies used the same specimens, same localities, or even nearby localities that I used for genomic and morphometric analyses. I performed the research on the Web of Science (https://.webofknowledge.com) and Google Scholar (https://scholar.google.com/) portals, searching for the name "*Proechimys"* in association to the terms "karyotyp*", "cytogenet*", "DNA", "mitochondria*", "nuclea*". I searched for information about the *Proechimys* specimens used in these studies as the catalog number, collector number, locality, analyzed data type, which nominal taxon was assigned to the voucher and after I compared this information with my genetic and morphological datasets.

2.2.5. Provisional name attribution

I was able to associate information of the lineages recovered on this study to cytogenetics, mtDNA, and nuDNA information on the *Proechimys* specimens available in the literature. Based on such knowledge, even though preliminarily, I associated the lineages from phylogenetic trees and species delimitation analyses with the names that had been currently in use for these individuals, localities and species in the literature – see Leite and Patton (2015) as the most recent catalogue of the species of the genus. At this moment, I was conservative, avoiding to include unnecessary noise on the taxonomy of an already complex group; hypothesis of new species established with these analyses, were named with the specific epithet of the closest known species preceded by the word *affinis* (aff.); if more than one putative new species is hypothesized to be related to this species, the name of this species is followed by numbers.

2.3. Results

2.3.1. Phylogenetic trees

After sequencing and cleaning the reads, 222 individuals were retained with more than 100,000 reads, with a mean coverage depth of 21.4 (Table S3). Conservative dataset consisted of 88,129 loci, and in both ML and MC approaches the current definition of the genus *Proechimys* was not recovered as monophyletic (Fig. 1 and 2). In the conservative tree with ML, one individual from the Tepui region in the Guyana Shield (TEP) was recovered as sister to a clade formed by the other specimens of *Proechimys* (namely afterhere as PRO) and *Hoplomys gymnurus* (HG) (Fig. 2). Conservative tree with CM recovered a great and basal polytomy, with the echimyid genera representing outgroups at different points in the tree (Fig. S3).

In the conservative tree with ML, the *Proechimys* (PRO) is structured in five main clades with bootstrap support greater than 95%, named here as clades A, B, C, D, and E. Sister group of PRO are the representatives of *Hoplomys gymnurus* (HG), and the Tepui individual (TEP) is sister to PRO + HG with a medium statistical support. Phylogenetic relationships among these five main clades presented low statistical support. We identified 41 subclades according to the branch lengths and geography (Fig. 1, Fig. 2, and Table S3). Clade A presented 12 subclades (Fig. 1 and Fig. 3a–b), clade B had 6 subclades, while clade C was formed by 3 subclades (Fig. 1 and Fig. 3e). Clade D was the most structured with 17 subclades (Fig. 2 and Fig. 3c–d) whereas clade E was formed by 3 subclades (Fig. 2 and Fig. 3f). Although the MC conservative tree did not recover the structure in five main clades in *Proechimys*, it recovered with low statistical support 36 of 41 subclades of the ML tree (Fig. S3).

Among the five main clades, the clade E was more restricted geographically in the Western Amazon (Fig. 3f) while the others, especially the clade D, were widely distributed. Clades B and C had an eastern distribution near to the Atlantic Ocean coast (Fig. 3e). There was a geographic distribution overlap between the clades, except among B, C, and E. Among the subclades, inside the clades B, C, D and E there is no geographical overlap, except in one case of clade D between the LOR and NSM lineages (Fig. 3d). On the other hand, clade A

presented a great overlap among the subclade distribution (Fig. 3a–b). In this phylogeny, the individuals from Western Andes were restricted to a unique clade: CAM + ECU + MAR (Fig. 2), and those that occupy the *várzea* forests (seasonally flooded lowland forests) in Western Amazon too: CLJ+MMR+UJR (Fig. 1). However, the individuals from Cerrado biome were scattered in three points of the phylogeny in different clades: PTI in the clade A (Fig. 1 and Fig. 3a), CER in the clade B (Fig. 1 and Fig. 3e), and SCE in the clade D (Fig. 2 and Fig. 3c).

2.3.2. Phylogenetic relationship among main clades

Although phylogenetic relationships were well resolved at the conservative tree tips, the relationships among the deeper clades were not. To resolve these relationships, I created a strict database with fewer individuals with good sequencing quality and different amount of missing data to evaluate its effect on tree topology. Strict datasets had 73 individuals representing all subclades except for five of them that were composed by individuals with less than 500,000 reads (LRX, MAR, STR, TEP, and UMD; Table S3). Number of loci varied in the strict datasets, from 741 in 30% of missing data matrix to 59,669 loci in 90% matrix. All trees topologies in the ML analysis and six of the seven in the CM recovered *Proechimys* as monophyletic by the TEP absence (Fig. 4). All subclades from conservative trees were recovered as monophyletic in all trees, only varying the bootstrap values (Fig. S4 – S17). The five main clades were recovered monophyletic in all strict trees with ML, and only in matrices with 60%, 80% and 90% CM analyses, when the number of loci was greater than 2,000 (Fig. 4). Phylogenetic relationships among the five main clades were different depending on missing data amount and the algorithm used for the trees inference. However, only four topologies with five main clades as monophyletic presented bootstrap values greater than 50% for all clades: (*i*) 30%, 40%, and 50% matrices in ML analyses presented the same topology, namely afterhere as Tree 1 (Fig. 4); (ii) 60% matrix in ML, as Tree 2 (Fig. 4); (*iii*) 70%, 80%, and 90% analyses under ML presented the same topology, identical to the conservative tree with ML, namely afterhere as Tree 3 (Fig. 4); and (*iv*) 90% matrix in CM, as Tree 4 (Fig. 4), and the position of clades C and E were the responsible for this variation.

Both bootstrap mean and standard deviation for the strict trees, as well the number of clades with low support, showed that Tree 2 (60% matrix with ML) and Tree 4 (90% matrix

with CM) were the trees with best resolution (Table 1), and between them, Tree 2 was the best one (Table 1). Under coalescent model, the largest distances among tree topologies were among matrices with the lowest number of missing data and loci (d = 57, between 30% and 40% matrices; Table 2), while the smallest distance is between the trees with greater numbers of loci and missing data amount (d = 6, between 90% and 80%; Table 2). On the other hand, trees inferred with maximum likelihood showed the largest distances between the most divergent matrices in number of missing data and loci (d = 12, between 30% and 90%; Table 2), and the smaller ones between matrices with similar dataset (d = 2, between 90% and 80%; Table 2). Considering the average of the topological distances among the best four topologies, the Tree 2 (60% matrix in ML) had the smaller mean distance (d = 5.6) from the other strict trees, while the Tree 1 (30%, 40%, and 50% matrices in ML) and Tree 3 (70%, 80%, and 90% matrices in ML) had d = 6.3. Tree 4 (90% matrix in CM) had a mean d = 23.2. In this way, I considered Tree 2, the topology of the tree with 60% of missing data under ML, the best topology and used it as a hypothesis of diversification for the genus.

2.3.3. Species delimitation

Seven morphometric variables had low repeatability and were eliminated from the analyses (Table S4). After data cleaning, I eliminated variables with non-normal distribution (Table S5), I removed the size effect from the remaining variables, and after I excluded variables with correlations greater than 0.5 (Table S6). In the end, 10 variables were used in species delimitation analyses: BaL: basilar length of Hensel; BuL: bullar length; CDM1: cranial depth at M1; D: diastema length; GSL: greatest length of skull; MB: greatest breadth across mastoid; RB: rostral breadth; RD: rostral depth; OccW: occipital condyle width, and ZB: zygomatic arch breadth (Fig. S2).

Analysis 1, testing the independence of *Hoplomys* (HG), Tepui individuals (TEP) and other *Proechimys* specimens (PRO), showed that the three lineages can be considered independent with molecular data and integrated data (Fig. 5). However, morphological data could only indicate independence between PRO and HG, and not between TEP and PRO or between TEP and HG (Fig. 5). Demographic scenario with less statistical support in the

models was with θ = G (2, 2000) and τ = G (2, 2000): with large ancestral population sizes with relatively deep divergence times (Fig. 5).

Analysis 2, testing the independence of the five main clades in the four different topologies found in strict datasets, presented all main clades of *Proechimys* isolated in all topologies with integrated data (Fig. 6). Models with only molecular data did not recover the independence of Clades A, B, and C in the demographic scenario with large ancestral population sizes and with relatively deep divergence times [θ = G (2, 2000) and τ = G (2, 2000)] in the most of the topologies (Fig. 6), and the same happened with models using only morphological data (Fig. 6). Tree 1 topology (30% matrix with ML) was the one with the most demographic models with low statistical support, regardless of data type (Fig. 6).

In Analysis 3, testing the independence of named clades found in conservative tree under ML, of the 41 subclades recovered, I had good quality data of 32 to test their independence. LRX, MAR, STR, and UMD were composed of samples with less than 500,000 reads, and SSR, IQT, PAB, and LOR had only one individual as representative, and they did not reach 500 loci for the analysis. We did not perform an analysis with clade E because only one of its subclades (WAM) had 500 loci, and this clade was already indicated as independent from the others in Analysis 2. In clade A, MMR + UJR + CLJ cannot be differentiated in most demographic scenarios nor in most datasets, as well the JMI + ATH, and GJR + MRL subclades (Fig. S18). In clade B, the subclades that did not show independence in most datasets and demographic scenarios were CER + XAI (Fig. S18). In Clade C all subclades were considered independent while in clade D two sets of subclades could not be considered isolated: PTC + SCE and SAR + NAR (Fig. S18).

During the literature review, I was able to associate cytogenetic data to 35 of the 41 subclades (except for LMR, LXR, PTC, PAB, UMD, and IQT; Fig. 7 and Table S7), and I found available data on mtDNA, or nuDNA, or both for 32 subclades (Table S7 and Fig. 7). Analyzing the results from species delimitation analyses, the different karyotypes, and if the same subclades were recovered with mtDNA and nuDNA data, I could identify 28 independent lineages and 4 untested lineages (IQT, LOR, PAB, and UMD) that need more data to confirm their independence (Fig. 7).

2.4. Discussion

2.4.1. Phylogeny of *Proechimys* and its taxonomic implications

The first multiloci phylogeny with genomic data for an echimyid rodent suggest that taxonomic rearrangements will be necessary on the taxa of the genus-group and species-group levels. On the genus level, the genomic analyses revealed that the current composition of *Proechimys* (sensu Patton and Leite, 2015) is not monophyletic. The individual ROM115116 included in this analysis as sample TEP was collected at Ridge Camp, in Mount Roraima, Guyana. I analyzed the morphology of the TEP specimen only by photo, and it has characteristics that I can associate it to the genus *Proechimys*, especially the tetralophodont upper teeth, while the genus *Hoplomys* [a possible candidate to attribute this species] presents pentalophodont upper teeth (Patton and Emmons, 2015). The only nominal taxa for *Proechimys* known to occur at the Tepui region, *Proechimys hoplomyoides* (Tate), was collected less than 20 Km away, also on the Mount Roraima Tepui but in the Rondon Camp in Venezuela (Tate, 1939). Thus, based on some morphological characters and on sampling locality, in the same Tepui, it is possible associate the TEP individual to the name *P. hoplomyoides*.

Proechimys hoplomyoides is a rare species, with six collected specimens (Patton and Leite, 2015), and its taxonomic position has been previously questioned. Moojen, (1948) and Cabrera (1961) had already associated the nominal taxa *hoplomyoides* to *Hoplomys* and not to *Proechimys*, due to heavy and dense spines on the dorsum. Even the validity of *Hoplomys* as a genus, fully separated from *Proechimys,* has already been discussed, based on morphological characters (Gardner and Emmons, 1984), and on allozyme analyses (Patton and Reig, 1990). However, this is the first study with a large genomic datasabe including these genera that brought evidence to question the taxonomic position and validity of *Hoplomys* and also of *P. hoplomyoides* by the inclusion of Tepui individuals.

It is possible to consider these results as a bias, since i) only an individual was analyzed, ii) the statistical support was not high, iii) the number of missing loci was high for the TEP individual. For this reason, I tested the isolation of these three lineages through simulations, and the results with genetic and integrated data showed that *Proechimys* (PRO), *Hoplomys* (HG) and *P. hoplomyoides* (TEP) can be considered isolated lineages, and that

morphologically representatives from Tepui region cannot be isolated from HG and PRO. These morphometric results may be explained by the variables chosen for the analysis, as they respect the premises for the simulations but did not encompass the variance needed to evaluate the speciation mechanisms in these lineages (Edwards and Knowles, 2014), or because different databases can tell different evolutionary histories for the same taxon (Solís-Lemus et al., 2015), and therefore it is necessary to invest in integrated analyzes (J. P. Huang and Knowles, 2016). I considered these three lineages as independent, even though their phylogenetic relationships still need to be better evaluated, because simulations with integrated and molecular data ratified the results I found in the phylogenetic trees. Besides the genetic evidence, without accessing with more details the morphology of these individuals from Tepui and comparing with the types and other individuals of *Hoplomys*, I decided to not change the status of these taxa. However, there are two possible taxonomic rearrangements: *i)* consider *Hoplomys* as a junior synonym of *Proechimys* or *ii)* create a new genus for *P. hoplomyoides*, and maintain the validity of *Hoplomys*.

On the taxa of the species-group level, several taxonomic changes will be mandatory. Clades A, B, C of *Proechimys* was not recovered as independent lineages using only molecular and only morphometric dataset, especially in the demographic scenarios with large ancestral populations size and deep divergence times. Their isolation was recovered only in the integrated analyses. In addition to the facts listed above for discordance between different dataset in the simulations, another factor is relevant. The origin of the echimyid genera is considered a case of rapid diversification (Leite and Patton, 2002), and *Proechimys* is not an exception. This may explain the difficulties of mitochondrial markers in resolving the basal polytomy of *Proechimys* (Da Silva, 1998; Patton et al., 2000), and of simulations with one dataset to recover the independence of these lineages. This emphasizes the importance of datasets and analyses integration to recover the evolutionary history of a lineage (Dayrat, 2005; Padial et al., 2010). Analysis 3 also had the same pattern obtained on Analyses 1 and 2, with models based on morphometric data only presenting the lowest statistical support but the use of morphometric data in an integrated framework improving the PP of models when compared with only molecular data models.

The 28 lineages recovered here surpasses the 22 species currently considered valid in the genus (Patton and Leite, 2015). In Clade A, 8 independent lineages were recovered: *Proechimys steerei*, *Proechimys* aff. *steerei*, *Proechimys goeldii*, *Proechimys quadruplicatus*,

Proechimys gardneri, *Proechimys kulinae*, *Proechimys echinothrix*, and *Proechimys* aff. *echinothrix* (Fig. 7 and Table S7). These lineages represented the species groups *goeldii*, *gardneri*, and *echinothrix* sensu Patton and Leite. Four samples in the genomic libraries were representative of holotypes, and they all were recovered in the clade A. The holotype of *Proechimys gardneri* was recovered in the JMI lineage, *Proechimys pattoni* in ATH, *Proechimys kulinae* in GJR, and *Proechimys echinothrix* in the SSR subclades. JMI and ATH subclades currently considered as distinct species, *P. gardneri* and *P. pattoni* respectively, need to have their taxonomic status reassessed because species delimitation analyses did not demonstrate them as isolated lineages. Additionally, only *P. gardneri* and *P. pattoni* have the same karyotype (Machado, 2017; Patton and Leite, 2015) among the 21 species of *Proechimys* with described karyotypes (except by the rare species *P. hoplomyoides*), indicating that these species could not be completely differentiated species.

Clade B presented 4 independent lineages: *Proechimys roberti*, *Proechimys* aff. *roberti* 1, *Proechimys* aff. *roberti* 2, and *Proechimys guairae* (Fig. 7 and Table S7). These lineages represented partially the species groups *guyannensis,* and *trinitatis* sensu Patton and Leite. Individuals from Cerrado biome and Eastern Amazon associated historically to the nominal taxon *oris* were recovered in the *Proechimys roberti* lineage, while the *P.* aff. *roberti* 1 included samples from Tapajós-Xingu interfluve, and *P.* aff. *roberti* 2 from Madeira-Tapajós interfluve. Furthermore, specimens from northern Venezuela, where the *Proechimys guairae* species complex occurs were recovered in the NVZ subclade, and I named as *Proechimys guairae*, following Pattom and Leite.

Clade C was formed by three lineages from the Guianan region (Fig. 7 and Table S7), namely *Proechimys guyannensis* from Guiana Shield, *Proechimys* aff. *guyannensis* 1 from the Rio Negro basin, and *Proechimys* aff. *guyannensis* 2 from Bolivar, Southeastern Venezuela. They were associated to a part of the species group *guyannensis* (Patton and Leite, 2015).

Clade D is one of the most complex groups, formed by individuals that occur in the entire distribution of the genus, and represent 12 independent lineages: *Proechimys brevicauda*, *Proechimys* aff. *brevicauda* 1, *Proechimys* aff. *brevicauda* 2, *Proechimys longicaudatus*, *Proechimys cuvieri*, *Proechimys* aff. *cuvieri* 1, *Proechimys* aff. *cuvieri* 2, *Proechimys* aff. *cuvieri* 3, *Proechimys* aff. *cuvieri* 4, *Proechimys semispinosus*, *Proechimys*

decumanus, and *Proechimys chrysaeolus* (Fig. 7 and Table S7). For three subclades (LOR, PAB, and UMD) I did not have enough data to test their independence. Samples from Pantanal biome, south of the Cerrado, Western Andes, Central America, and Magdalena River basin were recovered in clade D, and represented the species groups *decumanus*, *longicaudatus*, *semispinosus* and part of the *trinitatis.*

Clade E that is restricted to the *terra-firme* forests from Western Amazon have individuals in one lineage: *Proechimys simonsi*, and one subclade (IQT) that did not have enough data to test its independence (Fig. 7 and Table S7). *Proechimys simonsi* represents the species group *simonsi* (Patton and Leite2015).

These results showed that species groups *guyannesnis* and *trinitatis* sensu Patton and Leite (2015) are not monophyletic, and some morphological characters used to define them, as the bacular shape, should be used with caution, as pointed out in other studies with terrestrial echimyids (Dalapicolla and Leite, 2015). Formal association of the valid and available nominal taxa of species-group of genus *Proechimys* to the 28 independent lineages identified in this study is beyond its scope, as it would require a broad taxonomic revision with analyses of types and geographical variation.

Cases of sympatry among the subclades, especially in the Western Amazon, occur mainly between lineages from different main clades, according to the geographical distribution of the subclades (Fig. 3). Some studies indicated differences in the microhabitat occupied by sympatric species of *Proechimys* (Patton et al., 2000; Voss et al., 2001). These results corroborated the hypothesis of segregation in microhabitat because the greater the phylogenetic distance, the greater would be the chance of accumulation of differences, including in the use of microhabitat. Clade A have most of the cases of internal overlapping among subclades but these sympatric lineages presented differences in body sizes and in habitat use. *Proechimys steerei* (CLJ, MMR, UJR) and *P. quadruplicatus* (IJM) are typical of várzea forests while the other lineages use terra-firme forests (Matocq et al., 2000; Patton et al., 2000). *Proechimys gardneri* (JMI, ATH) and *Proechimys kulinae* (GJR, LMR) are the smaller species of *Proechimys* (Da Silva, 1998), while *P. goeldii* (EAM), *P.* aff. *steerei* (PTI), *P. steerei* (CLJ, MMR, UJR) and *P. quaduplicatus* (IJM) represented specimens with the larger body size in *Proechimys* (Da Silva and Patton, 1998; Patton and Leite, 2015). For the case of sympatry between lineages within clade D, between *P.* aff. *brevicauda* 1 (NSM) and

untested LOR subclade, it was not possible to associate to the microhabitat segregation because basic information about the ecology of these lineage is lacking.

2.4.2. Phylogenetic inference using RAD data

Proechimys phylogeny based on conservative RAD matrix recovered with high statistical support the most recent branches, but the support was smaller at the deeper divergences, such as, among the five main clades or among PRO+HG+TEP lineages. On the other hand, this is the first phylogeny that did not recover a basal polytomy for *Proechimys*. The poor resolution could be explained by the (*i*) amount of missing data, since our conservative matrix had 85% of missing data; (*ii*) the low resolution of RAD technique to solve deep divergences like in *Proechimys*; (*iii*) due to the evolutionary history of *Proechimys*, with a rapid diversification event, which makes it difficult to elucidate phylogenetic relationships, regardless the missing data amount or the sequencing technique applied.

Studies showed that high missing data amount in a RAD matrix may be more beneficial for inference of the evolutionary history of a taxon (Rubin et al., 2012; Wagner et al., 2013; Wessinger et al., 2016). For example, Tripp et al. (2017) recovered good statistical support for phylogenies with 90% of missing data matrices, and H. Huang and Knowles (2016) showed through simulations that include more individuals and loci in RAD matrices may be more informative than to remove loci with large amount of missing, whether in recent or older divergences. The increase of missing data amount leads to a greater loci number, which can encompass a larger mutational spectrum, allowing the sampling of different types of mutations in different parts of the genome (H. Huang and Knowles, 2016), and improving the resolution of phylogenetic inference.

My results demonstrated that in the trees inferred with CM, the levels of missing data influenced the bootstrap value, its standard deviation, the number of clades with low support, and the topology distance: these indices improves with the increase on the levels of missing data. On the other hand, in trees inferred with ML with concatenated data, the increase of the missing data was not proportional to the improvement of these indices, as the best trees were obtained with intermediate levels of missing data. In addition, missing data influenced the

bootstrap values and the topology of the trees in the CM, while for ML the influence was higher in the topology than in the bootstrap value. Thus, missing data amount in a RAD matrix can influence in different ways the phylogenetic reconstruction, depending on the algorithm, and tests using different matrices with varied values of missing data in phylogenomics using RAD data is recommended (H. Huang and Knowles, 2016).

RAD are known to be efficient in resolving phylogenetic relationships involved in rapid diversification and with shallow divergences times (Eaton, 2014; Lemmon and Lemmon, 2013). However, their resolution power decreased when they were used in deeper diversification groups, more phylogenetically distant (Tripp et al., 2017). In older groups, smaller is the chance to identify orthologous sequences during the bioinformatics steps (Rubin et al., 2012), due to accumulated mutations in the restriction sites used by the enzymes in DNA digestion, during libraries preparations. Nevertheless, several studies using RAD-Sequencing data have succeeded in recovering the phylogenetic relationships of taxa with varied divergence times: with cichlid fish with divergence time <15,000 years ago (Wagner et al., 2013); flowering plants with divergence around 4.8 Ma (Tripp et al., 2017); American live oaks around 7 Ma (Cavender-Bares et al., 2015); ground beetles around 17 Ma (Cruaud et al., 2014); and *Drosophila* between 5 – 63 Ma (Cariou et al., 2013). Therefore, RAD is effective enough to act at different levels of divergence, even for *Proechimys* species divergence time.

Since it was demonstrated by empirical data that RAD have the resolution power to recover deeper phylogenetic relationships, and since the large missing data amount is beneficial in these data in association with different algorithms of tree reconstruction, the low statistical support in conservative trees and the differences in topologies in strict trees (considering the five main clades in *Proechimys*), could be related to the evolutionary history *Proechimys*, with rapid diversification in Late Miocene, rather than to a lack of data power. However, more data from other NGS technique or divergence time estimates for origin of *Proechimys* and its main clades will be required to test this hypothesis.

2.5. Conclusions

In this contribution, I presented the first phylogeny to the most diverse genus of the Family Echimyidae, using multiloci from Restriction-site associated DNA sequencing (RAD).

RAD is known to solve recent diversification events, and here I applied this technique to infer the phylogeny of this genus with deep divergence times.

In this study, *Proechimys* does not represent a monophyletic genus, since a lineage from the Tepui region, *Proechimys hoplomyoides*, was considered an independent lineage of the rest of the genus in the species delimitation analyses, and in phylogenetic tree *P. hoplomyoides* was recovered as a sister group of another echimyid genus, *Hoplomys* and other *Proechimys* specimens. A broader phylogenetic assessment for both genera, *Hoplomys* and *Proechimys*, with the inclusion of additional samples of *Hoplomys* and *P. hoplomyoides*, is required.

Excluding *P. hoplomyoides*, I divided the genus *Proechimys* into five main clades, named A, B, C, D, and E, and I identified 28 lineages with independent evolutionary histories that could be considered putative species. Clade A is represented by individuals identified in previous studies as belonging to the species groups: *echinothrix*, *gardneri* and *goeldii*; Clade B: part of the *guyannensis* group and part of the *trinitatis* group; Clade C: part of the *guyannensis* group; Clade D: comprised species groups *decumanus, longicaudatus*, *semispisnosus*, and part of *trinitatis*. Group E was formed by *simonsi* group. Most of sympatry was among the lineages from different clades, and in the clade A, the sympatric lineages can be segregated at the microhabitat level. In general, these main clades were not in agreement with the 10 species groups suggested in the literature by morphological data.

In *Proechimys*, the resolution among the main clades was poor, and I cannot associate it to the missing data or low number of orthologous loci. Rather, the low resolution among *Proechimys* clades probably is due to the rapid and deep diversification event in the Miocene. However, I showed that using other tools with RAD data, such as, simulations, datasets integration, test different amount of missing data, it is possible to improve the resolution power, even for deeper and rapid diversification events.

References

ÁLVAREZ, A., R. L. M. ARÉVALO, AND D. H. VERZI. 2017. Diversification patterns and size evolution in caviomorph rodents. Biological Journal of the Linnean Society 121:907–922.

AMARAL, P. J. ET AL. 2013. *Proechimys* (Rodentia, Echimyidae): characterization and taxonomic considerations of a form with a very low diploid number and a multiple sex chromosome system. BMC Genetics 14:21.

CABRERA, A. 1961. Catalogo de los Mamiferos de America del Sur. Revista del Museo Argentino de Ciencias Naturales Bernardino Rivadavia IV:518–531.

CARIOU, M., L. DURET, AND S. CHARLAT. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. Ecology and Evolution 3:846–852.

CAVENDER-BARES, J., A. GONZÁLEZ-RODRÍGUEZ, D. A. R. EATON, A. A. L. HIPP, A. BEULKE, AND P. S. MANOS. 2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach. Molecular Ecology 24:3668–3687.

CHIFMAN, J., AND L. KUBATKO. 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. Journal of Theoretical Biology 374:35–47.

COURCELLE, M., M. TILAK, Y. L. R. LEITE, E. J. P. DOUZERY, AND P.-H. FABRE. 2019. Digging for the spiny rat and hutia phylogeny using a gene capture approach, with the description of a new mammal subfamily. Molecular Phylogenetics and Evolution 136:241–253.

CRUAUD, A. ET AL. 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. Molecular Biology and Evolution 31:1272–1274.

DALAPICOLLA, J., AND Y. L. R. LEITE. 2015. Taxonomic implications of morphological variation in three species of *Trinomys* (Rodentia: Echimyidae) from eastern Brazil. Zootaxa 3919:61–80.

DAYRAT, B. 2005. Towards integrative taxonomy. Biological Journal of the Linnean Society 85:407–415.

DRAY, S., AND A. B. DUFOUR. 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22:1–20.

EATON, D. A. R. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics (Oxford, England) 30:1844–9.

EATON, D. A. R., AND R. H. REE. 2013. Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Systematic biology 62:689–706.

EDWARDS, D. L., AND L. L. KNOWLES. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? Proceedings of the Royal Society B: Biological Sciences 281:20132765–20132765.

FABRE, P. H. ET AL. 2017. Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. Molecular Biology and Evolution 34:613–633.

FABRE, P. H., T. GALEWSKI, M. K. TILAK, AND E. J. P. DOUZERY. 2013. Diversification of South American spiny rats (Echimyidae): A multigene phylogenetic approach. Zoologica Scripta 42:117–134.

FABRE, P. H., J. L. PATTON, AND Y. L. R. LEITE. 2016. Family Echimyidae (hutias, South American spiny-rats and coypu). Pp. 552–641 in Handbook of the Mammals of the World. Vol 6. Lagomorphs and Rodents I (D. E. Wilson, T. E. J. Lacher & R. A. Mittermeier, eds.). Lynx Edicions, Barcelona.

GALEWSKI, T., J. F. MAUFFREY, Y. L. LEITE, J. L. PATTON, AND E. J. DOUZERY. 2005. Ecomorphological diversification among South American spiny rats (Rodentia; Echimyidae): a phylogenetic and chronological approach. Molecular Phylogenetics and Evolution 34:601–615.

GARDNER, A. L., AND L. H. EMMONS. 1984. Species groups in *Proechimys* (Rodentia, Echimyidae) as indicated by karyology and bullar morphology. Journal of Mammalogy 65:10–25.

HUANG, H., AND L. L. KNOWLES. 2016a. Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Systematic Biology 65:357–365.

HUANG, J. P., AND L. L. KNOWLES. 2016b. The species versus subspecies conundrum: quantitative delimitation from integrating multiple data types within a single bayesian approach in Hercules Beetles. Systematic biology 65:685–99.

KOBOLDT, D. C., K. M. STEINBERG, D. E. LARSON, R. K. WILSON, AND E. R. MARDIS. 2013. The next-generation sequencing revolution and its impact on genomics. Cell 155:27–38.

LARA, M. C., J. L. PATTON, AND M. N. F. DA SILVA. 1996. The simultaneous diversification of South American echimyid rodents (Hystricognathi) based on complete citochrome b sequences. Molecular Phylogenetics and Evolution 5:403–413.

LEITE, Y. L. R., AND J. L. PATTON. 2002. Evolution of South American spiny rats (Rodentia, Echimyidae): the star-phylogeny hypothesis revisited. Molecular Phylogenetics and Evolution 25:455–464.

LEMMON, E. M., AND A. R. LEMMON. 2013. High-throughput genomic data in systematics and phylogenetics. Annual Review of Ecology, Evolution, and Systematics 44:99–121.

MACHADO, M. X. 2017. Cariologia e filogenia molecular em *Trinomys* e *Proechimys* (Echimyidae, Rodentia). Universidade Federal do Espírito Santo.

MATOCQ, M. D., J. L. PATTON, AND M. N. F. DA SILVA. 2000. Population genetic structure of two ecologically distinct amazonian spiny rats: separating history and current ecology. Evolution 54:1423–1432.

MCCORMACK, J. E., S. M. HIRD, A. J. ZELLMER, B. C. CARSTENS, AND R. T. BRUMFIELD. 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution 66:526–538.

MOOJEN, J. 1948. Speciation in the Brazilian spiny rats (genus *Proechimys*, Family Echimyidae). University of Kansas Publications, Museum Natural History 1:303–401.

PADIAL, J. M., A. MIRALLES, I. DE LA RIVA, AND M. VENCES. 2010. The integrative future of taxonomy. Frontiers in Zoology 7:16.

PARADIS, E., J. CLAUDE, AND K. STRIMMER. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.

PATTON, J. L. 1987. Species groups of spiny rats, genus *Proechimys* (Rodentia : Echimyidae). Fieldiana Zoology 39:305–346.

PATTON, J. L., AND L. H. EMMONS. 2015. Genus *Hoplomys* J. A. Allen, 1908. Pp. 940–942 in Mammals of South America, Volume 2: Rodents (J. L. Patton, G. D'Elía & U. F. J. Pardiñas, eds.). The University of Chicago Press, Chicago and London.

PATTON, J. L., AND O. A. REIG. 1990. Genetic differentiation among echimyid rodents, with emphasis on spiny rats, genus *Proechimys*. Advances in Neotropical Mammalogy:75–96.

PATTON, J. L., AND M. A. ROGERS. 1983. Systematic implications of non-geographic variation in the spiny rat genus *Proechimys* (Echimyidae). Zeitschrift für Säugetierkunde 48:363–370.

PATTON, J. L., M. N. DA SILVA, AND J. R. MALCOLM. 2000. Mammals of the rio Juruá and the evolutionary and ecological diversification of Amazonia. Bulletin American Museum of Natural History 244:1–306.

PATTON, J. P., AND R. N. LEITE. 2015. Genus *Proechimys* J. A. Allen, 1889. Pp. 950–989 in Mammals of South America, Volume 2: Rodents (J. L. Patton, U. F. J. Pardiñas & G. D'Elía, eds.). University Of Chicago Press, Chicago.

PENNY, D., AND M. D. HENDY. 1985. The use of tree comparison metrics. Systematic Zoology 34:75–82.

PERCEQUILLO, A. R. ET AL. no date. Tempo and mode of evolution of Oryzomyine rodents (Rodentia, Cricetidae, Sigmodontinae): a phylogenomic approach. *In Prep*.

PETERSON, B. K., J. N. WEBER, E. H. KAY, H. S. FISHER, AND H. E. HOEKSTRA. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

PRADO, J. R., A. R. PERCEQUILLO, A. T. THOMAZ, AND L. L. KNOWLES. 2019. Similar but different: Revealing the relative roles of species-traits versus biome properties structuring genetic variation in South American marsh rats. Journal of Biogeography 46:770–783.

PUCKETT, E. E. ET AL. 2016. Global population divergence and admixture of the brown rat (*Rattus norvegicus*). Proceedings of the Royal Society of London B: Biological Sciences 283:20161762.

R-DEVELOPMENT CORETEAM. 2018. R Development Core Team. Viena. *R: A Language and Environment for Statistical Computing*.

RANNALA, B., AND Z. YANG. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–56.

REITZEL, A. M., S. HERRERA, M. J. LAYDEN, M. Q. MARTINDALE, AND T. M. SHANK. 2013. Going where traditional markers have not gone before: utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. Molecular Ecology 22:2953–2970.

RESENDE-MOREIRA, L. C. ET AL. 2019. Evolving in isolation: Genetic tests reject recent connections of Amazonian savannas with the central Cerrado. Journal of Biogeography 46:196–211.

RODRIGUES DA COSTA, M. J. ET AL. 2016. Cryptic species in *Proechimys goeldii* (Rodentia, Echimyidae)? A case of molecular and chromosomal differentiation in allopatric Populations. Cytogenetic and genome research 148:199–210.

ROUKOS, D. H. 2012. Biotechnological, genomics and systems–synthetic biology revolution: redesigning genetic code for a pragmatic systems medicine. Expert Review of Medical Devices 9:97–101.

RUBIN, B. E. R., R. H. REE, AND C. S. MOREAU. 2012. Inferring phylogenies from RAD sequence data. PLoS ONE 7:e33394.

DA SILVA, M. N., AND J. L. PATTON. 1998. Molecular phylogeography and the evolution and conservation of amazonian mammals. Molecular Ecology 7:475–486.

DA SILVA, M. N. F. 1998. Four new species of spiny rats of the genus *Proechimys* (Rodentia: Echimyidae) from the western Amazon of Brazil. Proceedings of the Biological Society of Washington 111:436–471.

SOLÍS-LEMUS, C., L. L. KNOWLES, AND C. ANÉ. 2015. Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution 69:492–507.

STAMATAKIS, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

SUKUMARAN, J., AND L. L. KNOWLES. 2017. Multispecies coalescent delimits structure, not species. Proceedings of the National Academy of Sciences of the United States of America 114:1607–1612.

TATE, G. H. H. 1939. The mammals of the Guiana Region. Bulletin of the American Museum of Natural History 76:151–229.

THOMAZ, A. T., T. P. CARVALHO, L. R. MALABARBA, AND L. L. KNOWLES. 2019. Geographic distributions, phenotypes, and phylogenetic relationships of *Phalloceros* (Cyprinodontiformes: Poeciliidae): Insights about diversification among sympatric species pools. Molecular Phylogenetics and Evolution 132:265–274.

TRIPP, E. A., Y.-H. E. TSAI, Y. ZHUANG, AND K. G. DEXTER. 2017. RADseq dataset with 90% missing data fully resolves recent radiation of *Petalidium* (Acanthaceae) in the ultra-arid deserts of Namibia. Ecology and Evolution 7:7920–7936.

UPHAM, N. S., R. OJALA-BARBOUR, J. BRITO M, P. M. VELAZCO, AND B. D. PATTERSON. 2013. Transitions between Andean and Amazonian centers of endemism in the radiation of some arboreal rodents. BMC Evolutionary Biology 13:191.

UPHAM, N. S., AND B. D. PATTERSON. 2015. Evolution of caviomorph rodents: a complete phylogeny and timetree for living genera. Pp. 63–120 in Biology of caviomorph rodents: diversity and evolution (A. I. Vassallo & D. Antenucci, eds.). SAREM Series A, Buenos Aires.

VOSS, R. S., D. P. LUNDE, AND N. B. SIMMONS. 2001. The mammals of Paracou, French Guiana: a neotropical lowland rainforest fauna part 2. Nonvolant species. Bulletin of the American Museum of Natural History 263:3–236.

WAGNER, C. E. ET AL. 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular Ecology 22:787–798.

WESSINGER, C. A., C. C. FREEMAN, M. E. MORT, M. D. RAUSHER, AND L. C. HILEMAN. 2016. Multiplexed shotgun genotyping resolves species relationships within the North American genus Penstemon; Multiplexed shotgun genotyping resolves species relationships within the North American genus Penstemon. American Journal of Botany 103:912–922.

WOLAK, M. E., D. J. FAIRBAIRN, AND Y. R. PAULSEN. 2012. Guidelines for estimating repeatability. Methods in Ecology and Evolution 3:129–137.

WOODS, AND KILPATRICK. 2005. Infraorder Hystricognathi Brandt, 1855. Pp. 1538–1600 in Mammal Species of the World. A Taxonomic and Geographic Reference (D. E. Wilson & D. M. Reeder, eds.). Volume 2. Johns Hopkins University Press, Baltimore.

ZUUR, A. F., E. N. IENO, AND C. S. ELPHICK. 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1:3–14.

Tables

Table 1: Bootstrap mean (Mean), standard deviation (SD) values, and clades number with bootstrap values below 50% (Clades with low Support) for each strict datasets with different amounts of missing data under Coalescent Model (CM) and Maximum Likelihood (ML). Loci = number of SNPs (CM) or loci (ML) used for the trees inference. Bold numbers indicate the best values for each variable. For trees details see Fig. S6-S19.

Missing Data	Loci	Mean		SD		Clades with Low Support	
		CM	ML	CM	ML	CM	ML
30%	741	70.7	97.99	37.36	9.19	52	3
40%	1,046	78.3	98.06	32.53	9.2	40	3
50%	1,580	85.22	97.36	28.02	9.47	27	3
60%	2,604	89.79	**99.93**	21.67	**0.43**	22	**1**
70%	6,754	95.32	98.3	12.42	8.14	9	3
80%	15,911	94.32	99.2	17.36	4.78	12	2
90%	59,669	**98.13**	98.83	**7.07**	5.93	**3**	2

Table 2: Topology distance (d) calculated under PH model among trees with different amount of missing data inferred under the Coalescent Model (lower diagonal) and Maximum Likelihood (upper diagonal). Values in bold indicate greater distances and values in italics the smaller distances for each phylogenetic inference method.

Missing Data	30%	40%	50%	60%	70%	80%	90%
30%	-	4	6	6	10	10	**12**
40%	**57**	-	*2*	4	8	6	8
50%	**57**	28	-	6	6	4	6
60%	55	36	32	-	4	6	8
70%	51	34	26	20	-	*2*	4
80%	53	32	28	16	10	-	*2*
90%	51	34	26	16	*6*	*6*	-

Figures

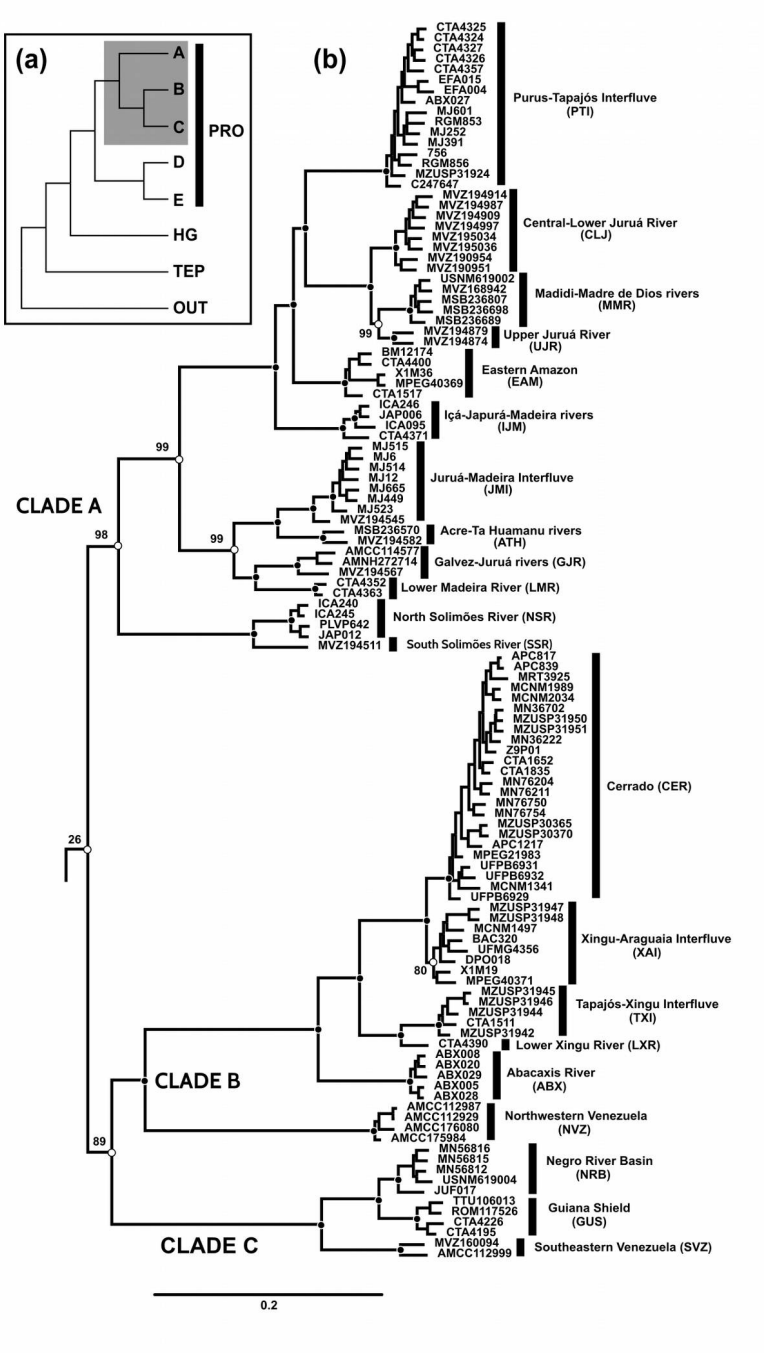


Figure 1: (a) Conservative phylogenetic tree built under the Maximum Likelihood (ML). (b) In details for clades A, B and C of the genus *Proechimys*, for clades D and E see Fig. 2. Black bars correspond to the subclades within each main clade with their respective names and acronyms. Black circles indicate 100% of bootstrap, white circles are bootstrap values below 100% and the value is shown. Bootstrap values within each subclade have been omitted for a better view. Further information on samples are in Table S1.

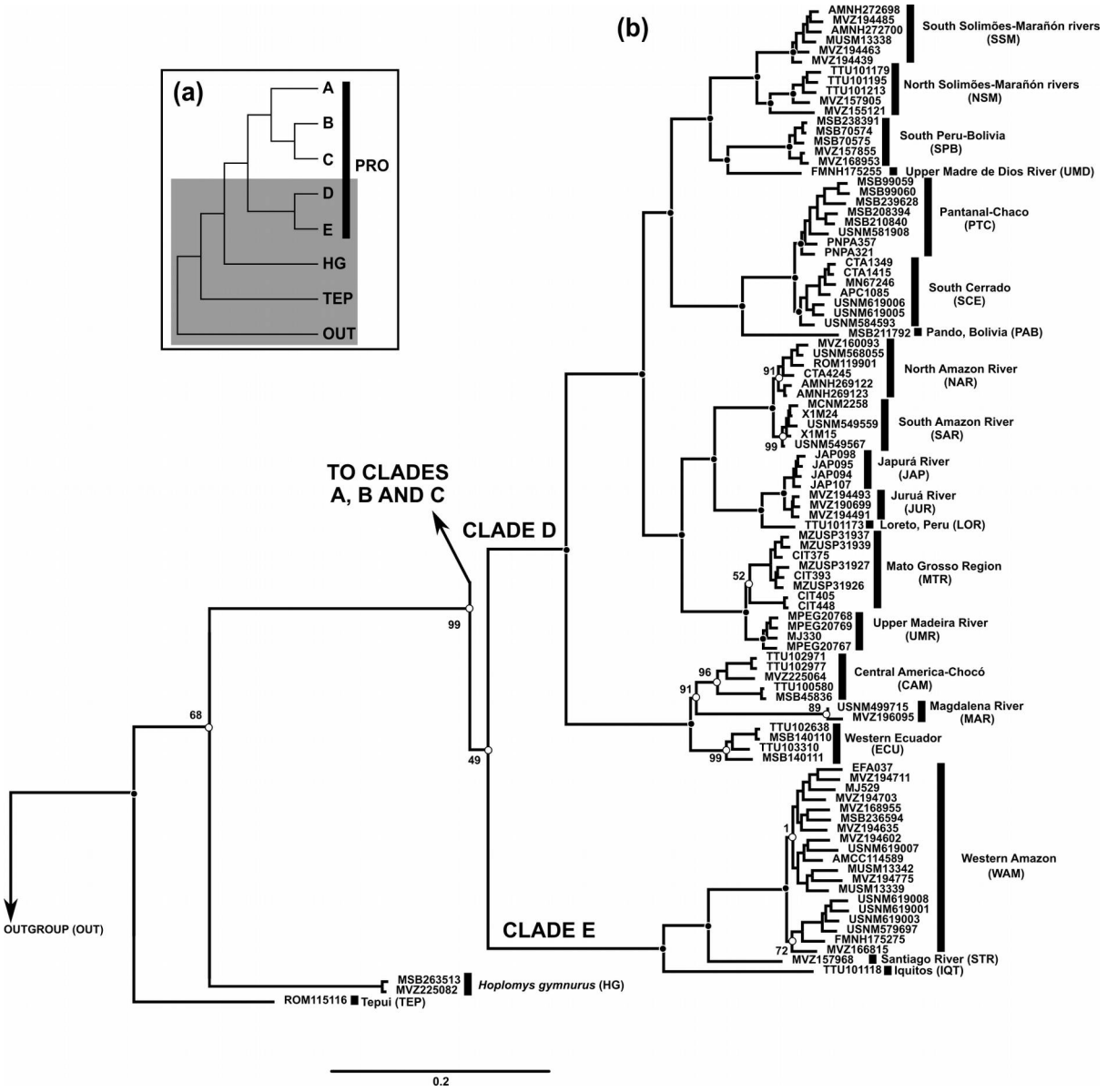


Figure 2: (a) Conservative phylogenetic tree built under the Maximum Likelihood (ML). (b) In details for clades D, E and outgroups, for clades A, B and C see Fig. 1. Black bars correspond to the subclades within each main clade with their respective names and acronyms. Black circles indicate 100% of bootstrap, white circles are bootstrap values below 100% and the value is shown. Bootstrap values within each subclade have been omitted for a better view. Further information on samples are in Table S1.

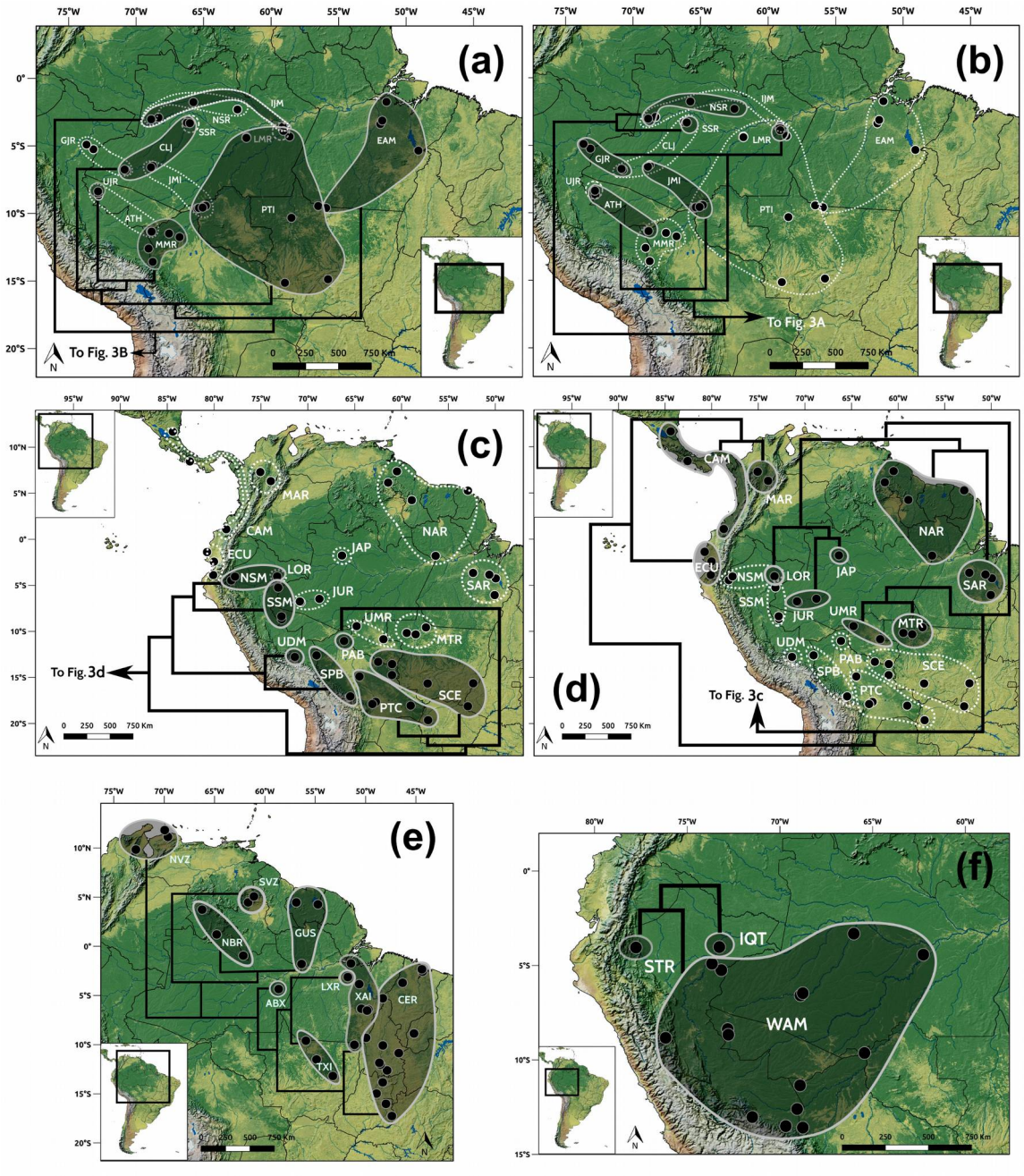


Figure 3: Geographic distribution of the subclades present in the main clades of the genus *Proechimys*. Due to the large subclades number in clades A and D, they were split into two figures each to improve visualization. (a) and (b) represent Clade A; (c) and (d) represent Clades D; (e) represents Clade B and C, and (f) Clade E. Black dots represent genetic samples, letters are the subclades acronyms (for meanings see Figures 1 and 2), the circles with solid line filled in gray represent the distribution area of a subclade for better visualization. Dashed circles represent subclades, from the same main clade, not highlighted in the figure to improve visualization. Black lines represent the phylogenetic relationships between the subclades, according to the Maximum Likelihood conservative tree (Fig 1 and 2). Rivers and mountains are represented on the map, dark green shades indicate closed vegetation, and light green shades is open vegetation.

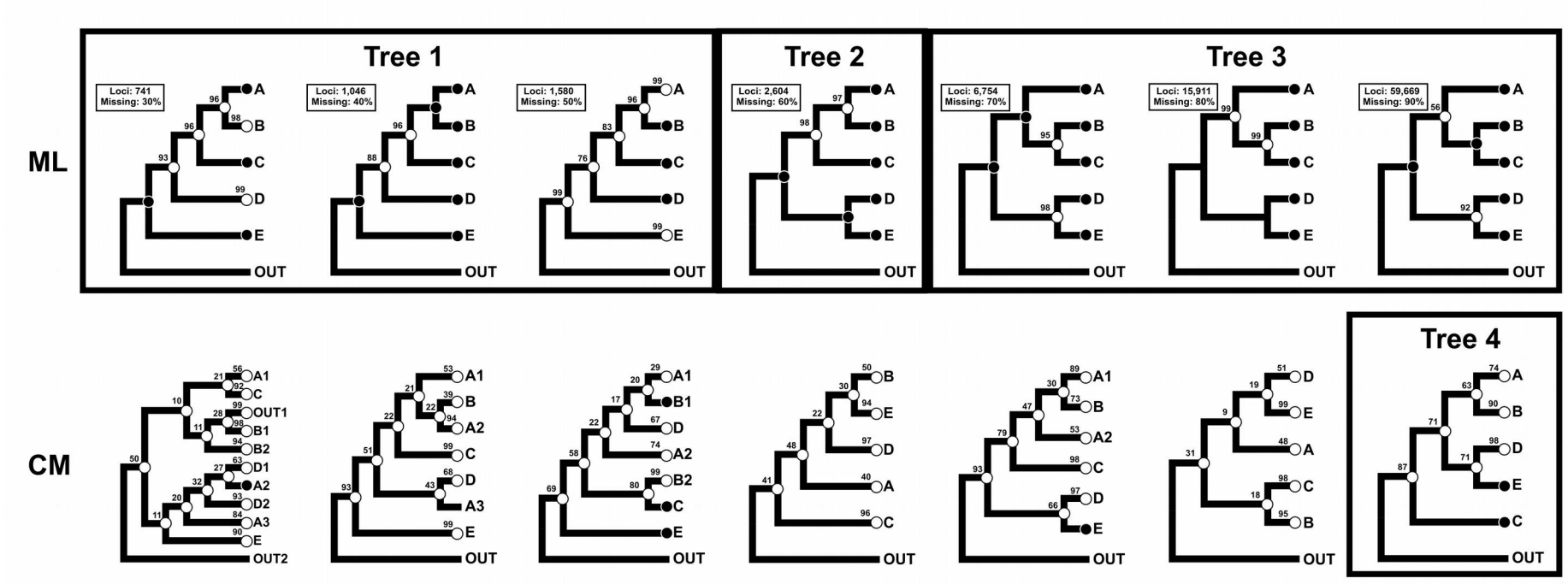


Figure 4: Trees topologies for the genus *Proechimys* based on the strict datasets, with RAD matrices from 30% to 90% of missing data and with indicative number of loci for each dataset. Maximum likelihood (ML) and coalescent model (CM) were used to build the phylogenetic trees. Subclades topologies were the same in all analyses, so we represent here the phylogenetic relationships among the five clades (A, B, C, D and E) and their bootstrap values. To see each topologies in details with individuals as OTU see Fig. S4 to S17. Black circles indicate 100% of bootstrap, white circles are bootstrap values below 100% and the value is shown. OUT = outgroups; subdivisions in the clades (A1, A2, D2) or in the outgroup (OUT1, OUT2) indicate that the clade was not recovered as monophyletic in that analyses. The topologies used in Analysis 2 of the species delimitation with iBPP are within the squares and named as Tree 1, Tree 2, Tree 3, and Tree 4.

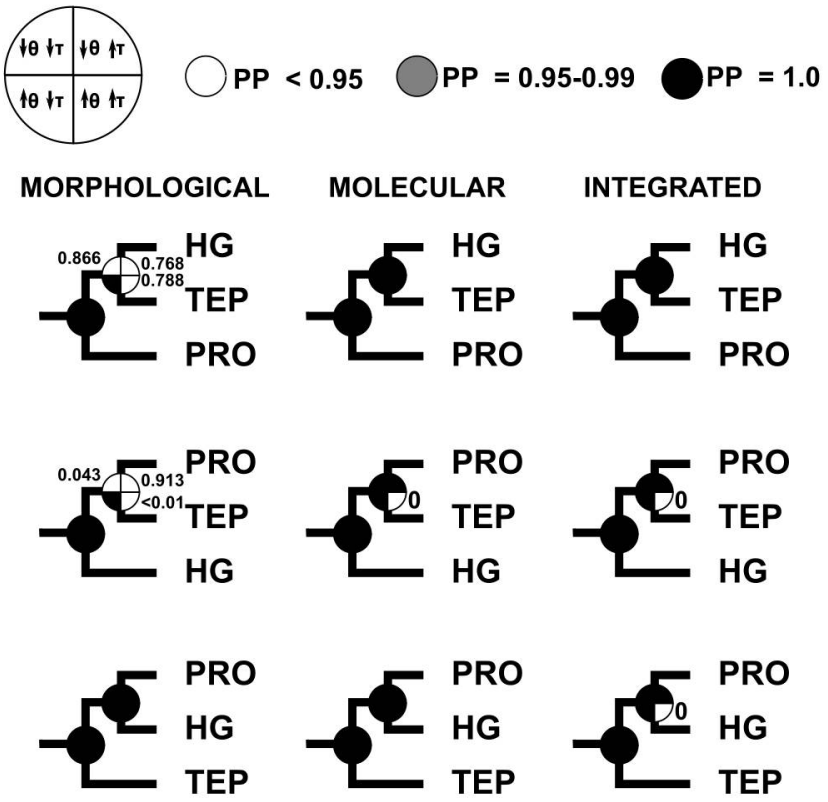


Figure 5: Posterior probabilities (PP) found in simulations for each node of the guide trees used in the Analysis 1 in the iBPP, testing the isolation of *Hoplomys gymnurus* (HG), *Proechimys sensu stricto* (PRO), and *Proechimys* individual from the Tepui region (TEP) in three different topologies. Circles are divided into four portions, each one representing a combination of θ (ancestral population sizes) and τ (divergence times) values, top left: θ = (1, 10) and τ = (1, 10); top right: θ = G (1, 10) and τ = G (2, 2000); bottom left: θ = G (2, 2000) and τ = G (1, 10); bottom right: θ = G (2, 2000) and τ = G (2, 2000). Columns indicates the datasets used in the simulations.

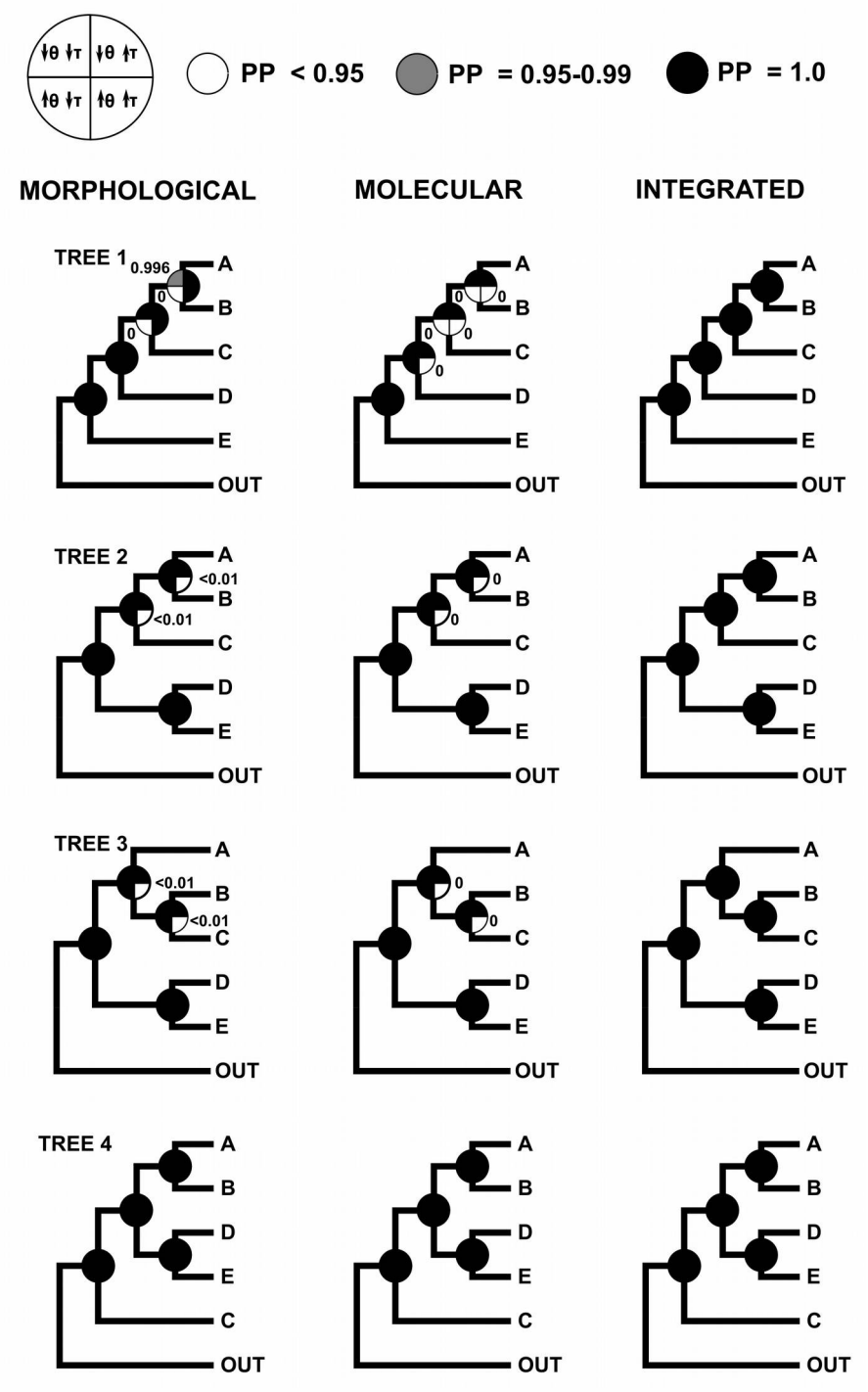


Figure 6: Posterior probabilities (PP) found in simulations for each node of the guide trees used in the Analysis 2 in the iBPP, testing the isolation of the five main clades (A, B, C, D, and E) from genus *Proechimys*. Circles are divided into four portions, each one representing a combination of θ (ancestral population sizes) and τ (divergence times) values, top left: θ = (1, 10) and τ = (1, 10); top right: θ = G (1, 10) and τ = G (2, 2000); bottom left: θ = G (2, 2000) and τ = G (1, 10); bottom right: θ = G (2, 2000) and τ = G (2, 2000). Columns indicates the datasets used in the simulations.

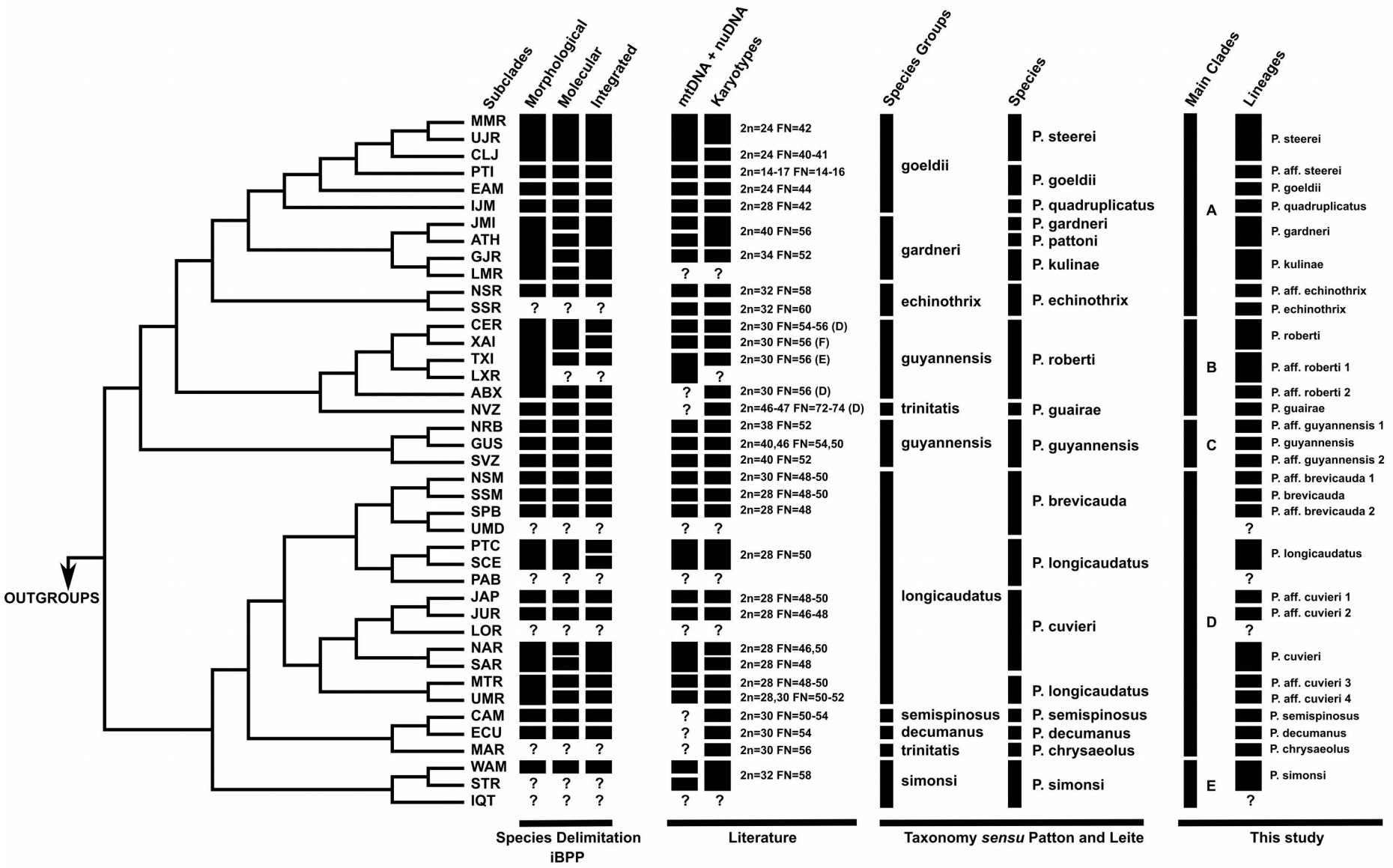

Subclades
Morphological
Molecular
Integrated
mtDNA + nuDNA
Karyotypes
Species Groups
Species
Main Clades
Lineages
OUTGROUPS
MMR
UJR
CLJ
PTI
EAM
IJM
JMI
ATH
GJR
LMR
NSR
SSR
CER
XAI
TXI
LXR
ABX
NVZ
NRB
GUS
SVZ
NSM
SSM
SPB
UMD
PTC
SCE
PAB
JAP
JUR
LOR
NAR
SAR
MTR
UMR
CAM
ECU
MAR
WAM
STR
IQT
2n=24 FN=42
2n=24 FN=40-41
2n=14-17 FN=14-16
2n=24 FN=44
2n=28 FN=42
2n=40 FN=56
2n=34 FN=52
2n=32 FN=58
2n=32 FN=60
2n=30 FN=54-56 (D)
2n=30 FN=56 (F)
2n=30 FN=56 (E)
2n=30 FN=56 (D)
2n=46-47 FN=72-74 (D)
2n=38 FN=52
2n=40,46 FN=54,50
2n=40 FN=52
2n=30 FN=48-50
2n=28 FN=48-50
2n=28 FN=48
2n=28 FN=50
2n=28 FN=48-50
2n=28 FN=46-48
2n=28 FN=46,50
2n=28 FN=48
2n=28 FN=48-50
2n=28,30 FN=50-52
2n=30 FN=50-54
2n=30 FN=54
2n=30 FN=56
2n=32 FN=58
goeldii
gardneri
echinothrix
guyannensis
trinitatis
guyannensis
longicaudatus
semispinosus
decumanus
trinitatis
simonsi
P. steerei
P. goeldii
P. quadruplicatus
P. gardneri
P. pattoni
P. kulinae
P. echinothrix
P. roberti
P. guairae
P. guyannensis
P. brevicauda
P. longicaudatus
P. cuvieri
P. longicaudatus
P. semispinosus
P. decumanus
P. chrysaeolus
P. simonsi
A
B
C
D
E
P. steerei
P. aff. steerei
P. goeldii
P. quadruplicatus
P. gardneri
P. kulinae
P. aff. echinothrix
P. echinothrix
P. roberti
P. aff. roberti 1
P. aff. roberti 2
P. guairae
P. aff. guyannensis 1
P. guyannensis
P. aff. guyannensis 2
P. aff. brevicauda 1
P. brevicauda
P. aff. brevicauda 2
P. longicaudatus
P. aff. cuvieri 1
P. aff. cuvieri 2
P. cuvieri
P. aff. cuvieri 3
P. aff. cuvieri 4
P. semispinosus
P. decumanus
P. chrysaeolus
P. simonsi
Species Delimitation
iBPP
Literature
Taxonomy sensu Patton and Leite
This study

Figure 7: Subclades found in the conservative tree under ML, aligned with (*i*) results of Analysis 3 of species delimitation, using iBPP and three datasets (morphological, molecular, and integrated); (*ii*) literature data on mtDNA + nuDNA, and karyotype information for the same localities or nearby localities, or same samples used in the species delimitation analyses; (*iii*) the current taxonomic proposal for *Proechimys* by Patton and Leite in species groups and valid species; (*iv*) the taxonomic proposal of this study. For Analysis 3: rectangles and squares indicate which subclades were recovered as independent lineages, each column indicates the database used, (?) indicate unused subclades for lack of data. Details for simulations values are shown in Fig. S18. For the literature data: rectangles and squares indicate whether subclades were recovered as a monophyletic clade with other genetic markers (mtDNA + nuDNA), and if whether the subclades had different karyotypes, and 2n and NF are provide. Details for cytogenetic data and how the association was performed are in Table S7. (?) indicate no association with karyotypes or other genetic markers. For this study: The proposal of 28 independent lineages and 4 that need to be evaluated with more data (?) were presented in the last two columns. Subclades were considered independent lineages when they showed PP = 1 most of demographic scenarios for two of the three datasets used in the Analysis 3 of species delimitation.

Supplementary Material

Appendices

Appendix A: Details on preparation and sequencing of the genomic libraries based on the Peterson et al. (2012) protocol.

Genomic DNA from liver and muscle samples were extracted with DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA), following the manufacturer's recommendations, except for the DNA elution step where we used double distilled water (ddH_20) instead of elution buffer. Genomic DNA from skin and dry muscle samples were extracted following the same protocol as fresh tissues but with some modifications before the digestion step. In a sterile environment the hairs were removed from the skin samples. Afterwards, both the skin and dry muscle samples were hydrated for three days with ddH_20, replacing the water every 24 hours. After the hydration, the material was washed twice with 1X STE buffer (Bi *et al.* 2013), and then cut into small pieces to facilitate the digestion. During the digestion step I added 1 mM of dithiotreitol (DTT), a reducing agent, in 20 μL of volume per sample (Rohland & Hofreiter 2007). Extracted DNA was quantified by Qubit fluorometer (Life Technologies, Grand Island, NY, USA), and it was diluted with ddH_20 or concentrated in the SpeedVac Concentrator (ThermoFisher Scientific, Waltham, MA, USA) at 43º C (medium temperature) to reach the concentration of 17.6 ng/μL.

I followed the protocol from Peterson *et al.* (2012) for the preparation of genomic libraries (see Material and Methods for details about the number of samples and libraries) using the ddRAD-Seq technique. In this approach 300 ng of genomic DNA (*i.e.*, 17 μL of extracted DNA) were cut in variable-sized fragments, using two restriction enzymes: *Eco*-RI and *Mse*-I. The resulting solution was cleaned with commercial Ampure XP Beads (Beckman Coulter, Brea, CA, USA) and quantified in Qubit fluorometer (Life Technologies, Grand Island, NY, USA). Then, I used 50 ng of fragmented DNA in a volume of 33 μL per sample for the ligation step, in which the ends of the fragmented DNA were bonded to the Illumina adapters and a unique barcode per sample. After the reaction samples were pooled together and the solution was cleaned again with commercial Ampure XP Beads (Beckman Coulter, Brea, CA, USA). DNA fragments were automatically selected by size (between 350 and 450 bp) through Pippin Prep (Sage Science, Beverly, MA, USA) and selected fragments were

amplified by PCR. The libraries were cleaned with the beads again, quantified and sequenced. All libraries were sequenced in three lanes of HiSeq2500 (Illumina, San Diego, CA, USA) according to instructions of the manufacturer to generate 150 base pairs, single-end reads in the Hospital for Sick Children (Toronto, Ontario, Canada).

Bi, K., Linderoth, T., Vanderpool, D., Good, J.M., Nielsen, R. & Moritz, C. (2013) Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* 22, 6018–6032.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species L. Orlando (Ed). *PLoS ONE* 7, e37135.

Rohland, N. & Hofreiter, M. (2007) Ancient DNA extraction from bones and teeth. *Nature Protocols* 2, 1756–1762.

Appendix B: Details on reads processing for phylogenetic analyses using the ipyrad pipeline.

For the phylogenetic approach, I performed the *ipyrad* pipeline (http://ipyrad.readthedocs.io/) in seven steps. In the first step the raw data was demultiplexed in individuals according to the barcode list I provided, without mismatches in barcodes (*max_barcode_mismatch*). In the second step I edited the reads, eliminating barcodes, adapters and the reads with more than 5 pb with low quality Q>20 (*max_low_qual_bases*) and with low quality scores for Illumina, Phed below 33 (*phred_Qscore_offset*). We allowed reads of variable sizes with 110 pb of minimum size (*filter_min_trim_len*). We grouped the reads as homologous in the third step if they presented a similarity ≥ 90% (*clust_threshold*), and the clusters were *de novo* aligned with Muscle (Edgar 2004). In the step 4, the heterozygosity indices and the sequencing error rates were calculated and used for the fifth step, the consensus step. A consensus sequence was created for each allele from the aligned reads, considering the values of the parameters calculated in the fourth step. In the fifth step the data was also filtered, allowing up to 5 Ns (uncalled bases) per allele consensus sequences (*max_Ns_consens*), and the number of alleles per locus was calculated. In step 6 the consensus sequences are aligned again with Muscle (Edgar 2004) using the parameters of step 3, and in step 7 the outputs were created for the subsequent analyzes with some filters: maximum of 6 indels per locus (*max_Indels_locus*); maximum of 50% of heterozygous sites per locus (*max_shared_Hs_locus*), and up to 2 unique alleles were allowed in an individual (*max_alleles_consens*); and I chose an amount of missing data values per loci (*min_samples_locus*), see Material and Methods to details about this number.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* 32, 1792–7.

Tables

Table S1: 278 Samples used in the genomics analyzes with information about the locality, species groups, and institution of origin of the samples. **AMNH-AMCC:** Ambrose Monell Cryo Collection, American Museum of Natural History, New York, USA; **CIT:** Coleção de Tecidos Miguel Trefaut Rodrigues, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil; **FMNH:** Field Museum of Natural History, Chicago, Illinois, USA; **LMUSP:** Laboratório de Mamíferos da Escola Superior de Agricultura "Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil; **MCN-M:** Coleção de Mastozoologia do Museu de Ciências Naturais da Pontíficia Universidade Católica de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; **MN:** Museu Nacional da Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; **MPEG**: Museu Paraense Emílio Goeldi, Belém, Pará, Brazil; **MSB:** Museum of Southwestern Biology, Alburqueque, New Mexico, USA; **MVZ:** Museum of Vertebrate Zoology, Berkeley, California, USA; **MZUSP:** Museu de Zoologia da Universidade de São Paulo, São Paulo, Brazil; **NMNH:** National Museum of Natural History, Washington, D.C., USA; **ROM:** Royal Ontario Museum, Toronto, Ontario, Canada; TTU: Texas Tech University, Lubbock, Texas, USA; **UFES-CTA:** Coleção de Tecido Animal da Universidade Federal do Espírito Santo, Vitória, Espírito Santo; Brazil; **UFMG:** Coleção de Mamíferos, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; **UFPB:** Univerisade Federal da Paraíba, João Pessoa, Paraíba, Brazil; **UMMZ:** University of Michigan Museum of Zoology, Ann Arbor, Michigan, USA. Table with more information about locality in ***.csv*** format is available on https://github.com/jdalapicolla.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
756		MZUSP	goeldii	-55.78638	-14.87316
ABX005		LMUSP	guyannensis	-58.63314	-4.31203
ABX008		LMUSP	guyannensis	-58.63442	-4.34328
ABX020		LMUSP	guyannensis	-58.63442	-4.34328
ABX027		LMUSP	goeldii	-58.63314	-4.31203
ABX028		LMUSP	guyannensis	-58.63314	-4.31203
ABX029		LMUSP	guyannensis	-58.63442	-4.34328
ABX077		LMUSP	guyannensis	-58.20922	-4.58208
AMCC112929	USNM448714	AMNH-AMCC	trinitatis	-72.8091	9.84363
AMCC112987	USNM448733	AMNH-AMCC	trinitatis	-72.8091	9.84363

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
AMCC112999	USNM448711	AMNH-AMCC	guyannensis	-61.09269	5.07502
AMCC114577	MUSM23828	AMNH-AMCC	gardneri	-73.66744	-4.90625
AMCC114589	MUSM23829	AMNH-AMCC	simonsi	-73.66744	-4.90625
AMCC175984	JOG4521; EBRG25376	AMNH-AMCC	trinitatis	-69.63328	11.13328
AMCC176080	JOG4488; EBRG25473	AMNH-AMCC	trinitatis	-69.95012	11.81811
AMNH23109		AMNH-AMCC	trinitatis	-74.11742	11.14226
AMNH235152		AMNH-AMCC	trinitatis	-61.46712	10.41568
AMNH269122		AMNH-AMCC	longicaudatus	-52.92366	5.27438
AMNH269123	LHE1163	AMNH-AMCC	longicaudatus	-52.92366	5.27438
AMNH272698	RSV2092	AMNH-AMCC	longicaudatus	-73.16208	-5.2495
AMNH272700	RSV2095	AMNH-AMCC	longicaudatus	-73.16208	-5.2495
AMNH272714	RSV2132	AMNH-AMCC	gardneri	-73.16208	-5.2495
AMNH64659		AMNH-AMCC	decumanus	-80.71113	-1.93681
AMNH78026		AMNH-AMCC	goeldii	-67.68961	4.03523
APC1085		MZUSP	longicaudatus	-57.21464	-15.65353
APC1217		MZUSP	guyannensis	-46.70895	-10.85524
APC817		MZUSP	guyannensis	-47.98038	-16.01777
APC839		MZUSP	guyannensis	-47.98038	-16.01777
BAC320	UFES1580	CTA-UFES	guyannensis	-50.42313	-6.34376
BM12174		LMUSP	goeldii	-51.90807	-3.38232
BMC111589		LMUSP	guyannensis	-51.81476	-3.13498
BMC111682		LMUSP	goeldii	-51.75418	-3.27342
C247647	C247647-6738	MZUSP	goeldii	-56.51269	-9.45183
CAM091		MZUSP	goeldii	-58.95777	-12.9896
CAM200		MZUSP	goeldii	-58.95777	-12.9896
CIT375	PEU960021	CIT	longicaudatus	-59.44718	-10.17484
CIT393	PEU960065	CIT	longicaudatus	-59.44718	-10.17484

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
CIT405	M968406	CIT	longicaudatus	-57.39563	-9.5678
CIT448	M968498	CIT	longicaudatus	-57.39563	-9.5678
CIT588	M088	CIT	goeldii	-53.25595	-13.18229
CTA1028	YL53	CTA-UFES	*E. spinosus*	-43.5	-20.083
CTA1349	MVZ197574; LPC394	CTA-UFES	longicaudatus	-52.35583	-15.63333
CTA1415	MVZ197661; LPC462	CTA-UFES	longicaudatus	-52.35583	-15.63333
CTA1511	UFMG3029	CTA-UFES	guyannensis	-55.93028	-9.59694
CTA1517	MVZ197575; LPC564	CTA-UFES	goeldii	-55.93028	-9.59694
CTA1652	UFMG3035	CTA-UFES	guyannensis	-48.63556	-11.84278
CTA1835	UFES1390	CTA-UFES	guyannensis	-49.95863	-9.30361
CTA4195	MPEG42380; SLF225	CTA-UFES	guyannensis	-56.36422	-1.79591
CTA4226	MPEG42412; SLF309	CTA-UFES	guyannensis	-56.36422	-1.79591
CTA4245	MPEG42367	CTA-UFES	longicaudatus	-56.36422	-1.79591
CTA4324	UFES2637	CTA-UFES	goeldii	-59.1306	-3.57972
CTA4325	UFES2638	CTA-UFES	goeldii	-59.1306	-3.57972
CTA4326	UFES2639	CTA-UFES	goeldii	-59.1306	-3.57972
CTA4327	UFES2640	CTA-UFES	goeldii	-59.1306	-3.57972
CTA4352	UFES2705	CTA-UFES	gardneri	-59.1306	-3.57972
CTA4357	UFES2710	CTA-UFES	goeldii	-59.0942	-3.88806
CTA4363	UFES2834	CTA-UFES	gardneri	-59.1306	-3.57972
CTA4371	UFES2842	CTA-UFES	goeldii	-59.1306	-3.57972
CTA4390	UFES2945; BM74724	CTA-UFES	guyannensis	-51.77012	-3.12428
CTA4400	UFES2962; BM101571	CTA-UFES	goeldii	-51.77012	-3.12428
DPO18	UFES1569	CTA-UFES	guyannensis	-50.44612	-6.34786
DPO19	UFES1570	CTA-UFES	guyannensis	-50.36045	-6.38747
EEB1013		LMUSP	*T. dimidiatus*	-44.368	-22.80657
EFA004		LMUSP	goeldii	-61.82882	-4.35058
EFA015		LMUSP	goeldii	-61.82744	-4.40803
EFA037		LMUSP	simonsi	-62.26745	-4.42802
EFA039		LMUSP	simonsi	-62.30707	-4.43154
FMNH175255	UPE133	FMNH	longicaudatus	-71.38542	-12.77165
FMNH175256	SS2146	FMNH	longicaudatus	-71.38542	-12.77165
FMNH175275	UPE231	FMNH	simonsi	-71.49185	-13.02362
FMNH26441		FMNH	guyannensis	-45.78333	-9.1
FMNH52618		FMNH	longicaudatus	-69.68333	-13.85

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
FMNH71184	PH6537	FMNH	trinitatis	-74.1	-5.5333
FMPS010		LMUSP	*C. laticeps*	-57.65317	-18.99968
ICA085		LMUSP	echinothrix	-68.34741	-2.87486
ICA095		LMUSP	goeldii	-68.33989	-2.90496
ICA240		LMUSP	echinothrix	-68.87975	-3.03818
ICA245		LMUSP	echinothrix	-68.88251	-3.03233
ICA246		LMUSP	goeldii	-68.88524	-3.02949
JAP006		LMUSP	goeldii	-65.75583	-1.76103
JAP012		LMUSP	echinothrix	-65.75583	-1.76103
JAP094		LMUSP	longicaudatus	-66.35717	-1.76422
JAP095		LMUSP	longicaudatus	-66.35717	-1.76422
JAP098		LMUSP	longicaudatus	-66.35717	-1.76422
JAP107		LMUSP	longicaudatus	-66.35717	-1.76422
JUF017		LMUSP	guyannensis	-62.15008	-0.94917
JUF158		LMUSP	guyannensis	-62.09181	-1.02889
LMUSP185		LMUSP	longicaudatus	-68.74681	-10.99838
LMUSP298		LMUSP	longicaudatus	-69.56473	-10.93071
MBR046		LMUSP	*T. pachyurus*	-54.61611	-20.46514
MCNM1341		MCN-M	guyannensis	-44.35094	-2.56688
MCNM1497	LGV161	MCN-M	guyannensis	-49.72934	-6.43589
MCNM1989	LOM37	MCN-M	guyannensis	-47.39486	-17.25773
MCNM2034	LOM35	MCN-M	guyannensis	-47.39486	-17.25773
MCNM2258	BM78	MCN-M	longicaudatus	-50.06717	-6.0548
MJ12		MZUSP	gardneri	-64.94222	-9.4139
MJ252		MZUSP	goeldii	-65.34884	-9.59557
MJ330		MZUSP	longicaudatus	-64.79292	-9.42976
MJ391		MZUSP	goeldii	-64.83327	-9.43604
MJ449		MZUSP	gardneri	-65.46861	-9.61306
MJ514		MZUSP	gardneri	-64.87107	-9.44449
MJ515		MZUSP	gardneri	-64.85128	-9.45414
MJ523		MZUSP	gardneri	-65.4533	-9.63504
MJ529		MZUSP	simonsi	-65.43986	-9.63406
MJ6		MZUSP	gardneri	-64.94222	-9.4139
MJ601		MZUSP	goeldii	-65.0713	-9.56918
MJ665		MZUSP	gardneri	-64.85128	-9.45414

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
MJ859		MZUSP	longicaudatus	-65.445	-9.62658
MN36222	MNLM236	MN	guyannensis	-48.30485	-13.83346
MN36702	MNLM262	MN	guyannensis	-48.30485	-13.83346
MN56812	MNLM519	MN	guyannensis	-64.78914	1.2086
MN56815	MNLM521	MN	guyannensis	-64.78914	1.2086
MN56816	MNLM522	MN	guyannensis	-64.78914	1.2086
MN67246	APC599	CIT	longicaudatus	-52.92449	-18.12468
MN76204	MNLM2306	MN	guyannensis	-48.3	-5.28333
MN76211	MNLM2312	MN	guyannensis	-48.3	-5.28333
MN76750	MNLM2337	MN	guyannensis	-48.29032	-10.0782
MN76754	MNLM2341	MN	guyannensis	-48.29032	-10.0782
MPEG10658		MPEG	simonsi	-69.26065	-8.84739
MPEG10811		MPEG	longicaudatus	-68.67144	-9.07918
MPEG10820		MPEG	longicaudatus	-68.67144	-9.07918
MPEG20767	BDP2122	FMNH	longicaudatus	-61.93104	-10.84726
MPEG20768	BDP2186	FMNH	longicaudatus	-61.93104	-10.84726
MPEG20769	BDP2177	FMNH	longicaudatus	-61.93104	-10.84726
MPEG21336		MPEG	longicaudatus	-61.93104	-10.84726
MPEG21338		MPEG	longicaudatus	-61.93104	-10.84726
MPEG21983		MPEG	guyannensis	-46.33483	-3.70759
MPEG22881		MPEG	gardneri	-64.71989	-3.3539
MPEG26357		MPEG	simonsi	-65.70879	-2.21689
MPEG34016		MPEG	goeldii	-63.07166	-7.55169
MPEG34407		MPEG	goeldii	-63.07166	-7.55169
MPEG40369	EPM07	MPEG	goeldii	-51.45549	-1.7374
MPEG40371	EPM04	MPEG	guyannensis	-51.45549	-1.7374
MRT3925		CIT	guyannensis	-47.87264	-12.61673
MSA-MC123		MZUSP	goeldii	-63.9505	-8.80237
MSA-SA110		MZUSP	longicaudatus	-63.9505	-8.80237
MSB140110		MSB	decumanus	-80.11667	-3.88333
MSB140111		MSB	decumanus	-80.75	-1.38333
MSB208394		MSB	longicaudatus	-65.51084	-14.01672
MSB210840		MSB	longicaudatus	-62.75	-17.65
MSB211792		MSB	longicaudatus	-66.13381	-11.03798
MSB211815		MSB	goeldii	-66.21667	-11.01667

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
MSB236570		MSB	gardneri	-68.91681	-11.3501
MSB236594		MSB	simonsi	-68.91681	-11.3501
MSB236689		MSB	goeldii	-67.56023	-11.49004
MSB236698		MSB	goeldii	-68.84981	-11.35059
MSB236806		MSB	longicaudatus	-66.77966	-11.74947
MSB236807		MSB	goeldii	-66.77966	-11.74947
MSB238391		MSB	longicaudatus	-65.55	-17.1
MSB239628		MSB	longicaudatus	-63.11667	-17.88333
MSB263513		MSB	*H. gymnurus*	-81.15074	8.53204
MSB45836		MSB	semispinosus	-84.7178	15.84003
MSB70574		MSB	longicaudatus	-65.55	-17.1
MSB70575		MSB	longicaudatus	-65.46667	-17.05
MSB99059		MSB	longicaudatus	-64.49139	-14.89556
MSB99060		MSB	longicaudatus	-64.49139	-14.89556
MUSM13338	RSV2120	AMNH-AMCC	longicaudatus	-73.16208	-5.2495
MUSM13339	RSV2033	AMNH-AMCC	simonsi	-73.16208	-5.2495
MUSM13342	RSV2076	AMNH-AMCC	simonsi	-73.16208	-5.2495
MVZ136648		MVZ	longicaudatus	-71.2166	-10.1333
MVZ155121		MVZ	longicaudatus	-78.16833	-4.45667
MVZ157855	JLP8271	MVZ	longicaudatus	-69.068	-12.63333
MVZ157905		MVZ	longicaudatus	-77.751	-4.022
MVZ157968		MVZ	simonsi	-77.751	-4.022
MVZ160093	JLP9039	MVZ	longicaudatus	-61.43333	6.15
MVZ160094	JLP9044	MVZ	guyannensis	-61.66667	4.46667
MVZ162309		MVZ	*H. gymnurus*	-77.0185	3.93699
MVZ166815	JLP11091	MVZ	simonsi	-71.26166	-12.68001
MVZ168942	RMW675	MVZ	goeldii	-69.07289	-12.6
MVZ168945	I080	MVZ	trinitatis	-69.01667	8.95
MVZ168948	AN121	MVZ	canicollis	-72.08333	10.98333
MVZ168949	I230	MVZ	trinitatis	-61.63333	10.6833
MVZ168953	EY624	MVZ	longicaudatus	-69.07289	-12.6
MVZ168955	EY631	MVZ	simonsi	-69.07289	-12.6
MVZ190699	JLP15922	MVZ	longicaudatus	-68.76672	-6.46666
MVZ190951		MVZ	goeldii	-66.2333	-3.2833
MVZ190954		MVZ	goeldii	-66	-3.31667

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
MVZ194439	INPA3442	MVZ	longicaudatus	-72.78304	-8.66663
MVZ194463	MPEG28360; MNFS1606	MVZ	longicaudatus	-72.81667	-8.36666
MVZ194474	MPEG28366; JUR236	MVZ	longicaudatus	-72.81667	-8.36666
MVZ194485	MPEG28371; MNFS1486	MVZ	longicaudatus	-72.81662	-8.36666
MVZ194491	MPEG28377; JLP15638	MVZ	longicaudatus	-70.85003	-6.75
MVZ194492	MPEG28364; JUR187	MVZ	longicaudatus	-68.76672	-6.46666
MVZ194493	MPEG28367; MNFS331	MVZ	longicaudatus	-70.75008	-6.83344
MVZ194511	MPEG25501; JUR297; MPEG28378	MVZ	echinothrix	-66.01666	-3.31666
MVZ194545	MPEG25512; MNFS857; MPEG28381	MVZ	gardneri	-68.9002	-6.58334
MVZ194567	MPEG25505; MNFS541; MPEG28385	MVZ	gardneri	-70.85003	-6.75
MVZ194582	MPEG25509; MNFS1166; MPEG28392	MVZ	gardneri	-72.78304	-8.66663
MVZ194602	MPEG28409; JUR250	MVZ	simonsi	-72.81662	-8.36666
MVZ194635	MPEG28447; MNFS1316	MVZ	simonsi	-72.78304	-8.66663
MVZ194703	MPEG28398; JUR195	MVZ	simonsi	-68.89219	-6.58282
MVZ194711	MPEG28417; JUR302	MVZ	simonsi	-66.01666	-3.31666
MVZ194775	MPEG28427; MNFS724	MVZ	simonsi	-68.76672	-6.46666
MVZ194874	MPEG28572; MNFS1507	MVZ	goeldii	-72.81662	-8.36666
MVZ194879	MPEG28575; MNFS1548	MVZ	goeldii	-72.81667	-8.36666
MVZ194909	MPEG28486; JUR68	MVZ	goeldii	-70.73359	-6.80001
MVZ194914	MPEG28491; JUR81	MVZ	goeldii	-70.73359	-6.80001
MVZ194987	MPEG28540; MNFS497	MVZ	goeldii	-70.75008	-6.83344
MVZ194997	MPEG28543; MNFS599	MVZ	goeldii	-70.85003	-6.75
MVZ195034	MPEG28562; MNFS711	MVZ	goeldii	-68.76667	-6.46669
MVZ195036	MPEG28564; MNFS715	MVZ	goeldii	-68.76667	-6.46669
MVZ196095		MVZ	trinitatis	-73.9511	6.31417
MVZ225064		MVZ	semispinosus	-82.58217	8.46342
MVZ225082		MVZ	*H. gymnurus*	-79.92578	8.68753
MZUSP30365	UUPI326; CIT1449	CIT	guyannensis	-45.20264	-8.86342
MZUSP30370	UUPI412; CIT1465	CIT	guyannensis	-45.20264	-8.86342
MZUSP31924	APC176; CIT648	CIT	goeldii	-58.49231	-10.32276
MZUSP31926	APC209; CIT680	CIT	longicaudatus	-59.44718	-10.17484
MZUSP31927	APC251; CIT703	CIT	longicaudatus	-59.44718	-10.17484
MZUSP31937	M000142; CIT622	CIT	longicaudatus	-58.49231	-10.32276

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
MZUSP31939	APC169; CIT642	CIT	longicaudatus	-58.49231	-10.32276
MZUSP31942	M97032; CIT483	CIT	guyannensis	-54.87472	-11.50302
MZUSP31944	M97180; CIT511	CIT	guyannensis	-54.87472	-11.50302
MZUSP31945	M000090; CIT589	CIT	guyannensis	-53.25595	-13.18229
MZUSP31946	M000097; CIT591	CIT	guyannensis	-53.25595	-13.18229
MZUSP31947	APC273; CIT714	CIT	guyannensis	-51.11932	-10.01433
MZUSP31948	APC274; CIT753	CIT	guyannensis	-51.11932	-10.01433
MZUSP31950	APC825	MZUSP	guyannensis	-48.54	-12.02688
MZUSP31951	APC847	MZUSP	guyannensis	-48.54	-12.02688
NUTRIA289		CTA-UFES	*M. coypus*	NA	NA
PLVP642		LMUSP	echinothrix	-62.48921	-2.30486
PNPA321		MZUSP	longicaudatus	-57.16668	-19.63333
PNPA357		MZUSP	longicaudatus	-57.16668	-19.63333
RGM853		MZUSP	goeldii	-58.97905	-15.13767
RGM856		MZUSP	goeldii	-58.97905	-15.13767
ROM115116		ROM	trinitatis	-60.76667	5.33333
ROM117526		ROM	guyannensis	-56.88556	4.46694
ROM119901		ROM	longicaudatus	-58.90933	4.24963
TL17240		MZUSP	goeldii	-51.98652	-7.70104
TL17273		MZUSP	guyannensis	-51.98652	-7.70104
TTU100580	TK119156	TTU	semispinosus	-84.39952	11.67685
TTU103310	TK134958	TTU	decumanus	-80.02158	-2.42728
TTU102638	TK135304	TTU	decumanus	-80.09286	-3.8795
TTU102971	TK135716	TTU	semispinosus	-78.70597	1.09356
TTU102977	TK135943	TTU	semispinosus	-78.70597	1.09356
TTU106013	TK145304	TTU	guyannensis	-54.73945	4.26732
TTU34990	TK14603	TTU	longicaudatus	-67.53333	-15.46667
TTU34991	TK14609	TTU	longicaudatus	-67.53333	-15.46667
TTU46355	TK22895	TTU	longicaudatus	-75.95	-9.2
TTU46356	TK22911	TTU	longicaudatus	-75.95	-9.2
TTU101118	TK73760	TTU	simonsi	-73.26836	-4.02398
TTU101173	TK73888	TTU	longicaudatus	-73.26836	-4.02398
TTU101179	TK73909	TTU	longicaudatus	-73.26836	-4.02398
TTU101195	TK73940	TTU	longicaudatus	-73.26836	-4.02398
TTU101213	TK73977	TTU	longicaudatus	-73.26836	-4.02398

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
UFMG4356		UFMG	guyannensis	-49.85198	-6.53241
UFPB1015		UFPB	longicaudatus	-59.499	-1.93314
UFPB1265		UFPB	longicaudatus	-63.45693	-8.75272
UFPB1266		UFPB	longicaudatus	-63.45693	-8.75272
UFPB2990		UFPB	simonsi	-76.1094	-5.89482
UFPB6734		UFPB	guyannensis	-48.33851	-10.71415
UFPB6929		UFPB	guyannensis	-44.41654	-2.33712
UFPB6931		UFPB	guyannensis	-44.41654	-2.33712
UFPB6932		UFPB	guyannensis	-44.41654	-2.33712
UMMZ80045		UMMZ	simonsi	-76.81724	-0.43999
UMMZ80079		UMMZ	goeldii	-76.81724	-0.43999
UMMZ92712		UMMZ	semispinosus	-73.61667	4.15
UNIBAN1392		CIT	longicaudatus	-63.22963	-10.21504
USNM499715		USNM	trinitatis	-75.07118	7.30036
USNM549559		USNM	longicaudatus	-52.37	-3.65
USNM549567		USNM	longicaudatus	-52.37	-3.65
USNM568055		USNM	longicaudatus	-60.4911	7.3706
USNM579697		USNM	simonsi	-69.6844	-13.5044
USNM581908		USNM	longicaudatus	-59.0267	-18.0583
USNM584593		USNM	longicaudatus	-61.0347	-14.7672
USNM619001	LHE0742	USNM	simonsi	-68.7667	-13.5833
USNM619002	LHE0747	USNM	goeldii	-68.7667	-13.5833
USNM619003	LHE0820	USNM	simonsi	-69.68355	-13.49957
USNM619004	ALG14009	USNM	guyannensis	-67.01537	1.91987
USNM619005	TTS382	USNM	longicaudatus	-57.86666	-19.18331
USNM619006	NBH1305	USNM	longicaudatus	-60.85286	-14.3699
USNM619007	LLW424	USNM	simonsi	-73.34073	-11.77955
USNM619008	ACF076	USNM	simonsi	-68.88173	-12.95664
X1M15		LMUSP	longicaudatus	-49.94611	-4.25168
X1M19		LMUSP	guyannensis	-50.63753	-3.83923
X1M24		LMUSP	longicaudatus	-50.63753	-3.83923
X1M36		LMUSP	goeldii	-49.09108	-5.34573
Z9P01		LMUSP	guyannensis	-48.91626	-14.97213

Table S2: 479 Morphometric samples with information about the about the locality, species groups, institution of origin of the samples, locality, sample groupings for the three analyzes carried out in the iBPP. **AMNH:** American Museum of Natural History, New York, USA; **FMNH:** Field Museum of Natural History, Chicago, Illinois, USA; **LMUSP:** Laboratório de Mamíferos da Escola Superior de Agricultura "Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil; **MCN-M:** Coleção de Mastozoologia do Museu de Ciências Naturais da Pontíficia Universidade Católica de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; **MN:** Museu Nacional da Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; **MPEG:** Museu Paraense Emílio Goeldi, Belém, Pará, Brazil; **MVZ:** Museum of Vertebrate Zoology, Berkeley, California, USA; **MZUSP**: Museu de Zoologia da Universidade de São Paulo, São Paulo, Brazil; **NMNH:** National Museum of Natural History, Washington, D.C., USA; **UFES-MAM:** Coleção de Mamíferos da Universidade Federal do Espírito Santo, Vitória, Espírito Santo; Brazil; **UFMG:** Coleção de Mamíferos, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; **UFPB:** Univerisade Federal da Paraíba, João Pessoa, Paraíba, Brazil; **UMMZ:** University of Michigan Museum of Zoology, Ann Arbor, Michigan, USA. Table with raw data for the 22 variables in ***.csv*** format is available on https://github.com/jdalapicolla.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
663	LMUSP	goeldii	PRO	A	EAM	-51.77435	-3.12264
756	MZUSP	goeldii	PRO	A	PTI	-55.78638	-14.87316
ABX005	LMUSP	guyannensis	PRO	B	ABX	-58.63314	-4.31203
ABX008	LMUSP	guyannensis	PRO	B	ABX	-58.63442	-4.34328
ABX020	LMUSP	guyannensis	PRO	B	ABX	-58.63442	-4.34328
ABX027	LMUSP	goeldii	PRO	A	PTI	-58.63314	-4.31203
ABX077	LMUSP	guyannensis	PRO	B	ABX	-58.20922	-4.58208
AMNH141874	AMNH	outgroup	HG	-	-	-83.61152	9.96607
AMNH18811	AMNH	semispinosus	PRO	D	CAM	-82.57032	8.5078
AMNH18813	AMNH	semispinosus	PRO	D	CAM	-82.57032	8.5078
AMNH18817	AMNH	semispinosus	PRO	D	CAM	-82.57032	8.5078
AMNH18818	AMNH	semispinosus	PRO	D	CAM	-82.57032	8.5078
AMNH214680	AMNH	longicaudatus	PRO	D	PTC	-64.7884	-15978
AMNH214681	AMNH	longicaudatus	PRO	D	PTC	-64.84949	-16.0618
AMNH214700	AMNH	longicaudatus	PRO	D	PTC	-64.8935	-16.14017
AMNH235203	AMNH	guyannensis	PRO	C	GUS	-54.39421	4.31816
AMNH262314	AMNH	longicaudatus	PRO	D	PTC	-63.15	-15.71667

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
AMNH262316	AMNH	longicaudatus	PRO	D	PTC	-63.15	-15.71667
AMNH262317	AMNH	longicaudatus	PRO	D	PTC	-63.15	-15.71667
AMNH262320	AMNH	longicaudatus	PRO	D	PTC	-63.15	-15.71667
AMNH262321	AMNH	longicaudatus	PRO	D	PTC	-63.15	-15.71667
AMNH263054	AMNH	longicaudatus	PRO	D	SPB	-66.77966	-11.74947
AMNH263056	AMNH	goeldii	PRO	A	MMR	-67.2	-11.38333
AMNH263057	AMNH	goeldii	PRO	A	MMR	-67.21667	-11.4
AMNH263060	AMNH	simonsi	PRO	E	WAM	-68.57543	-12.39863
AMNH263062	AMNH	simonsi	PRO	E	WAM	-68.91681	-11.3501
AMNH263108	AMNH	goeldii	PRO	A	MMR	-66.21667	-11.01667
AMNH263109	AMNH	goeldii	PRO	A	MMR	-67.2	-11.38333
AMNH263112	AMNH	goeldii	PRO	A	MMR	-68.57543	-12.39863
AMNH263113	AMNH	goeldii	PRO	A	MMR	-68.57543	-12.39863
AMNH263122	AMNH	longicaudatus	PRO	D	SPB	-66.77966	-11.74947
AMNH264881	AMNH	longicaudatus	PRO	D	PTC	-67.06667	-15.28333
AMNH266594	AMNH	longicaudatus	PRO	D	NAR	-52.92366	5.27438
AMNH266595	AMNH	guyannensis	PRO	C	GUS	-52.92366	5.27438
AMNH267025	AMNH	longicaudatus	PRO	D	NAR	-52.92366	5.27438
AMNH267027	AMNH	longicaudatus	PRO	D	NAR	-52.92366	5.27438
AMNH267028	AMNH	longicaudatus	PRO	D	NAR	-52.92366	5.27438
AMNH267029	AMNH	longicaudatus	PRO	D	NAR	-52.92366	5.27438
AMNH267037	AMNH	guyannensis	PRO	C	GUS	-52.92366	5.27438
AMNH267039	AMNH	longicaudatus	PRO	D	NAR	-52.92366	5.27438
AMNH267047	AMNH	guyannensis	PRO	C	GUS	-52.92366	5.27438
AMNH267602	AMNH	longicaudatus	PRO	D	NAR	-52.92366	5.27438
AMNH268278	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
AMNH268281	AMNH	longicaudatus	PRO	D	SSM	-73.16208	-5.2495
AMNH272699	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
AMNH272700	AMNH	longicaudatus	PRO	D	SSM	-73.16208	-5.2495
AMNH272716	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
AMNH272717	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
AMNH28517	AMNH	semispinosus	PRO	D	CAM	-86.57038	12.98484
AMNH29548	AMNH	outgroup	HG	-	-	-87.22868	12.87533
AMNH30189	AMNH	guyannensis	PRO	C	SVZ	-65.11667	7.15
AMNH30205	AMNH	guyannensis	PRO	C	SVZ	-65.11667	7.15

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
AMNH30722	AMNH	guyannensis	PRO	C	SVZ	-61.68548	6.7465
AMNH30723	AMNH	guyannensis	PRO	C	SVZ	-61.68548	6.7465
AMNH30724	AMNH	guyannensis	PRO	C	SVZ	-61.68548	6.7465
AMNH33212	AMNH	semispinosus	PRO	D	CAM	-79.65856	0.96441
AMNH33213	AMNH	semispinosus	PRO	D	CAM	-79.65856	0.96441
AMNH34211	AMNH	outgroup	HG	-	-	-78.08333	1.48333
AMNH63081	AMNH	decumanus	PRO	D	ECU	-80.07814	-2.23388
AMNH64663	AMNH	decumanus	PRO	D	ECU	-80.4267	-0.61975
AMNH64666	AMNH	decumanus	PRO	D	ECU	-80.4267	-0.61975
AMNH75633	AMNH	trinitatis	TEP	-	-	-60.76443	5.17078
APC1034	MZUSP	longicaudatus	PRO	D	SCE	-57.21464	-15.65353
APC1069	MZUSP	longicaudatus	PRO	D	SCE	-57.21464	-15.65353
APC1072	MZUSP	longicaudatus	PRO	D	SCE	-57.21464	-15.65353
APC1085	MZUSP	longicaudatus	PRO	D	SCE	-57.21464	-15.65353
APC1095	MZUSP	longicaudatus	PRO	D	SCE	-57.21464	-15.65353
APC342	MZUSP	guyannensis	PRO	B	XAI	-51.11932	-10.01433
APC344	MZUSP	guyannensis	PRO	B	XAI	-51.11932	-10.01433
CIT588	MZUSP	goeldii	PRO	A	EAM	-53.25595	-13.18229
CIT644	MZUSP	goeldii	PRO	A	PTI	-58.49231	-10.32276
EFA004	LMUSP	goeldii	PRO	A	PTI	-61.82882	-4.35058
EFA015	LMUSP	goeldii	PRO	A	PTI	-61.82744	-4.40803
EFA038	LMUSP	simonsi	PRO	E	WAM	-62.26745	-4.42802
ICA027	LMUSP	echinothrix	PRO	A	NSR	-68.35518	-2.89186
ICA095	LMUSP	goeldii	PRO	A	IJM	-68.33989	-2.90496
ICA240	LMUSP	echinothrix	PRO	A	NSR	-68.87975	-3.03818
ICA246	LMUSP	goeldii	PRO	A	IJM	-68.88524	-3.02949
ICA269	LMUSP	echinothrix	PRO	A	NSR	-68.88154	-3.03794
JAP006	LMUSP	goeldii	PRO	A	IJM	-65.75583	-1.76103
JAP013	LMUSP	goeldii	PRO	A	IJM	-65.75583	-1.76103
JAP085	LMUSP	echinothrix	PRO	A	NSR	-66.35717	-1.76422
JAP086	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP094	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP095	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP098	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP099	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
JAP102	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP107	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP108	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP109	LMUSP	echinothrix	PRO	A	NSR	-66.35717	-1.76422
JAP110	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP116	LMUSP	longicaudatus	PRO	D	JAP	-65.75918	-1.76611
JAP117	LMUSP	longicaudatus	PRO	D	JAP	-65.75918	-1.76611
JAP118	LMUSP	echinothrix	PRO	A	NSR	-65.75918	-1.76611
JAP122	LMUSP	echinothrix	PRO	A	NSR	-65.75918	-1.76611
JAP124	LMUSP	longicaudatus	PRO	D	JAP	-65.75918	-1.76611
JAP133	LMUSP	echinothrix	PRO	A	NSR	-66.35717	-1.76422
JAP145	LMUSP	echinothrix	PRO	A	NSR	-65.75918	-1.76611
JAP152	LMUSP	echinothrix	PRO	A	NSR	-66.35717	-1.76422
JAP155	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP160	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP161	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP162	LMUSP	longicaudatus	PRO	D	JAP	-65.75918	-1.76611
JAP168	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP173	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP179	LMUSP	longicaudatus	PRO	D	JAP	-65.75918	-1.76611
JAP192	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JAP193	LMUSP	echinothrix	PRO	A	NSR	-66.35717	-1.76422
JAP208	LMUSP	echinothrix	PRO	A	NSR	-65.75918	-1.76611
JAP209	LMUSP	echinothrix	PRO	A	NSR	-66.35717	-1.76422
JAP223	LMUSP	echinothrix	PRO	A	NSR	-66.35717	-1.76422
JAP225	LMUSP	longicaudatus	PRO	D	JAP	-66.35717	-1.76422
JUF009	LMUSP	guyannensis	PRO	C	NRB	-62.13603	-0.95611
JUF017	LMUSP	guyannensis	PRO	C	NRB	-62.15008	-0.94917
JUF028	LMUSP	guyannensis	PRO	C	NRB	-62.13603	-0.95611
JUF164	LMUSP	guyannensis	PRO	C	NRB	-62.09181	-1.02889
M968564	MZUSP	guyannensis	PRO	B	TXI	-54.87472	-11.50302
M97029	MZUSP	guyannensis	PRO	B	TXI	-54.87472	-11.50302
M97124	MZUSP	guyannensis	PRO	B	TXI	-54.87472	-11.50302
M97184	MZUSP	goeldii	PRO	A	EAM	-54.87472	-11.50302
M976265	MZUSP	guyannensis	PRO	B	TXI	-54.87472	-11.50302

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
MCNM1380	MCN-M	guyannensis	PRO	B	XAI	-50.06717	-6.0548
MCNM1461	MCN-M	guyannensis	PRO	B	XAI	-49.72934	-6.43589
MCNM1490	MCN-M	guyannensis	PRO	B	XAI	-50.06717	-6.0548
MCNM1497	MCN-M	guyannensis	PRO	B	XAI	-49.72934	-6.43589
MCNM980	MCN-M	longicaudatus	PRO	D	NAR	-56.37939	-1.46522
MJ008	LMUSP	gardneri	PRO	A	JMI	-64.87035	-9.44602
MJ018	LMUSP	gardneri	PRO	A	JMI	-64.83393	-9.44944
MJ070	LMUSP	gardneri	PRO	A	JMI	-64.84313	-9.45063
MJ158	LMUSP	simonsi	PRO	E	WAM	-64.82581	-9.4455
MJ252	LMUSP	goeldii	PRO	A	PTI	-65.34884	-9.59557
MJ272	LMUSP	gardneri	PRO	A	JMI	-64.83465	-9.44823
MJ330	LMUSP	longicaudatus	PRO	D	UMR	-64.79292	-9.42976
MJ391	LMUSP	goeldii	PRO	A	PTI	-64.83327	-9.43604
MJ397	LMUSP	simonsi	PRO	E	WAM	-64.83294	-9.43781
MJ402	LMUSP	gardneri	PRO	A	JMI	-64.84323	-9.45019
MJ449	LMUSP	gardneri	PRO	A	JMI	-65.46861	-9.61306
MJ515	LMUSP	gardneri	PRO	A	JMI	-64.85128	-9.45414
MJ537	LMUSP	gardneri	PRO	A	JMI	-64.87124	-9.44407
MJ540	LMUSP	gardneri	PRO	A	JMI	-64.87124	-9.44407
MJ549	LMUSP	simonsi	PRO	E	WAM	-64.86	-9.45673
MJ601	LMUSP	goeldii	PRO	A	PTI	-65.0713	-9.56918
MJ636	LMUSP	simonsi	PRO	E	WAM	-64.85177	-9.4416
MJ709	LMUSP	gardneri	PRO	A	JMI	-64.83478	-9.4478
MJ774	LMUSP	simonsi	PRO	E	WAM	-64.83314	-9.43693
MN19620	MN	guyannensis	PRO	B	XAI	-48.43832	-1.41888
MN19622	MN	guyannensis	PRO	B	XAI	-48.43832	-1.41888
MN19625	MN	guyannensis	PRO	B	XAI	-48.43832	-1.41888
MN1974	MN	guyannensis	PRO	B	XAI	-48.49244	-1.45968
MN21910	MN	guyannensis	PRO	B	CER	-47.90605	-15.7862
MN21912	MN	guyannensis	PRO	B	CER	-47.90605	-15.7862
MN21918	MN	guyannensis	PRO	B	CER	-47.90605	-15.7862
MN21919	MN	guyannensis	PRO	B	CER	-47.90605	-15.7862
MN21920	MN	guyannensis	PRO	B	CER	-47.90605	-15.7862
MN22258	MN	guyannensis	PRO	B	CER	-47.90605	-15.7862
MN34460	MN	guyannensis	PRO	B	XAI	-48.43832	-1.41888

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
MN34462	MN	guyannensis	PRO	B	XAI	-48.43832	-1.41888
MN35875	MN	longicaudatus	PRO	D	NSM	-73.4774	-4.09507
MN35994	MN	guyannensis	PRO	B	CER	-48.30485	-13.83346
MN36021	MN	guyannensis	PRO	B	CER	-48.30485	-13.83346
MN36036	MN	guyannensis	PRO	B	CER	-48.30485	-13.83346
MN36049	MN	guyannensis	PRO	B	CER	-48.30485	-13.83346
MN36063	MN	guyannensis	PRO	B	CER	-48.30485	-13.83346
MN36064	MN	guyannensis	PRO	B	CER	-48.30485	-13.83346
MN36081	MN	guyannensis	PRO	B	CER	-48.30485	-13.83346
MN69017	MN	guyannensis	PRO	C	NRB	-63.51361	0.16639
MN69018	MN	guyannensis	PRO	C	NRB	-63.51361	0.16639
MN69025	MN	guyannensis	PRO	C	NRB	-63.51361	0.16639
MN69031	MN	guyannensis	PRO	C	NRB	-63.26194	0.16444
MN69044	MN	guyannensis	PRO	C	NRB	-63.26194	0.16444
MN69046	MN	guyannensis	PRO	C	NRB	-63.26194	0.16444
MN69059	MN	guyannensis	PRO	C	NRB	-63.26194	0.16444
MN69135	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN69136	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN69138	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN69144	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN69163	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69164	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69165	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69166	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69171	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69172	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69182	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69187	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69188	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69189	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69194	MN	guyannensis	PRO	C	NRB	-63.44567	0.74031
MN69242	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN69244	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN69263	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN69264	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
MN69344	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN69347	MN	guyannensis	PRO	C	NRB	-64.00778	0.3475
MN76211	MN	guyannensis	PRO	B	CER	-48.3	-5.28333
MN76215	MN	guyannensis	PRO	B	CER	-48.3	-5.28333
MN76741	MN	guyannensis	PRO	B	CER	-48.29032	-10.0782
MN76750	MN	guyannensis	PRO	B	CER	-48.29032	-10.0782
MN76754	MN	guyannensis	PRO	B	CER	-48.29032	-10.0782
MPEG20766	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20767	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20769	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20770	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20772	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20775	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20779	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20787	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20790	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20791	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20792	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20794	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20796	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20797	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG20798	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG21339	MPEG	longicaudatus	PRO	D	UMR	-61.93104	-10.84726
MPEG21983	MPEG	guyannensis	PRO	B	CER	-46.33483	-3.70759
MPEG25502	MPEG	gardneri	PRO	A	GJR	-70.85003	-6.75
MPEG25504	MPEG	gardneri	PRO	A	GJR	-70.85003	-6.75
MPEG25505	MPEG	gardneri	PRO	A	GJR	-70.85003	-6.75
MPEG25506	MPEG	gardneri	PRO	A	GJR	-70.85003	-6.75
MPEG25507	MPEG	gardneri	PRO	A	ATH	-72.78304	-8.66663
MPEG25508	MPEG	gardneri	PRO	A	ATH	-72.78304	-8.66663
MPEG25509	MPEG	gardneri	PRO	A	ATH	-72.78304	-8.66663
MPEG25511	MPEG	gardneri	PRO	A	ATH	-72.78304	-8.66663
MPEG25512	MPEG	gardneri	PRO	A	JMI	-68.9002	-6.58334
MPEG25515	MPEG	gardneri	PRO	A	JMI	-68.9002	-6.58334
MPEG25516	MPEG	gardneri	PRO	A	JMI	-68.9002	-6.58334

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
MPEG28342	MPEG	longicaudatus	PRO	D	SSM	-72.81667	-8.36666
MPEG28343	MPEG	longicaudatus	PRO	D	SSM	-72.81662	-8.36666
MPEG28344	MPEG	longicaudatus	PRO	D	SSM	-72.81662	-8.36666
MPEG28346	MPEG	longicaudatus	PRO	D	SSM	-72.85094	-8.60044
MPEG28347	MPEG	longicaudatus	PRO	D	SSM	-72.8	-8.56644
MPEG28349	MPEG	longicaudatus	PRO	D	SSM	-72.85094	-8.60044
MPEG28350	MPEG	longicaudatus	PRO	D	SSM	-72.78304	-8.66663
MPEG28351	MPEG	longicaudatus	PRO	D	SSM	-72.78304	-8.66663
MPEG28353	MPEG	longicaudatus	PRO	D	SSM	-72.78304	-8.66663
MPEG28354	MPEG	longicaudatus	PRO	D	SSM	-72.78304	-8.66663
MPEG28357	MPEG	longicaudatus	PRO	D	SSM	-72.81667	-8.36666
MPEG28358	MPEG	longicaudatus	PRO	D	SSM	-72.81662	-8.36666
MPEG28360	MPEG	longicaudatus	PRO	D	SSM	-72.81667	-8.36666
MPEG28361	MPEG	longicaudatus	PRO	D	SSM	-72.81662	-8.36666
MPEG28362	MPEG	longicaudatus	PRO	D	SSM	-72.81662	-8.36666
MPEG28363	MPEG	longicaudatus	PRO	D	SSM	-72.81662	-8.36666
MPEG28364	MPEG	longicaudatus	PRO	D	JUR	-68.76672	-6.46666
MPEG28368	MPEG	longicaudatus	PRO	D	JUR	-70.75008	-6.83344
MPEG28369	MPEG	goeldii	PRO	A	CLJ	-70.75008	-6.83344
MPEG28372	MPEG	longicaudatus	PRO	D	JUR	-70.75008	-6.83344
MPEG28373	MPEG	longicaudatus	PRO	D	JUR	-70.75008	-6.83344
MPEG28374	MPEG	longicaudatus	PRO	D	JUR	-70.75008	-6.83344
MPEG28377	MPEG	longicaudatus	PRO	D	JUR	-70.85003	-6.75
MPEG28397	MPEG	simonsi	PRO	E	WAM	-68.76672	-6.46666
MPEG28398	MPEG	simonsi	PRO	E	WAM	-68.89219	-6.58282
MPEG28400	MPEG	simonsi	PRO	E	WAM	-72.81662	-8.36666
MPEG28408	MPEG	simonsi	PRO	E	WAM	-72.81662	-8.36666
MPEG28409	MPEG	simonsi	PRO	E	WAM	-72.81662	-8.36666
MPEG28414	MPEG	simonsi	PRO	E	WAM	-66.01666	-3.31666
MPEG28416	MPEG	simonsi	PRO	E	WAM	-66.01666	-3.31666
MPEG28417	MPEG	simonsi	PRO	E	WAM	-66.01666	-3.31666
MPEG28419	MPEG	simonsi	PRO	E	WAM	-66.01666	-3.31666
MPEG28424	MPEG	simonsi	PRO	E	WAM	-70.85003	-6.75
MPEG28427	MPEG	simonsi	PRO	E	WAM	-68.76672	-6.46666
MPEG28429	MPEG	simonsi	PRO	E	WAM	-68.89219	-6.58282

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
MPEG28430	MPEG	simonsi	PRO	E	WAM	-68.89219	-6.58282
MPEG28432	MPEG	simonsi	PRO	E	WAM	-68.89219	-6.58282
MPEG28433	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28437	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28438	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28443	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28444	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28447	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28448	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28449	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28451	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28452	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28453	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28455	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28456	MPEG	simonsi	PRO	E	WAM	-72.78304	-8.66663
MPEG28459	MPEG	simonsi	PRO	E	WAM	-72.81662	-8.36666
MPEG28460	MPEG	longicaudatus	PRO	D	SSM	-72.81662	-8.36666
MPEG28486	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28491	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28492	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28493	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28529	MPEG	goeldii	PRO	A	CLJ	-70.75008	-6.83344
MPEG28531	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28532	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28533	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28534	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28535	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28537	MPEG	goeldii	PRO	A	CLJ	-70.73359	-6.80001
MPEG28540	MPEG	goeldii	PRO	A	CLJ	-70.75008	-6.83344
MPEG28543	MPEG	goeldii	PRO	A	CLJ	-70.85003	-6.75
MPEG28546	MPEG	goeldii	PRO	A	CLJ	-70.85003	-6.75
MPEG28558	MPEG	goeldii	PRO	A	CLJ	-70.85003	-6.75
MPEG28559	MPEG	goeldii	PRO	A	CLJ	-68.76667	-6.46669
MPEG28562	MPEG	goeldii	PRO	A	CLJ	-68.76667	-6.46669
MPEG28564	MPEG	goeldii	PRO	A	CLJ	-68.76667	-6.46669

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
MPEG28567	MPEG	goeldii	PRO	A	CLJ	-68.91651	-6.53356
MPEG28569	MPEG	goeldii	PRO	A	UJR	-72.85094	-8.60044
MPEG28570	MPEG	goeldii	PRO	A	UJR	-72.85094	-8.60044
MPEG28571	MPEG	goeldii	PRO	A	UJR	-72.81662	-8.36666
MPEG28572	MPEG	goeldii	PRO	A	UJR	-72.81662	-8.36666
MPEG28573	MPEG	goeldii	PRO	A	UJR	-72.81667	-8.36666
MPEG28574	MPEG	goeldii	PRO	A	UJR	-72.81667	-8.36666
MPEG28575	MPEG	goeldii	PRO	A	UJR	-72.81667	-8.36666
MPEG28576	MPEG	goeldii	PRO	A	UJR	-72.81662	-8.36666
MPEG33879	MPEG	goeldii	PRO	A	EAM	-51.45549	-1.7374
MPEG33880	MPEG	guyannensis	PRO	B	XAI	-51.45549	-1.7374
MPEG33881	MPEG	goeldii	PRO	A	EAM	-51.45549	-1.7374
MPEG33882	MPEG	guyannensis	PRO	B	XAI	-51.45549	-1.7374
MPEG33997	MPEG	guyannensis	PRO	B	XAI	-51.45549	-1.7374
MPEG42408	MPEG	longicaudatus	PRO	D	NAR	-56.36422	-1.79591
MPEG42431	MPEG	longicaudatus	PRO	D	NAR	-56.36422	-1.79591
MUSM11258	AMNH	longicaudatus	PRO	D	SSM	-73.16208	-5.2495
MUSM11262	AMNH	longicaudatus	PRO	D	SSM	-73.16208	-5.2495
MUSM11283	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
MUSM11297	AMNH	longicaudatus	PRO	D	SSM	-73.16208	-5.2495
MUSM11299	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
MUSM11300	AMNH	gardneri	PRO	A	GJR	-73.16208	-5.2495
MUSM11314	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
MUSM13342	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
MUSM13343	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
MUSM13344	AMNH	simonsi	PRO	E	WAM	-73.16208	-5.2495
MVZ153616	MVZ	longicaudatus	PRO	D	NSM	-78.1572	-4.45772
MVZ155034	MVZ	longicaudatus	PRO	D	NSM	-78.16854	-4.45236
MVZ155036	MVZ	longicaudatus	PRO	D	NSM	-78.16123	-4.45563
MVZ157855	MVZ	longicaudatus	PRO	D	SPB	-69068	-12.63333
MVZ157934	MVZ	longicaudatus	PRO	D	NSM	-77751	-4022
MVZ157948	MVZ	longicaudatus	PRO	D	NSM	-77751	-4022
MVZ157966	MVZ	longicaudatus	PRO	D	NSM	-77751	-4022
MVZ160093	MVZ	longicaudatus	PRO	D	NAR	-61.43333	6.15
MVZ166814	MVZ	simonsi	PRO	E	WAM	-71.26166	-12.68001

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
MVZ168942	MVZ	goeldii	PRO	A	MMR	-69.07289	-12.6
MVZ168943	MVZ	goeldii	PRO	A	MMR	-69.07289	-12.6
MVZ168947	MVZ	trinitatis	PRO	B	NVZ	-72.16666	10
MVZ168955	MVZ	simonsi	PRO	E	WAM	-69.07289	-12.6
MVZ168956	MVZ	longicaudatus	PRO	D	SPB	-69.07289	-12.6
MVZ168958	MVZ	longicaudatus	PRO	D	SPB	-69.07289	-12.6
MVZ187184	MVZ	gardneri	PRO	A	GJR	-70.85003	-6.75
MVZ187185	MVZ	gardneri	PRO	A	GJR	-70.85003	-6.75
MVZ187186	MVZ	gardneri	PRO	A	GJR	-70.85003	-6.75
MVZ187187	MVZ	gardneri	PRO	A	GJR	-70.85003	-6.75
MVZ187188	MVZ	gardneri	PRO	A	GJR	-70.85003	-6.75
MVZ187191	MVZ	gardneri	PRO	A	GJR	-70.85003	-6.75
MVZ187193	MVZ	gardneri	PRO	A	GJR	-68.76672	-6.46666
MVZ187194	MVZ	gardneri	PRO	A	ATH	-72.78304	-8.66663
MVZ187195	MVZ	gardneri	PRO	A	ATH	-72.78304	-8.66663
MVZ187203	MVZ	gardneri	PRO	A	JMI	-68.89219	-6.58282
MVZ187204	MVZ	gardneri	PRO	A	JMI	-68.89219	-6.58282
MVZ187205	MVZ	gardneri	PRO	A	JMI	-68.89219	-6.58282
MVZ187206	MVZ	gardneri	PRO	A	JMI	-68.89219	-6.58282
MVZ187207	MVZ	gardneri	PRO	A	JMI	-68.89219	-6.58282
MVZ190668	MVZ	longicaudatus	PRO	D	SSM	-72.78304	-8.66663
MVZ190678	MVZ	longicaudatus	PRO	D	SSM	-72.81667	-8.36666
MVZ190684	MVZ	longicaudatus	PRO	D	JUR	-70.75008	-6.83344
MVZ190689	MVZ	longicaudatus	PRO	D	JUR	-70.75008	-6.83344
MVZ190692	MVZ	longicaudatus	PRO	D	JUR	-70.85003	-6.75
MVZ190693	MVZ	longicaudatus	PRO	D	JUR	-70.85003	-6.75
MVZ190696	MVZ	longicaudatus	PRO	D	JUR	-68.76672	-6.46666
MVZ197574	MVZ	longicaudatus	PRO	D	SCE	-52.35583	-15.63333
MVZ197576	MVZ	guyannensis	PRO	B	TXI	-55.93028	-9.59694
MVZ197578	MVZ	guyannensis	PRO	B	TXI	-55.93028	-9.59694
MVZ197579	MVZ	guyannensis	PRO	B	TXI	-55.93028	-9.59694
MVZ197580	MVZ	guyannensis	PRO	B	TXI	-55.93028	-9.59694
MVZ197581	MVZ	guyannensis	PRO	B	TXI	-55.93028	-9.59694
MVZ197582	MVZ	guyannensis	PRO	B	TXI	-55.93028	-9.59694
MVZ197585	MVZ	guyannensis	PRO	B	CER	-48.63556	-11.84278

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
MZUSP21255	MZUSP	longicaudatus	PRO	D	SAR	-52.3666	-3.65
MZUSP21257	MZUSP	longicaudatus	PRO	D	SAR	-52.3666	-3.65
MZUSP21258	MZUSP	longicaudatus	PRO	D	SAR	-52.3666	-3.65
MZUSP21260	MZUSP	longicaudatus	PRO	D	SAR	-52.3666	-3.65
MZUSP21276	MZUSP	longicaudatus	PRO	D	SAR	-52.3666	-3.65
MZUSP21278	MZUSP	longicaudatus	PRO	D	SAR	-52.3666	-3.65
MZUSP26694	MZUSP	guyannensis	PRO	B	XAI	-48.49036	-1.45555
MZUSP26696	MZUSP	guyannensis	PRO	B	XAI	-48.49036	-1.45555
MZUSP30365	MZUSP	guyannensis	PRO	B	CER	-45.20264	-8.86342
MZUSP30368	MZUSP	guyannensis	PRO	B	CER	-45.20264	-8.86342
MZUSP30370	MZUSP	guyannensis	PRO	B	CER	-45.20264	-8.86342
MZUSP30376	MZUSP	guyannensis	PRO	B	CER	-45.20264	-8.86342
MZUSP30377	MZUSP	guyannensis	PRO	B	CER	-45.20264	-8.86342
MZUSP30379	MZUSP	guyannensis	PRO	B	CER	-45.20264	-8.86342
MZUSP30382	MZUSP	guyannensis	PRO	B	CER	-45.20264	-8.86342
MZUSP30384	MZUSP	guyannensis	PRO	B	CER	-45.20264	-8.86342
MZUSP31924	MZUSP	goeldii	PRO	A	PTI	-58.49231	-10.32276
MZUSP31925	MZUSP	longicaudatus	PRO	D	MTR	-59.44718	-10.17484
MZUSP31928	MZUSP	longicaudatus	PRO	D	MTR	-59.44718	-10.17484
MZUSP31937	MZUSP	longicaudatus	PRO	D	MTR	-58.49231	-10.32276
MZUSP31939	MZUSP	longicaudatus	PRO	D	MTR	-58.49231	-10.32276
MZUSP31945	MZUSP	guyannensis	PRO	B	TXI	-53.25595	-13.18229
MZUSP31946	MZUSP	guyannensis	PRO	B	TXI	-53.25595	-13.18229
MZUSP31950	MZUSP	guyannensis	PRO	B	CER	-48.54	-12.02688
MZUSP31951	MZUSP	guyannensis	PRO	B	CER	-48.54	-12.02688
PNPA321	MZUSP	longicaudatus	PRO	D	PTC	-57.16668	-19.63333
PNPA357	MZUSP	longicaudatus	PRO	D	PTC	-57.16668	-19.63333
RGM044	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767
RGM156	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767
RGM529	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767
RGM530	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767
RGM531	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767
RGM848	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
RGM855	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767
RGM857	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767
RGM883	MZUSP	goeldii	PRO	A	PTI	-58.97905	-15.13767
UFES1388	UFES-MAM	guyannensis	PRO	B	CER	-49.95863	-9.30361
UFES1390	UFES-MAM	guyannensis	PRO	B	CER	-49.95863	-9.30361
UFES1392	UFES-MAM	guyannensis	PRO	B	CER	-49.95863	-9.30361
UFES1395	UFES-MAM	guyannensis	PRO	B	CER	-49.9755	-9.3843
UFES1396	UFES-MAM	guyannensis	PRO	B	CER	-49.95863	-9.30361
UFES1397	UFES-MAM	guyannensis	PRO	B	CER	-49.95863	-9.30361
UFES1400	UFES-MAM	guyannensis	PRO	B	CER	-49.9755	-9.3843
UFES1402	UFES-MAM	guyannensis	PRO	B	CER	-49.9755	-9.3843
UFES1403	UFES-MAM	guyannensis	PRO	B	CER	-49.9755	-9.3843
UFES1569	UFES-MAM	guyannensis	PRO	B	XAI	-50.44612	-6.34786
UFES1570	UFES-MAM	guyannensis	PRO	B	XAI	-50.36045	-6.38747
UFES1580	UFES-MAM	guyannensis	PRO	B	XAI	-50.42313	-6.34376
UFES1583	UFES-MAM	guyannensis	PRO	B	XAI	-50.40877	-6.34475
UFES1857	UFES-MAM	guyannensis	PRO	B	XAI	-50.42313	-6.34376
UFES1858	UFES-MAM	guyannensis	PRO	B	XAI	-50.42313	-6.34376
UFES1859	UFES-MAM	guyannensis	PRO	B	XAI	-50.42313	-6.34376
UFES1860	UFES-MAM	guyannensis	PRO	B	XAI	-50.42313	-6.34376
UFES1873	UFES-MAM	guyannensis	PRO	B	XAI	-50.43528	-6.40018
UFES1887	UFES-MAM	guyannensis	PRO	B	XAI	-50.12758	-6.32946
UFES2637	UFES-MAM	goeldii	PRO	A	PTI	-59.1306	-3.57972
UFES2638	UFES-MAM	goeldii	PRO	A	PTI	-59.1306	-3.57972
UFES2640	UFES-MAM	goeldii	PRO	A	PTI	-59.1306	-3.57972
UFES2649	UFES-MAM	gardneri	PRO	A	LMR	-59.1306	-3.57972
UFES2834	UFES-MAM	gardneri	PRO	A	LMR	-59.1306	-3.57972
UFES2835	UFES-MAM	gardneri	PRO	A	LMR	-59.1306	-3.57972
UFES2839	UFES-MAM	gardneri	PRO	A	LMR	-59.1306	-3.57972
UFES2962	UFES-MAM	goeldii	PRO	A	EAM	-51.77012	-3.12428
UFMG3018	UFMG	longicaudatus	PRO	D	SCE	-52.35583	-15.63333
UFMG3023	UFMG	guyannensis	PRO	B	TXI	-55.93028	-9.59694
UFMG3026	UFMG	guyannensis	PRO	B	TXI	-55.93028	-9.59694
UFMG3029	UFMG	guyannensis	PRO	B	TXI	-55.93028	-9.59694
UFMG3032	UFMG	guyannensis	PRO	B	TXI	-55.93028	-9.59694

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
UFMG3033	UFMG	guyannensis	PRO	B	TXI	-55.93028	-9.59694
UFMG3035	UFMG	guyannensis	PRO	B	CER	-48.63556	-11.84278
UFPB6929	UFPB	guyannensis	PRO	B	CER	-44.41654	-2.33712
UFPB6931	UFPB	guyannensis	PRO	B	CER	-44.41654	-2.33712
UFPB6932	UFPB	guyannensis	PRO	B	CER	-44.41654	-2.33712
UFSM305	LMUSP	longicaudatus	PRO	D	SCE	-52.1493	-15.88295
UMMZ125182	UMMZ	outgroup	HG	-	-	-84.02	10.4244
UMMZ80069	UMMZ	goeldii	PRO	A	IJM	-77817	-983
UMMZ80073	UMMZ	goeldii	PRO	A	IJM	-77817	-983
USNM102731	NMNH	trinitatis	PRO	B	NVZ	-66.92276	10.59241
USNM172949	NMNH	longicaudatus	PRO	D	NAR	-58.27163	6.4857
USNM364151	NMNH	simonsi	PRO	E	WAM	-69.23741	-12.59659
USNM388133	NMNH	trinitatis	TEP	-	-	-61.47003	5.95496
USNM390367	NMNH	simonsi	PRO	E	WAM	-69.18341	-12.58804
USNM391009	NMNH	longicaudatus	PRO	D	JUR	-63.17175	-17.51258
USNM392887	NMNH	semispinosus	PRO	D	CAM	-83.23064	14.98815
USNM442697	NMNH	trinitatis	PRO	B	NVZ	-72.63333	9.20002
USNM442699	NMNH	trinitatis	PRO	B	NVZ	-72.63333	9.20002
USNM442880	NMNH	trinitatis	PRO	B	NVZ	-72.63333	9.20002
USNM442883	NMNH	trinitatis	PRO	B	NVZ	-72.63333	9.20002
USNM513633	NMNH	decumanus	PRO	D	ECU	-80.06913	-2.24016
USNM513645	NMNH	decumanus	PRO	D	ECU	-79.93627	-3.48507
USNM530931	NMNH	goeldii	PRO	A	MMR	-71.0451	-12.15612
USNM530932	NMNH	goeldii	PRO	A	MMR	-71.0451	-12.15612
USNM530935	NMNH	simonsi	PRO	E	WAM	-69.20697	-12.72047
USNM559415	NMNH	goeldii	PRO	A	MMR	-71.11199	-12.09053
USNM559426	NMNH	goeldii	PRO	A	MMR	-71.11199	-12.09053
USNM578000	NMNH	simonsi	PRO	E	WAM	-72.94991	-11.58344
USNM579259	NMNH	simonsi	PRO	E	WAM	-68.77	-13.58
USNM579616	NMNH	simonsi	PRO	E	WAM	-66.7333	-10.7667
USNM579694	NMNH	simonsi	PRO	E	WAM	-69.6122	-13.1472
USNM579695	NMNH	simonsi	PRO	E	WAM	-69.6122	-13.1472
USNM579697	NMNH	simonsi	PRO	E	WAM	-69.6844	-13.5044
USNM581983	NMNH	longicaudatus	PRO	D	SCE	-61.0083	-13.5517
USNM582772	NMNH	simonsi	PRO	E	WAM	-72.94991	-11.58344

Table S2: Continuation.

Catalog Number	Source	Species Group	Analysis 1	Analysis 2	Analysis 3	Longitude	Latitude
USNM582897	NMNH	simonsi	PRO	E	WAM	-72.94991	-11.58344
USNM584593	NMNH	longicaudatus	PRO	D	SCE	-61.0347	-14.7672
USNM584594	NMNH	longicaudatus	PRO	D	SCE	-61.0347	-14.7672
USNM584595	NMNH	longicaudatus	PRO	D	SCE	-61.0347	-14.7672
USNM584596	NMNH	longicaudatus	PRO	D	SCE	-61.0347	-14.7672
USNM584597	NMNH	longicaudatus	PRO	D	SCE	-61.0347	-14.7672
USNM584598	NMNH	longicaudatus	PRO	D	SCE	-61.0347	-14.7672
USNM588195	NMNH	longicaudatus	PRO	D	SCE	-61.0347	-14.7672
X1M15	LMUSP	longicaudatus	PRO	D	SAR	-49.94611	-4.25168

Table S3: RAD processing results of the three genomic libraries, showing in the initial number and post-filters of reads, clusters, values of heterozygosity (H) and its error (HE), the final number of clusters after correction by heterozygosity, the number of loci retained and samples used for the conservative trees (CT) analysis, and information on the samples that were used in strict analyses (ST). Of the 278 initial samples (270 *Proechimys* + 8 Outgroups), 235 were sequenced, the others had little amount of extracted or digested DNA and they needed to be removed from the libraries. Some samples were excluded from the phyloegenetic analyses because they had less than 100,000 filtered reads or less than 50 retained loci. In addition, there is information about the five main clades and 41 subclasses where the sample was recovered in the conservative tree. Table in ***.csv*** format available on https://github.com/jdalapicolla.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
756	A	PTI	876,183	875,176	299,176	10,472	0.0115	0.0053	9,076	1,776	YES	-
ABX005	B	ABX	4,250,491	4,246,405	365,088	79,939	0.0081	0.0018	76,583	14,464	YES	YES
ABX008	B	ABX	762,982	762,306	88,703	30,486	0.0062	0.0017	29,701	5,191	YES	-
ABX020	B	ABX	2,723,645	2,721,069	310,633	74,213	0.007	0.0018	71,727	14,005	YES	YES
ABX027	A	PTI	2,018,660	2,016,780	240,582	63,552	0.0073	0.002	61,267	12,697	YES	-
ABX028	B	ABX	2,362,504	2,360,317	239,439	69,703	0.0066	0.0018	67,567	13,142	YES	-
ABX029	B	ABX	2,372,097	2,370,015	255,708	69,657	0.0058	0.0017	67,708	13,070	YES	-
ABX077	-	-	7,162,966	7,155,903	755,307	136,625	0.0096	0.0025	128,073	23	-	-
AMCC112929	B	NVZ	851,170	850,361	159,706	42,923	0.0046	0.0015	41,655	3,405	YES	-
AMCC112987	B	NVZ	3,039,552	3,036,575	387,817	70,208	0.006	0.0016	67,500	5,984	YES	YES
AMCC112999	C	SVZ	1,493,735	1,492,180	188,042	47,498	0.0076	0.0014	45,881	6,433	YES	YES
AMCC114577	A	GJR	2,609,974	2,607,328	367,853	69,937	0.011	0.0017	66,700	6,416	YES	YES
AMCC114589	E	WAM	1,151,708	1,150,426	123,083	6,592	0.0178	0.0011	5,823	579	YES	-

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
AMCC175984	B	NVZ	1,593,994	1,592,345	253,049	57,651	0.007	0.0015	55,700	4,703	YES	-
AMCC176080	B	NVZ	2,627,971	2,625,323	420,811	70,249	0.0056	0.0016	67,277	5,787	YES	YES
AMNH23109	-	-	134	103	98	0	0.01	0.001	0	0	-	-
AMNH235152	-	-	17,108	16,920	13,642	87	0.014	0.0041	72	1	-	-
AMNH269122	D	NAR	446,760	446,305	168,094	24,787	0.0065	0.0042	22,730	3,877	YES	-
AMNH269123	D	NAR	4,709,606	4,704,845	440,479	81,567	0.0092	0.0015	77,209	12,105	YES	YES
AMNH272698	D	SSM	1,100,686	1,099,620	237,619	55,802	0.0075	0.0029	53,195	8,183	YES	-
AMNH272700	D	SSM	1,608,196	1,606,633	316,822	62,693	0.0078	0.0027	59,824	9,042	YES	YES
AMNH272714	A	GJR	1,646,703	1,644,970	242,247	61,240	0.0104	0.0016	58,710	5,566	YES	-
APC1085	D	SCE	1,275,032	1,273,833	195,288	61,834	0.0053	0.0019	59,832	11,950	YES	-
APC1217	B	CER	1,223,385	1,222,263	196,789	60,278	0.0044	0.0021	58,622	16,177	YES	-
APC817	B	CER	507,218	506,705	136,624	32,196	0.0039	0.0019	31,162	8,333	YES	-
APC839	B	CER	954,670	953,833	138,773	46,630	0.0042	0.002	45,412	12,212	YES	-
BAC320	B	XAI	1,307,660	1,306,432	215,425	60,063	0.0065	0.0019	58,348	14,849	YES	-
BM12174	A	EAM	654,113	653,474	196,741	41,609	0.0082	0.0024	39,770	5,890	YES	-
C247647	A	PTI	3,778,844	3,775,227	378,714	85,432	0.0077	0.002	81,855	16,712	YES	YES
CAM200	-	-	298,222	297,866	74,669	473	0.0355	0.0158	263	30	-	-
CIT375	D	MTR	1,141,773	1,140,616	262,513	58,679	0.0066	0.002	56,502	9,753	YES	-
CIT393	D	MTR	2,300,707	2,298,561	374,711	80,194	0.0073	0.0019	77,089	13,523	YES	YES

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
CIT405	D	MTR	1,264,816	1,263,620	183,971	57,586	0.0049	0.0018	55,895	9,538	YES	-
CIT448	D	MTR	2,766,677	2,764,014	314,952	72,701	0.0057	0.0018	70,109	12,523	YES	YES
CIT588	-	-	70,010	69,934	50,881	306	0.0198	0.0069	229	22	-	-
CIT648	A	PTI	2,895,615	2,892,827	260,567	72,153	0.0077	0.002	69,315	14,378	YES	YES
CTA1028	OUT	OUT	3,141,959	3,138,855	425,166	73,175	0.007	0.0019	70,068	678	YES	-
CTA1349	D	SCE	580,562	580,005	199,175	33,686	0.0044	0.0021	32,287	5,795	YES	-
CTA1415	D	SCE	2,175,679	2,173,754	259,666	71,385	0.005	0.0019	69,166	13,739	YES	YES
CTA1511	B	TXI	2,672,598	2,669,989	311,918	78,351	0.0068	0.002	75,906	16,978	YES	YES
CTA1517	A	EAM	2,568,226	2,565,803	267,401	77,763	0.009	0.0019	74,971	12,610	YES	YES
CTA1652	B	CER	2,780,902	2,778,288	289,468	75,105	0.005	0.0018	72,984	20,180	YES	-
CTA1835	B	CER	382,986	382,578	108,012	25,465	0.0046	0.0021	24,602	6,183	YES	-
CTA4195	C	GUS	2,021,054	2,018,923	304,134	63,369	0.0066	0.0015	61,033	9,195	YES	YES
CTA4226	C	GUS	1,355,311	1,354,063	227,575	56,566	0.0057	0.0027	54,486	8,126	YES	-
CTA4245	D	NAR	2,214,988	2,212,738	215,530	53,681	0.0073	0.0016	51,529	8,307	YES	YES
CTA4324	A	PTI	2,049,704	2,047,800	223,033	58,921	0.0067	0.0022	56,777	11,679	YES	-
CTA4325	A	PTI	1,481,203	1,479,790	210,688	49,174	0.0067	0.0023	47,176	9,408	YES	-
CTA4326	A	PTI	1,361,471	1,360,246	162,081	52,563	0.0067	0.002	50,760	10,250	YES	-
CTA4327	A	PTI	2,508,069	2,505,756	272,538	67,742	0.0071	0.002	65,265	13,620	YES	-
CTA4352	A	LMR	3,599,480	3,595,923	402,105	75,248	0.0087	0.0015	71,955	7,219	YES	YES

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
CTA4357	A	PTI	1,383,550	1,382,312	163,220	59,376	0.0066	0.002	57,491	11,573	YES	-
CTA4363	A	LMR	2,440,084	2,437,512	323,691	66,449	0.0086	0.0018	63,515	6,295	YES	YES
CTA4371	A	IJM	1,124,626	1,123,499	213,095	52,486	0.0074	0.0017	50,455	6,500	YES	YES
CTA4390	B	TXI	198,766	198,558	124,352	3,750	0.0126	0.0044	3,320	554	YES	-
CTA4400	A	EAM	737,976	737,170	185,395	42,822	0.0082	0.0022	41,033	6,014	YES	-
DPO018	B	XAI	1,203,365	1,202,146	173,078	51,983	0.0067	0.002	50,494	12,678	YES	-
EEB1013	OUT	OUT	1,245,322	1,244,143	201,547	55,354	0.006	0.0018	53,616	544	YES	-
EFA004	A	PTI	938,298	937,485	184,554	49,867	0.0064	0.003	47,632	8,250	YES	-
EFA015	A	PTI	613,347	612,818	146,081	34,908	0.0064	0.0031	33,082	5,765	YES	-
EFA037	E	WAM	639,589	639,065	150,881	41,869	0.006	0.0029	40,036	8,627	YES	-
FMNH175255	D	UMD	408,058	407,613	65,103	17,918	0.0062	0.0015	17,330	2,175	YES	-
FMNH175275	E	WAM	3,271,144	3,268,074	456,885	78,046	0.0073	0.0029	74,229	13,989	YES	YES
FMNH26441	-	-	28,994	28,855	4,524	304	0.0074	0.0016	287	2	-	-
FMNH52618	-	-	31,829	31,752	25,249	64	0.0319	0.0143	37	2	-	-
FMPS010	OUT	OUT	3,561,188	3,557,750	377,580	71,386	0.0093	0.0021	66,980	629	YES	YES
ICA085	-	-	53,220	53,147	13,790	3,480	0.005	0.0021	3,370	250	-	-
ICA095	A	IJM	2,819,142	2,816,301	399,671	75,502	0.0101	0.0018	71,378	9,231	YES	YES

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
ICA240	A	NSR	1,175,331	1,174,206	263,133	50,911	0.0066	0.003	48,428	3,514	YES	-
ICA245	A	NSR	2,414,180	2,411,946	390,167	67,468	0.0082	0.0029	63,649	4,985	YES	YES
ICA246	A	IJM	275,704	275,390	90,728	17,505	0.0084	0.0023	16,493	2,042	YES	-
JAP006	A	IJM	1,001,447	1,000,454	193,933	52,284	0.0079	0.0018	50,243	6,437	YES	-
JAP012	A	NSR	1,507,100	1,505,662	240,879	57,940	0.0069	0.0017	55,696	4,376	YES	YES
JAP094	D	JAP	1,350,168	1,348,909	290,864	58,209	0.0056	0.0029	55,642	8,841	YES	YES
JAP095	D	JAP	769,517	768,890	204,641	46,628	0.0047	0.003	44,565	7,062	YES	-
JAP098	D	JAP	1,189,912	1,188,835	227,822	55,642	0.0052	0.0027	53,485	8,583	YES	-
JAP107	D	JAP	1,476,001	1,474,595	352,015	61,519	0.0058	0.0029	58,300	8,854	YES	YES
JUF017	C	NRB	4,061,708	4,057,600	480,217	90,502	0.0086	0.0015	85,990	12,192	YES	YES
MBR046	OUT	OUT	1,674,854	1,673,033	238,582	58,226	0.0061	0.0018	55,953	1,101	YES	-
MCNM1341	B	CER	295,266	294,991	73,954	21,701	0.0041	0.002	21,060	5,071	YES	-
MCNM1497	B	XAI	2,400,373	2,398,159	291,650	72,106	0.007	0.0019	69,765	18,141	YES	YES
MCNM1989	B	CER	181,793	181,559	125,972	2,598	0.0091	0.0059	2,272	536	YES	-
MCNM2034	B	CER	228,003	227,725	127,512	6,616	0.0062	0.0044	6,067	1,476	YES	-
MCNM2258	D	SAR	185,154	184,762	108,418	3,090	0.011	0.0058	2,689	501	YES	-
MJ12	A	JMI	580,625	579,987	129,822	34,993	0.0081	0.0021	33,580	3,439	YES	-
MJ252	A	PTI	2,092,078	2,090,078	251,012	68,095	0.0071	0.0017	65,781	13,499	YES	-
MJ330	D	UMR	424,341	424,007	117,955	32,302	0.0065	0.0019	31,033	4,646	YES	-

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
MJ391	A	PTI	1,981,924	1,980,001	212,667	67,813	0.0072	0.0019	65,465	13,447	YES	-
MJ449	A	JMI	860,798	859,951	187,551	48,334	0.0082	0.003	46,090	4,741	YES	-
MJ514	A	JMI	1,252,271	1,251,100	260,847	55,979	0.0085	0.0029	53,392	5,527	YES	-
MJ515	A	JMI	2,144,733	2,142,758	319,340	65,331	0.0086	0.0029	62,354	6,504	YES	YES
MJ523	A	JMI	733,361	732,759	148,057	46,451	0.008	0.002	44,721	4,665	YES	-
MJ529	E	WAM	1,373,019	1,371,819	278,694	58,446	0.0071	0.0027	56,239	11,676	YES	-
MJ6	A	JMI	538,303	537,813	159,877	34,824	0.0079	0.0035	32,522	3,531	YES	-
MJ601	A	PTI	2,064,457	2,062,519	390,807	68,453	0.0098	0.0034	64,044	11,466	YES	-
MJ665	A	JMI	577,034	576,582	177,430	35,784	0.0078	0.0033	33,674	3,420	YES	-
MN36222	B	CER	1,565,955	1,564,411	263,020	64,175	0.0048	0.0018	62,330	17,218	YES	-
MN36702	B	CER	212,056	211,857	55,182	13,804	0.0046	0.0021	13,380	3,265	YES	-
MN56812	C	NRB	3,830,255	3,826,355	514,924	82,841	0.0094	0.0017	78,403	10,995	YES	YES
MN56815	C	NRB	1,976,755	1,974,759	318,122	65,649	0.0078	0.0017	63,155	9,494	YES	-
MN56816	C	NRB	2,233,523	2,231,269	338,417	66,922	0.0079	0.0016	64,204	9,601	YES	-
MN67246	D	SCE	564,796	564,243	153,577	38,291	0.004	0.002	36,961	7,043	YES	-
MN76204	B	CER	1,399,039	1,397,706	148,027	44,948	0.0048	0.0018	43,714	11,758	YES	-
MN76211	B	CER	3,103,532	3,100,828	299,906	74,991	0.0056	0.0019	72,637	20,146	YES	-
MN76750	B	CER	3,399,309	3,396,093	309,899	71,700	0.006	0.0018	69,258	18,669	YES	YES
MN76754	B	CER	2,233,270	2,231,186	235,670	63,209	0.0056	0.0018	61,362	16,814	YES	-

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
MPEG20767	D	UMR	1,219,017	1,218,000	217,180	55,313	0.0073	0.0018	53,368	8,137	YES	-
MPEG20768	D	UMR	3,596,375	3,592,865	443,107	79,054	0.0093	0.0016	75,153	11,352	YES	YES
MPEG20769	D	UMR	3,871,135	3,867,308	471,710	81,980	0.0099	0.0016	77,608	11,246	YES	YES
MPEG21983	B	CER	390,555	390,116	113,193	20,755	0.0063	0.0026	19,941	4,438	YES	-
MPEG40369	A	EAM	566,495	565,899	178,251	37,010	0.0063	0.0027	35,372	5,415	YES	-
MPEG40371	B	XAI	374,503	374,070	147,263	21,565	0.0072	0.0034	20,428	5,064	YES	-
MRT3925	B	CER	267,355	267,105	77,981	18,972	0.0043	0.002	18,276	4,384	YES	-
MSB140110	D	ECU	3,710,546	3,706,843	421,001	75,722	0.0071	0.0015	72,397	7,666	YES	YES
MSB140111	D	ECU	2,771,091	2,768,271	354,251	67,979	0.0067	0.0016	65,207	6,793	YES	YES
MSB208394	D	PTC	1,681,886	1,680,194	233,896	59,112	0.0062	0.0015	57,091	10,353	YES	YES
MSB210840	D	PTC	1,489,934	1,488,429	257,721	56,443	0.0063	0.0016	54,429	9,886	YES	-
MSB211792	D	PAB	999,198	998,289	158,650	51,920	0.0049	0.0027	50,058	7,585	YES	YES
MSB236570	A	ATH	2,558,413	2,556,048	382,635	70,321	0.0097	0.0028	66,827	6,524	YES	YES
MSB236594	E	WAM	496,118	495,673	140,709	35,769	0.0067	0.003	33,919	7,320	YES	-
MSB236689	A	MMR	806,223	805,423	153,452	48,627	0.0066	0.0028	46,589	7,611	YES	YES
MSB236698	A	MMR	2,605,256	2,602,807	345,824	70,480	0.0082	0.0029	67,312	11,175	YES	-
MSB236807	A	MMR	1,131,093	1,130,084	247,090	53,394	0.0069	0.003	50,907	8,310	YES	YES
MSB238391	D	SPB	1,421,451	1,420,325	271,360	55,518	0.006	0.0018	53,567	7,691	YES	-
MSB239628	D	PTC	1,285,043	1,283,948	238,296	56,120	0.0048	0.0018	54,262	9,812	YES	-

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
MSB263513	HG	HG	1,774,930	1,773,026	372,885	65,644	0.0071	0.0022	62,329	3,793	YES	YES
MSB45836	D	SEM	1,495,950	1,494,764	239,216	54,149	0.0038	0.0017	52,519	5,380	YES	YES
MSB70574	D	SPB	3,163,614	3,160,405	411,678	75,158	0.0076	0.0016	71,696	10,258	YES	YES
MSB70575	D	SPB	1,509,094	1,507,733	249,987	59,055	0.0061	0.0028	56,797	8,020	YES	-
MSB99059	D	PTC	1,574,661	1,573,182	321,540	59,536	0.0057	0.0027	56,960	10,301	YES	YES
MSB99060	D	PTC	1,084,800	1,083,721	243,208	48,781	0.0053	0.0029	46,671	8,442	YES	-
MUSM13338	D	SSM	601,726	601,176	149,745	40,357	0.007	0.003	38,282	5,807	YES	-
MUSM13339	E	WAM	464,635	464,214	147,965	33,224	0.0074	0.0033	31,088	6,636	YES	-
MUSM13342	E	WAM	1,264,796	1,263,564	265,383	59,592	0.0076	0.0027	57,254	11,850	YES	-
MVZ155151	D	NSM	1,534,418	1,533,053	312,250	58,583	0.007	0.0029	55,866	8,239	YES	-
MVZ157855	D	SPB	3,537,288	3,533,940	385,874	74,723	0.0071	0.0028	71,117	10,370	YES	YES
MVZ157905	D	NSM	2,862,935	2,860,277	406,759	74,530	0.0085	0.0028	70,955	10,632	YES	YES
MVZ157968	E	STR	184,713	184,529	90,165	7,409	0.0075	0.0042	6,671	1,157	YES	-
MVZ160093	D	NAR	2,106,276	2,104,366	327,408	64,296	0.0055	0.0027	61,564	9,976	YES	-
MVZ160094	C	SVZ	3,725,061	3,721,306	411,353	78,198	0.01	0.0015	74,657	10,102	YES	YES

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
MVZ166815	E	WAM	2,047,998	2,046,209	287,882	63,448	0.0075	0.0028	60,876	12,275	YES	-
MVZ168942	A	MMR	761,192	760,521	134,547	45,864	0.0068	0.0032	43,554	7,021	YES	-
MVZ168948	-	-	12,189	12,120	8,003	33	0.0239	0.0107	22	NaN	-	-
MVZ168949	-	-	20,168	20,108	17,230	43	0.0303	0.0213	24	2	-	-
MVZ168953	D	SPB	1,232,239	1,231,094	253,321	54,686	0.0051	0.0028	52,500	7,295	YES	-
MVZ168955	E	WAM	3,325,897	3,322,862	419,611	77,869	0.0082	0.0027	74,242	13,987	YES	YES
MVZ190699	D	JUR	2,215,186	2,213,146	394,036	70,819	0.0065	0.0028	67,316	10,863	YES	YES
MVZ190951	A	CLJ	2,206,125	2,204,071	411,744	72,759	0.0101	0.0032	68,253	11,308	YES	-
MVZ190954	A	CLJ	838,035	837,273	165,903	48,127	0.0076	0.0032	45,890	7,498	YES	-
MVZ194439	D	SSM	1,898,970	1,897,183	327,530	68,419	0.0084	0.0029	65,213	10,268	YES	YES
MVZ194463	D	SSM	1,564,475	1,563,088	312,188	61,919	0.0081	0.0028	59,034	8,990	YES	-
MVZ194474	-	-	99,366	99,228	61,815	1,176	0.0143	0.0074	994	148	-	-
MVZ194485	D	SSM	487,623	487,216	155,954	32,236	0.0075	0.0032	30,156	4,630	YES	-
MVZ194491	D	JUR	535,718	535,252	150,172	36,321	0.0049	0.003	34,519	5,354	YES	-
MVZ194493	D	JUR	941,957	941,104	240,458	50,590	0.0051	0.0029	48,366	7,719	YES	YES
MVZ194511	A	SSR	760,766	760,024	97,646	21,155	0.0072	0.0021	20,228	1,283	YES	YES
MVZ194545	A	JMI	2,183,678	2,181,345	229,409	62,652	0.0106	0.0017	60,051	6,344	YES	YES
MVZ194567	A	GJR	1,884,446	1,882,495	325,434	66,022	0.0108	0.0019	63,035	5,933	YES	YES
MVZ194582	A	ATH	839,115	838,414	159,305	48,548	0.0095	0.002	46,541	4,506	YES	YES

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
MVZ194602	E	WAM	200,474	200,266	94,043	8,618	0.0084	0.0038	7,799	1,587	YES	-
MVZ194635	E	WAM	376,080	375,731	77,506	23,096	0.0073	0.0031	22,019	4,584	YES	-
MVZ194703	E	WAM	1,124,815	1,123,796	272,101	56,960	0.0073	0.0027	54,706	11,490	YES	-
MVZ194711	E	WAM	1,567,176	1,565,740	328,818	62,192	0.007	0.0028	59,526	12,219	YES	-
MVZ194775	E	WAM	1,092,033	1,091,058	245,858	56,383	0.0076	0.0028	54,036	11,240	YES	-
MVZ194874	A	UJR	2,626,650	2,624,245	474,978	79,458	0.011	0.003	74,193	11,762	YES	YES
MVZ194879	A	UJR	1,759,530	1,757,966	312,783	63,610	0.0095	0.0033	60,231	9,827	YES	YES
MVZ194909	A	CLJ	2,252,447	2,250,279	389,587	69,294	0.0084	0.0029	65,788	10,905	YES	-
MVZ194914	A	CLJ	3,147,232	3,144,201	473,314	80,987	0.0094	0.003	76,063	11,983	YES	YES
MVZ194987	A	CLJ	2,649,630	2,647,221	439,430	78,430	0.0091	0.0031	73,994	12,946	YES	YES
MVZ194997	A	CLJ	857,157	856,421	189,878	47,150	0.0073	0.0032	44,884	7,379	YES	-
MVZ195034	A	CLJ	353,110	352,818	120,168	23,300	0.0074	0.0038	21,533	3,357	YES	-
MVZ195036	A	CLJ	1,179,854	1,178,846	285,308	56,034	0.0081	0.0033	53,087	8,769	YES	-
MVZ196095	D	MAR	151,445	151,231	72,935	5,230	0.0063	0.0035	4,868	437	YES	-
MVZ225064	D	SEM	3,403,305	3,399,963	476,625	80,606	0.0067	0.0017	76,861	8,076	YES	YES
MVZ225082	HG	HG	3,673,839	3,670,151	455,668	84,417	0.0077	0.0016	80,392	4,621	YES	YES
MZUSP30365	B	CER	1,800,847	1,799,198	175,257	48,241	0.0052	0.0018	46,847	12,404	YES	-
MZUSP30370	B	CER	3,192,286	3,189,447	301,742	73,339	0.0053	0.0018	71,071	19,316	YES	YES
MZUSP31926	D	MTR	1,194,053	1,192,881	173,448	57,420	0.0065	0.0019	55,655	9,396	YES	-

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
MZUSP31927	D	MTR	227,536	227,290	107,363	9,369	0.0075	0.003	8,811	1,354	YES	-
MZUSP31937	D	MTR	1,058,864	1,057,854	150,208	53,781	0.0057	0.0019	52,130	8,941	YES	-
MZUSP31939	D	MTR	653,205	652,644	144,974	42,525	0.0057	0.0023	40,806	6,589	YES	-
MZUSP31942	B	TXI	2,980,107	2,977,260	300,271	74,560	0.0069	0.0019	72,083	15,857	YES	YES
MZUSP31944	B	TXI	1,227,139	1,225,962	211,715	57,908	0.0061	0.002	56,254	11,413	YES	-
MZUSP31945	B	TXI	1,010,552	1,009,656	168,453	53,242	0.0046	0.002	51,889	10,531	YES	-
MZUSP31946	B	TXI	1,781,377	1,779,723	205,133	64,885	0.0047	0.0018	63,311	13,055	YES	-
MZUSP31947	B	XAI	2,852,271	2,849,567	309,353	73,524	0.006	0.0018	71,224	18,264	YES	YES
MZUSP31948	B	XAI	2,285,354	2,283,143	241,812	64,300	0.0058	0.0017	62,472	15,914	YES	-
MZUSP31950	B	CER	1,282,946	1,281,710	187,577	59,812	0.0047	0.0017	58,297	15,948	YES	-
MZUSP31951	B	CER	468,984	468,544	133,021	33,395	0.0042	0.0019	32,304	8,353	YES	-
NUTRIA289	OUT	OUT	2,242,635	2,240,340	328,132	68,584	0.0049	0.0017	66,073	1,412	YES	-
PLVP642	A	NSR	787,514	786,738	178,944	42,602	0.0053	0.0029	40,633	2,961	YES	-
PNPA321	D	PTC	938,349	937,507	147,346	47,426	0.006	0.002	45,925	8,767	YES	-
PNPA357	D	PTC	808,389	807,577	182,576	51,617	0.0058	0.0023	49,762	9,779	YES	-
RGM853	A	PTI	123,404	123,300	48,415	3,482	0.0082	0.0044	3,212	534	YES	-
RGM856	A	PTI	2,279,300	2,277,207	258,347	64,778	0.007	0.002	62,392	12,652	YES	-
ROM115116	TEP	TEP	238,801	238,591	70,571	10,508	0.0078	0.0025	9,977	411	YES	-
ROM117526	C	GUS	866,998	866,295	169,553	46,251	0.0053	0.0016	44,776	7,295	YES	-

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
ROM119901	D	NAR	1,560,168	1,558,720	248,007	59,858	0.0055	0.0027	57,605	9,396	YES	-
TTU100580	D	SEM	3,093,367	3,090,214	397,422	73,030	0.0058	0.0015	70,134	7,716	YES	-
TTU101118	E	IQT	1,836,157	1,834,449	286,697	63,188	0.0069	0.0027	60,790	8,712	YES	YES
TTU101173	D	LOR	1,636,788	1,635,185	352,978	60,685	0.0051	0.003	57,877	9,016	YES	YES
TTU101179	D	NSM	855,803	855,032	222,638	45,829	0.0069	0.0032	43,551	6,554	YES	-
TTU101195	D	NSM	1,310,786	1,309,607	301,533	55,506	0.0078	0.0029	52,837	7,987	YES	-
TTU101213	D	NSM	1,915,489	1,913,686	378,740	68,192	0.0076	0.003	64,779	10,250	YES	YES
TTU102638	D	ECU	1,382,556	1,381,189	211,174	53,709	0.0059	0.0016	51,865	4,933	YES	-
TTU102971	D	SEM	1,440,289	1,438,765	264,596	61,803	0.0066	0.0017	59,645	6,079	YES	-
TTU102977	D	SEM	3,233,333	3,230,335	304,958	62,083	0.0079	0.0021	59,344	5,615	YES	-
TTU103310	D	ECU	892,003	891,314	157,133	31,164	0.0061	0.0021	29,649	2,726	YES	-
TTU106013	C	GUS	2,039,625	2,037,504	294,435	62,661	0.0063	0.0016	60,233	9,207	YES	YES
UFMG4356	B	XAI	833,062	832,334	155,481	44,778	0.006	0.0019	43,533	9,536	YES	-
UFPB6929	B	CER	287,208	286,904	137,079	11,526	0.0067	0.0032	10,830	2,972	YES	-
UFPB6931	B	CER	383,739	383,329	154,586	21,081	0.0062	0.0028	19,943	5,653	YES	-
UFPB6932	B	CER	385,782	385,360	154,533	21,726	0.0062	0.0027	20,600	5,642	YES	-
UMMZ92712	-	-	1,501	1,497	341	44	0	0.001	44	NaN	-	-
UNIBAN1392	-	-	2,272,923	2,270,803	424,651	101,336	0.0066	0.0031	95,969	18	-	-

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
USNM499715	D	MAR	94,827	94,713	59,833	934	0.0151	0.007	781	68	YES	-
USNM549559	D	SAR	1,413,419	1,412,033	279,142	54,892	0.005	0.0031	52,545	8,395	YES	YES
USNM549567	D	SAR	155,285	155,147	45,256	10,454	0.0045	0.0024	10,053	1,501	YES	-
USNM568055	D	NAR	2,147,286	2,145,116	328,274	63,699	0.006	0.0016	61,240	9,971	YES	-
USNM579697	E	WAM	732,374	731,663	184,914	38,285	0.0057	0.0029	36,507	7,855	YES	-
USNM581908	D	PTC	460,827	460,435	134,748	31,038	0.0059	0.0027	29,538	5,384	YES	-
USNM584593	D	SCE	2,479,653	2,477,161	337,134	66,992	0.0076	0.0015	64,331	11,416	YES	YES
USNM619001	E	WAM	547,292	546,776	169,348	38,171	0.0047	0.0029	36,361	8,004	YES	-
USNM619002	A	MMR	590,866	590,284	187,808	39,019	0.0063	0.0031	36,796	6,228	YES	-
USNM619003	E	WAM	1,977,765	1,975,941	374,247	65,072	0.0066	0.0027	62,080	12,218	YES	-
USNM619004	C	NRB	2,387,856	2,385,490	301,986	65,780	0.0073	0.0015	63,319	9,595	YES	-
USNM619005	D	SCE	1,026,413	1,025,374	204,138	52,318	0.0062	0.0016	50,657	9,054	YES	-
USNM619006	D	SCE	1,905,902	1,903,979	252,959	64,256	0.0066	0.0015	62,111	11,114	YES	-
USNM619007	E	WAM	1,510,365	1,508,785	239,575	58,534	0.0068	0.0015	56,715	11,576	YES	-
USNM619008	E	WAM	2,476,190	2,473,754	351,297	68,969	0.0063	0.0015	66,527	13,159	YES	-
X1M15	D	SAR	1,357,606	1,356,316	198,154	62,148	0.0048	0.0018	60,294	10,836	YES	YES
X1M19	B	XAI	422,439	422,024	111,943	31,681	0.0066	0.0021	30,603	7,224	YES	-
X1M24	D	SAR	1,072,492	1,071,421	158,582	51,832	0.0048	0.0019	50,288	8,825	YES	-

Table S3: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained	CT	ST
X1M36	A	EAM	2,394,628	2,392,385	319,232	71,479	0.007	0.0019	68,925	11,300	YES	YES
Z9P01	B	CER	2,539,137	2,536,748	331,317	71,054	0.0055	0.0018	68,857	19,039	YES	-

Table S4: ICC test for cranial measurements with upper and lower values for the confidence interval. Bold variables were discarded for subsequent analyzes because they had a lower limit of the confidence interval below 0.8. N = number of measured skulls; k = number of times each skull was measured; ICC = ICC value, closer to 1, greater ability to correctly repeat measurement; Lower IC = lower limit of ICC confidence interval; Upper IC = upper limit of the ICC confidence interval.

Variables	N	k	ICC	Lower IC	Upper IC
RB	40	5	0.916	0.877	0.955
IOC	40	5	0.985	0.978	0.992
ZB	40	5	0.995	0.992	0.997
MB	40	5	0.987	0.980	0.993
GSL	40	5	1.000	1.000	1.000
NL	40	5	0.994	0.991	0.997
RL	40	5	0.997	0.995	0.998
OL	40	5	0.972	0.958	0.985
CD	**40**	**5**	**0.825**	**0.751**	**0.900**
RD	40	5	0.978	0.968	0.989
CDM1	40	5	0.994	0.991	0.997
BaL	40	5	0.998	0.998	0.999
PLa	40	5	0.992	0.988	0.996
PLb	**40**	**5**	**0.827**	**0.753**	**0.900**
PPL	**40**	**5**	**0.811**	**0.770**	**0.852**
CIL	40	5	0.999	0.999	1.000
MTRL	40	5	0.997	0.995	0.998
MaxB	40	5	0.970	0.956	0.985
BUM	**40**	**5**	**0.824**	**0.749**	**0.899**
D	40	5	0.963	0.945	0.981
IFL	40	5	0.972	0.959	0.986
IFW	40	5	0.987	0.980	0.993
BuL	40	5	0.940	0.912	0.968
MFW	**40**	**5**	**0.796**	**0.712**	**0.881**
OccW	40	5	0.943	0.916	0.970
DBU	**40**	**5**	**0.808**	**0.727**	**0.888**
BUH	**40**	**5**	**0.737**	**0.635**	**0.840**
LMD	40	5	0.926	0.891	0.960
MD	40	5	0.982	0.973	0.991

Table S5: Shapiro-Wilk Normality test values (W) and their p-values for the 22 morphometric variables used for the species delimitation analyzes. I also created QQ-plots for each variable. I did not show them here because they presented same results then Shapiro-Wilk tests. Values in bold indicate variables with non-normal distribution and they were eliminated.

Variables	W	p-value
RB	0.995	0.109
IOC	**0.991**	**0.004**
ZB	0.997	0.625
MB	0.997	0.537
GSL	0.998	0.849
NL	0.994	0.060
RL	0.995	0.182
OL	**0.982**	**0.000**
RD	0.996	0.259
CDM1	0.997	0.599
BaL	0.998	0.806
PLa	**0.993**	**0.025**
CIL	**0.916**	**0.000**
MTRL	**0.974**	**0.000**
MaxB	**0.986**	**0.000**
D	0.997	0.385
IFL	**0.993**	**0.035**
IFW	**0.992**	**0.008**
BuL	0.995	0.138
OccW	0.997	0.428
LMD	**0.993**	**0.019**
MD	**0.993**	**0.021**

Table S6: Pearson correlation coefficient (r) among the 12 normal-distributed variables used for the species delimitation analyzes. Values in bold indicate variables with r larger than 0.5 and we eliminated the NL and RL variables.

	RB	**ZB**	**MB**	**GSL**	**NL**	**RL**	**RD**	**CDM1**	**BaL**	**D**	**BuL**
RB											
ZB	0.27										
MB	0.20	0.25									
GSL	0.13	0.20	0.09								
NL	0.08	0.03	0.02	**0.57**							
RL	0.04	0.16	0.05	**0.68**	**0.72**						
RD	0.11	0.12	0.06	0.32	0.28	0.30					
CDM1	0.01	0.26	0.11	0.08	0.01	0.01	0.45				
BaL	0.14	0.21	0.00	0.41	0.26	0.35	0.01	0.20			
D	0.05	0.16	0.04	0.31	0.25	0.39	0.16	0.05	0.43		
BuL	0.35	0.20	0.43	0.03	0.23	0.15	0.10	0.09	0.13	0.05	
OccW	0.29	0.20	0.31	0.05	0.09	0.03	0.14	0.14	0.04	0.07	0.30

Table S7: Available cytogenetic, mtDNA, and nuDNA data in the literature for the 41 subclades from the conservative trees of *Proechimys*, organized by the five main clades. Names and acronyms of the subclades, original source for the data (References), and how the association with the present study was made (Associated by): the same specimen used in the conservative trees, other individuals from the same locality or by specimens from a nearby locality. Locality from original source is informed as well as the taxa name given by the authors. Diploid number (2n) and fundamental number (NF) are reported, and when available the name of the clade and marker recovered by mitochondrial and nuclear DNA. Questions marks (?) indicate that it was not possible to associate the literature information to the subclade.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
A	Purus-Tapajós Interfluve (PTI)	*P.* gr. *longicaudatus*	BRAZIL: Mato Grosso: Cotriguaçu and Querência	cytogenetics	2n = 16-17; NF = 14	nearby locality	Amaral et al. 2013
		Proechimys sp.3	BRAZIL: Pará: Jacareacanga and Flexal	cytogenetics	2n = 14-16; NF = 16	nearby locality	Barros 1978
		Proechimys sp.	BRAZIL: Mato Grosso: Juruena	cytogenetics	2n = 17; NF = 16	same locality	Machado 2017
		P. gr. *goeldii*	BRAZIL: Mato Grosso: Juruena	cytogenetics	2n = 15; NF = 16	same specimen (MZUSP31924)	Machado et al. 2005
		P. cf. *longicaudatus*	BRAZIL: Pará: Itaituba and Jacareacanga	cytogenetics	2n = 16-17; NF = 14	nearby locality	Rodrigues da Costa et al. 2016
		Proechimys sp.	BRAZIL: Rondônia: Morrinhos	mtDNA	Cyt b	nearby locality	Schetino 2008
	Central-Lower Juruá River (CLJ)	*P. steerei*	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b	same locality	Da Silva and Patton 1998; Patton et al. 2000
		P. steerei	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b	same specimen (MVZ194914)	Matocq et al. 2000
		P. steerei	BRAZIL: Amazonas: Central and Lower Juruá River	cytogenetics	2n = 24; NF = 40-41	same specimen (MVZ194914)	Patton et al. 2000
		P. steerei	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b	same specimen (MVZ190951)	Schetino 2008

Table S7: Continuation.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
A	Madidi-Madre de Dios rivers (MMR)	*P. steerei*	PERU: Madre de Dios: Pakitza	cytogenetics	2n = 24; NF = 42	nearby locality	Gardner and Emmons 1984
		P. steerei	PERU: Madre de Dios: Cusco Amazônico	mtDNA	Cyt b	nearby locality	Patton et al. 2000
	Upper Juruá River (UJR)	*P. steerei*	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b	same locality	Da Silva and Patton 1998; Patton et al. 2000
		P. steerei	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b	same specimen (MVZ194879)	Matocq et al. 2000
		P. steerei	BRAZIL: Amazonas: Upper Juruá River	cytogenetics	2n = 24; NF = 42	same specimen (MVZ194879)	Patton et al. 2000
	Eastern Amazon (EAM)	*P. goeldii*	BRAZIL: Pará: Altamira, Xingu River	mtDNA	Cyt b	same locality	Da Silva and Patton 1998; Patton et al. 2000; Schetino 2008
		P. goeldii	BRAZIL: Pará: Altamira, Xingu River	cytogenetics	2n = 24; NF = 44	same locality	Patton et al. 2000
	Içá-Japurá-Madeira rivers (IJM)	*P. quadruplicatus*	BRAZIL: Amazonas: Santa Isabel do Rio Negro	cytogenetics	2n = 28; NF = 42	nearby locality	Bonvicino et al. 2005
		P. quadruplicatus	BRAZIL: Amazonas: Lago Meduiním, Negro River	mtDNA	Cyt b	nearby locality	Da Silva and Patton 1998; Patton et al. 2000; Schetino 2008
		P. quadruplicatus	BRAZIL: Amazonas: Lago Meduiním, Negro River	cytogenetics	2n = 28; NF = 42	nearby locality	Patton et al. 2000

Table S7: Continuation.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
A	Juruá-Madeira Interfluve (JMI)	*P. gardneri*	BRAZIL: Amazonas: Juruá River	cytogenetics	2n = 40; NF = 56	same specimen (MVZ194545)	Da Silva 1998; Patton et al. 2000
		P. gardneri	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b	same locality	Da Silva 1998; Patton et al. 2000; Schetino 2008
	Acre-Ta Huamanu rivers (ATH)	*P. pattoni*	BRAZIL: Amazonas: Juruá River	cytogenetics	2n = 40; NF = 56	same specimen (MVZ194582)	Da Silva 1998; Patton et al. 2000
		P. pattoni	BRAZIL: Amazonas: Juruá River	mtDNA		same specimen (MVZ194582)	Da Silva 1998; Patton et al. 2000; Schetino 2008
	Galvez-Juruá rivers (GJR)	*P. kulinae*	BRAZIL: Amazonas: Juruá River	cytogenetics	2n = 34; NF = 52	same specimen (MVZ194567)	Da Silva 1998; Patton et al. 2000
		P. kulinae	BRAZIL: Amazonas: Juruá River	mtDNA		same specimen (MVZ194567)	Da Silva 1998; Patton et al. 2000; Schetino 2008
	Lower Madeira rivers (LMR)	?	?	?	?	?	?
	North Solimões River (NSR)	-	BRAZIL: Amazonas: Japurá and Santo Antônio do Içá	cytogenetics	2n = 32; NF = 58	same specimen (JAP012)	Moreira, *pers. comm.*
		P. echinothrix	BRAZIL: Amazonas: São Gabriel da Cachoeira	mtDNA	Cyt b (a-d)	nearby locality	Patton et al. 2000
		P. echinothrix	BRAZIL: Amazonas: Novo Airão	mtDNA	Cyt b (Jau + Tiquiê)	nearby locality	Schetino 2008

Table S7: Continuation.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
A	South Solimões River (SSR)	*P. echinothrix*	BRAZIL: Amazonas: Juruá River	cytogenetics	2n = 32; NF = 60	same specimen (MVZ194511)	Da Silva 1998; Patton et al. 2000
		P. echinothrix	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b (e)	same locality	Da Silva 1998; Patton et al. 2000
		P. echinothrix	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b	same specimen (MVZ194511)	Schetino 2008
B	Cerrado (CER)	*P. roberti*	BRAZIL: Tocantins: Peixe	mtDNA + nuDNA	Clade East	same specimen (MZUSP31950)	Leite 2013
		P. roberti	BRAZIL: Piauí: Bom Jesus	cytogenetics	2n = 30; NF = 56	same specimen (MZUSP30365)	Machado et al. 2005
		P. roberti	BRAZIL: Tocantins: Peixe	mtDNA	Cyt b (Santa Teresa + Fiandeiras)	same specimen (CTA1652)	Schetino 2008
		P. roberti	BRAZIL: Tocantins: São Sebastião	cytogenetics		same specimen (MN76211)	Weksler et al., 2001
	Xingu-Araguaia Interfluve (XAI)	*P. roberti*	BRAZIL: Mato Grosso: Vila Rica	mtDNA + nuDNA	Clade Plateau	same specimen (MZUSP31947)	Leite 2013
		P. roberti	BRAZIL: Mato Grosso: Vila Rica	cytogenetics	2n = 30; NF = 56	same specimen (MZUSP31947);	Machado et al. 2005
		P. roberti	BRAZIL: Pará: Paraupebas	mtDNA	Cyt b	same locality	Schetino 2008
	Tapajós-Xingu Interfluve (TXI)	*P. roberti*	BRAZIL: Mato Grosso: Claúdia	mtDNA + nuDNA	Clade Upper West	same specimen (MZUSP31944)	Leite 2013
		P. roberti	BRAZIL: Mato Grosso: Gaúcha	cytogenetics	2n = 30; NF = 56	same specimen (MZUSP31945);	Machado et al. 2005
		P. roberti	BRAZIL:Mato Grosso: A. Floresta	mtDNA	Cyt b (Rio Cristalino)	nearby locality	Schetino 2008

Table S7: Continuation.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
B	Lower Xingu River (LXR)	?	?	?	?	?	?
	Abacaxis River (ABX)	*Proechimys* sp.1	BRAZIL: Pará: Jacareacanga and Flexal	cytogenetics	2n = 30; NF = 56	nearby locality	Barros 1978
	Northwestern Venezuela (NVZ)	*P. guairae* "Falcon subspecies"	VENEZUELA: Falcón: Serranía de San Luis, Cabure	cytogenetics	2n = 46-47; NF = 72-74	same locality	Aguilera et al. 1995
	Negro River Basin (NRB)	*Proechimys* sp.A	BRAZIL: Amazonas: Barcelos and Santa Isabel do Rio Negro	cytogenetics	2n = 38; NF = 52	same locality	Bonvicino et al. 2005
		P. guyannensis	BRAZIL: Amazonas: Pico da Neblina	mtDNA	Cyt b (Pico da Neblina)	nearby locality	Schetino 2008
C	Guiana Shield (GUS)	*P. guyannensis*	BRAZIL: Pará: Oriximiná	cytogenetics	2n = 46; NF = 50	same specimen (CTA4226)	Machado 2017
		P. guyannensis	FRENCH GUYANA: Cayenne and Saül	cytogenetics	2n = 40; NF = 54	nearby locality	Patton et al. 2000
		P. guyannensis	BRAZIL: Pará: Curuá	mtDNA	Cyt b	nearby locality	Silva et al. 2018
		P. cayennensis	FRENCH GUYANA: Nouragues	mtDNA	Cyt b; Control region	nearby locality	Steiner et al. 2000
		P. guyannensis	FRENCH GUYANA: Cayenne	mtDNA	Cyt b; Control region	nearby locality	Van Vuuren et al. 2004
	Southeastern Venezuela (SVZ)	*P. guyannensis*	VENEZUELA: Bolivar: Cairara del Orinoco	cytogenetics	2n = 40; NF = 52	nearby locality	Reig and Useche 1976
		P. guyannensis	VENEZUELA: Bolivar: El Pauji	mtDNA	Cyt b (El Pauji)	same specimen (MVZ160094)	Schetino 2008

Table S7: Continuation.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
D	South Solimões-Marañón rivers (SSM)	*P. brevicauda*	BRAZIL: Amazonas: Juruá River	cytogenetics	2n = 28; NF = 48-50	same specimen (MVZ194485)	Patton et al. 2000
		P. brevicauda	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b (a+b+Juruá)	same locality	Patton et al. 2000
		P. brevicauda	PERU: Loreto: Gálvez River	mtDNA	Cyt b	same specimen (MUSM13338)	Schetino 2008
	North Solimões-Marañón rivers (NSM)	*Proechimys* sp.2	PERU: Loreto: Aupauayo	cytogenetics	2n = 30; NF = 50	nearby locality	Aniskin 1994
		P. brevicauda	PERU: Amazonas: Huampani, Cenepa River	cytogenetics	2n = 30; NF = 48	same locality	Gardner and Emmons 1984
		P. brevicauda	PERU: Amazonas: Huampani, Cenepa River	mtDNA	Cyt b (a+b+Juruá)	same locality	Patton et al. 2000
		P. brevicauda	PERU: Amazonas: Huampani, Cenepa River	mtDNA	Cyt b (Rio Cenepa)	same specimen (MVZ155121)	Schetino 2008
	South Peru-Bolivia (SPB)	*P. brevicauda*	PERU: Madre de Dios: Tambopata	cytogenetics	2n = 28; NF = 48	nearby locality	Gardner and Emmons 1984
		P. brevicauda	PERU: Madre de Dios: Lago Sandoval	mtDNA	Cyt b (c+d)	same specimen (MVZ157855)	Patton et al. 2000
	Upper Madre de Dios River (UMD)	?	?	?	?	?	?
	Pantanal-Chaco (PTC)	*P. longicaudatus*	BOLIVIA: SANTA CRUZ: El Refugio	mtDNA	Cyt b	nearby locality	Schetino 2008
	South Cerrado (SCE)	*P. longicaudatus*	BRAZIL: Mato Grosso: Porto Estrela	cytogenetics	2n = 28; NF = 50	same specimen (APC1085)	Machado 2017

Table S7: Continuation.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
D	South Cerrado (SCE)	*P. longicaudatus*	BRAZIL: Goiás: Parque Nacional das Emas	cytogenetics	2n = 28; NF = 50	same specimen (MN67246)	Machado et al. 2005
	Pando, Bolívia (PAB)	?	?	?	?	?	?
	North Amazon River (NAR)	*P. cuvieri*	BRAZIL: Pará: Oriximiná	cytogenetics	2n = 28; NF = 46	same specimen (CTA4245)	Machado 2017
		P. cuvieri	FRENCH GUYANA: La Trinité Mountains	mtDNA	Cyt b (c-g)	nearby locality	Patton et al. 2000
		P. cuvieri	FRENCH GUYANA: La Trinité Mountains	cytogenetics	2n = 28; NF = 50	nearby locality	Reig et al. 1979
		P. cuvieri	FRENCH GUYANA: Nouragues	mtDNA	Cyt b; Control region	nearby locality	Steiner et al. 2000
		P. cuvieri	GUYANA: Barima-Waini: Baramita	mtDNA	Cyt b; Control region	same specimen (USNM568055)	Van Vuuren et al. 2004; Silva et al. 2018
	South Amazon River (SAR)	*P. cuvieri*	BRAZIL: Pará: Altamira, Xingu River	cytogenetics	2n = 28; NF = 48	same specimen (USNM549559)	Patton et al. 2000
		P. cuvieri	BRAZIL: Pará: Altamira, Xingu River	mtDNA	Cyt b (c-g)	same specimen (USNM549559)	Patton et al. 2000
	Japurá River (JAP)	-	BRAZIL: Amazonas: Japurá	cytogenetics	2n = 28; NF = 48-50	same specimen (JAP098)	Moreira, *pers. comm.*
		P. cuvieri	BRAZIL: Amazonas: São Gabriel da Cachoeira	mtDNA	Cyt b (b)	nearby locality	Patton et al. 2000
		P. cuvieri	BRAZIL: Amazonas: São Gabriel da Cachoeira	mtDNA	Cyt b (Tiquiê + Santiago)	nearby locality	Schetino 2008

Table S7: Continuation.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
D	Juruá River (JUR)	*P. cuvieri*	BRAZIL: Amazonas: Juruá River	cytogenetics	2n = 28; NF = 46-48	same specimen (MVZ194491)	Patton et al. 2000
		P. cuvieri	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b	same locality	Patton et al. 2000
		P. cuvieri	BRAZIL: Amazonas: Juruá River	mtDNA	Cyt b (Juruá + Rio Blanco)	same locality	Schetino 2008
	Loreto, Peru (LOR)	?	?	?	?	?	?
	Mato Grosso region (MTR)	*P. longicaudatus*	BRAZIL: Mato Grosso: Aripuanã	cytogenetics	2n = 28; NF = 48-50	same specimen (MZUSP31926)	Machado et al. 2005
		Proechimys sp.	BRAZIL: Mato Grosso: Aripuanã	mtDNA + nuDNA	Clade TX	same specimen (MZUSP31926)	Parra 2017
		Proechimys sp.	BRAZIL: Amazonas: Margem direita do Rio Aripuanã	mtDNA	Cyt b	nearby locality	Schetino 2008
	Upper Madeira River (UMR)	*P.* gr. *longicaudatus*	BRAZIL: Rondônia: Jamari River	cytogenetics	2n = 30; NF = 52	nearby locality	Leal-Mesquita 1991
		P. longicaudatus	BRAZIL: Rondônia: Jamari River	cytogenetics	2n = 28; NF = 50	nearby locality	Machado et al. 2005
		Proechimys sp.	BRAZIL: Rondônia: Jamari River	mtDNA + nuDNA	Clade MJ	nearby locality	Parra 2017
	Central America-Chocó (CAM)	*P. semispinosus*	ECUADOR: Esmeraldas: 3 Km W Majua	cytogenetics	2n = 30; NF = 50-54	nearby locality	Gardner and Emmons 1984
		P. semispinosus	COSTA RICA: Limón: Cariari	cytogenetics	2n = 30; NF = 54	nearby locality	Patton and Gardner 1972

Table S7: Continuation.

Clades	Subclades	Taxa	Locality	Datatype	Data	Associated by	References
D	Magdalena River (MAR)	*P. chrysaeolus*	COLOMBIA: Nariño: Tumaco	cytogenetics	2n = 32; NF = 56	nearby locality	Bueno and Gomez-Laverde 1989
	Western Ecuador (ECU)	*P. decumanus*	ECUADOR: El Oro: 4 Km SE Santa Rosa	cytogenetics	2n = 30; NF = 54	nearby locality	Gardner and Emmons 1984
E	Western Amazon WAM	*P. simonsi*	BRAZIL: Amazonas: Juruá River	cytogenetics	2n = 32; NF = 58	same specimen (MVZ194711)	Patton et al. 2000
		P. simonsi	BOLIVIA: La Paz: Madidi River	mtDNA	Cyt b (v-z)	same specimen (USNM619001)	Matocq et al. 2000; Patton et al. 2000
		P. simonsi	PERU: Loreto: Gálvez River	mtDNA	Cyt b	same specimen (MUSM133339)	Schetino 2008
	Santiago River STR	*P. simonsi*	PERU: Amazonas: La Poza, Santiago River	mtDNA	Cyt b (u)	same locality	Patton et al. 2000
		P. simonsi	ECUADOR: Napo: Limoncocha	cytogenetics	2n = 32; NF = 58	nearby locality	Gardner and Emmons 1984
	Iquitos IQT	?	?	?	?	?	?

References:

Aguilera, M., Aldo Reig, O.A., Perez-Zapata, A., 1995. G-and C-banding karyotypes of spiny rats (*Proechimys*) of Venezuela. Rev. Chil. Hist. Nat. 68, 185–196.

Amaral, P.J., Nagamachi, C.Y., Noronha, R.C., Costa, M.J., Pereira, A.L., Rossi, R. V, Mendes-Oliveira, A.C., Pieczarka, J.C., 2013. *Proechimys* (Rodentia, Echimyidae): characterization and taxonomic considerations of a form with a very low diploid number and a multiple sex chromosome system. BMC Genet. 14, 21. https://doi.org/10.1186/1471-2156-14-21

Aniskin VM (1994). Evolutionary Cytogenetics of Spiny Rats from Peruvian Amazonia (Rodentia, Echimyidae). In: Mlekopitayushchie Peruanskoi Amazonii (Socolov VE, ed.). Nauka, Moscow, 62-82.

Barros RMS (1978). Variabilidade Cromossômica em Proechimys e Oryzomys (Rodentia) da Amazônia. Master's thesis, Instituto de Biociências, USP, São Paulo.

Bonvicino, C.R., Otazú, I.B., Vilela, J.F., 2005. Karyologic and molecular analysis of *Proechimys* Allen, 1899 (Rodentia, Echimyidae) from Amazonian region. Arq. do Mus. Nac. 63, 191–2000.

Bueno ML, Gomez-Laverde M and Morales A (1989). Caracterización cariológica de *Proechimys* sp (Rodentia: Echimyidae) de una colônia experimental. Biomedica 9: 13-22.

Da Silva, M.N., Patton, J.L., 1998. Molecular phylogeography and the evolution and conservation of amazonian mammals. Mol Ecol 7, 475–486.

Da Silva, M.N.F., 1998. Four new species of spiny rats of the genus *Proechimys* (Rodentia: Echimyidae) from the western Amazon of Brazil. Proc. Biol. Soc. Washingt. 111, 436–471.

Gardner, A.L., Emmons, L.H., 1984. Species groups in *Proechimys* (Rodentia, Echimyidae) as indicated by karyology and bullar morphology. J. Mammal. 65, 10–25.

Leal-Mesquita ERRBP (1991). Estudos Citogenéticos em Dez Espécies de Roedores Brasileiros da Família Echimyidae. Master's thesis, Instituto de Biociências, USP, São Paulo.

Leite, R.N., 2013. Biogeography and evolution of Neotropical small mammals, with emphasis on hystricognath spiny rats of the genus *Proechimys* (Family Echimyidae). Brigham Young University.

Machado, M.X., 2017. Cariologia e filogenia molecular em *Trinomys* e *Proechimys* (Echimyidae, Rodentia). Universidade Federal do Espírito Santo.

Machado, T., Silva, M.J. de J., Leal-Mesquita, E.R., Carmignotto, A.P., Yonenaga-Yassuda, Y., 2005. Nine karyomorphs for spiny rats of the genus *Proechimys* (Echimyidae, Rodentia) from North and Central Brazil. Genet. Mol. Biol. 28, 682–692.

Matocq, M.D., Patton, J.L., da Silva, M.N.F., 2000. Population genetic structure of two ecologically distinct amazonian spiny rats: separating history and current ecology. Evolution (N. Y). 54, 1423–1432. https://doi.org/10.1111/j.0014-3820.2000.tb00574.x

Parra, D. B., 2017. Filogeografia de um rato-de-espinho (*Proechimys* sp., Echimyidae) do sul da Amazônia sob o contexto das alterações ambientais do Quaternário. Instituto Nacional de Pesquisas da Amazônia.

Patton, J.L., Da Silva, M.N., Malcolm, J.R., 2000. Mammals of the rio Juruá and the evolutionary and ecological diversification of Amazonia. Bull. Am. Museum Nat. Hist. 244, 306.

Patton, J.L., Gardner, A.L., 1972. Notes on the systematics of *Proechimys* (Rodentia, Echimyidae), with emphasis on peruvian forms. Occas. Pap. 13, 1–30.

Reig OA and Useche M (1976). Diversidad cariotípica y sistemática en poblaciones venezolanas de *Proechimys* (Rodentia, Echimyidae), con datos adicionales sobre poblaciones de Perú y Colombia. Acta Cient. Venez. 27: 132-140.

Reig OA, Trainer M and Barros MA (1979). Sur l'ídentification chromosomique de *Proechimys guyannensis* (E. Geoffroy, 1803) et de *Proechimys cuvieri* Petter, 1978 (Rodentia, Echimyidae). Mammalia 43: 501-505.

Rodrigues da Costa, M.J., Siqueira do Amaral, P.J., Pieczarka, J.C., Sampaio, M.I., Rossi, R. V, Mendes-Oliveira, A.C., Rodrigues Noronha, R.C., Nagamachi, C.Y., 2016. Cryptic Species in *Proechimys goeldii* (Rodentia, Echimyidae)? A Case of Molecular and Chromosomal Differentiation in Allopatric Populations. Cytogenet. Genome Res. 148, 199–210. https://doi.org/10.1159/000446562

Schetino, M.A.A., 2008. Código de barras de DNA em espécies de *Proechimys* (Rodentia: Echimyidae) da Amazônia. Instituto Nacional de Pesquisas da Amazônia.

Silva, C.R., Ribas, C.C., Da Silva, M.N.F., Leite, R.N., Catzeflis, F., Rogers, D.S., De Thoisy, B., 2018. The role of Pleistocene climate change in the genetic variability, distribution and demography of Proechimys cuvieri and P. guyannensis (Rodentia: Echimyidae) in northeastern Amazonia. PLoS One 13, e0206660. https://doi.org/10.1371/journal.pone.0206660

Steiner, C., Sourrouille, P., Catzeflis, F., 2000. Molecular characterization and mitochondrial sequence variation in two sympatric species of Proechimys (Rodentia: Echimyidae) in French Guiana. Biochem. Syst. Ecol. 28, 963–973. https://doi.org/10.1016/S0305-1978(00)00021-1

Van Vuuren, B.J., Kinet, S., Chopelet, J., Catzeflis, F., 2004. Geographic patterns of genetic variation in four Neotropical rodents: conservation implications for small game mammals in French Guiana. Biol. J. Linn. Soc. 81, 203–218. https://doi.org/10.1111/j.1095-8312.2003.00278.x

Weksler, M., Bonvicino, C.R., Otazu, I.B., Silva, J.S., 2001. Status of *Proechimys roberti* and *P. oris* (Rodentia: Echimyidae) from eastern Amazonia and Central Brazil. J. Mammal. 82, 109–122. https://doi.org/10.1644/1545-1542(2001)082<0109:SOPRAP>2.0.CO;2

Figures

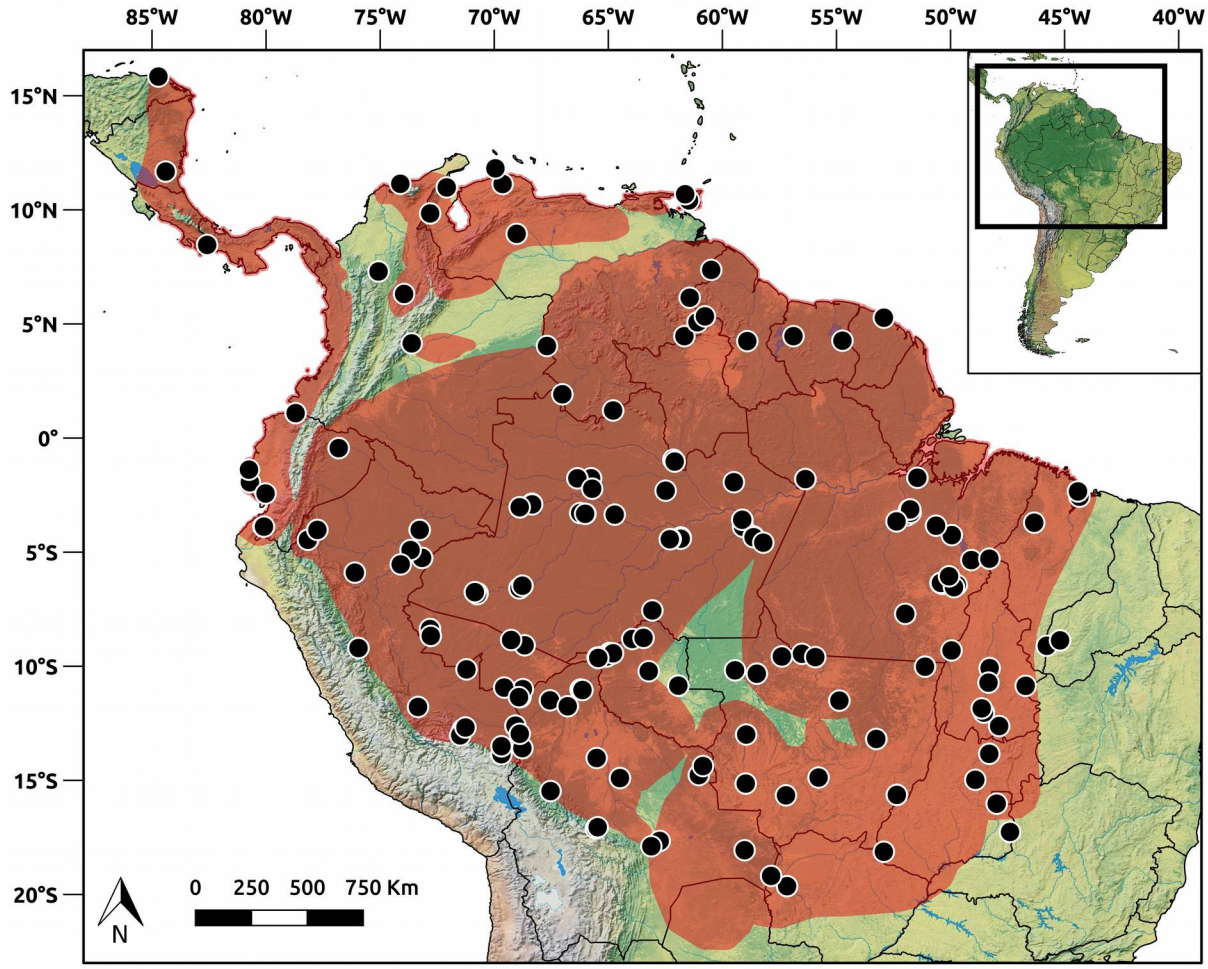


Figure S1: Geographical distribution of the 270 genetic samples used for the genomics analyzes. Red areas are the known distribution for *Proechimys*, according to IUCN (www.iucnredlist.org). Shades of green represent vegetation, lighter tones is open areas and dark tones forests. Main rivers and the mountains are also represented.

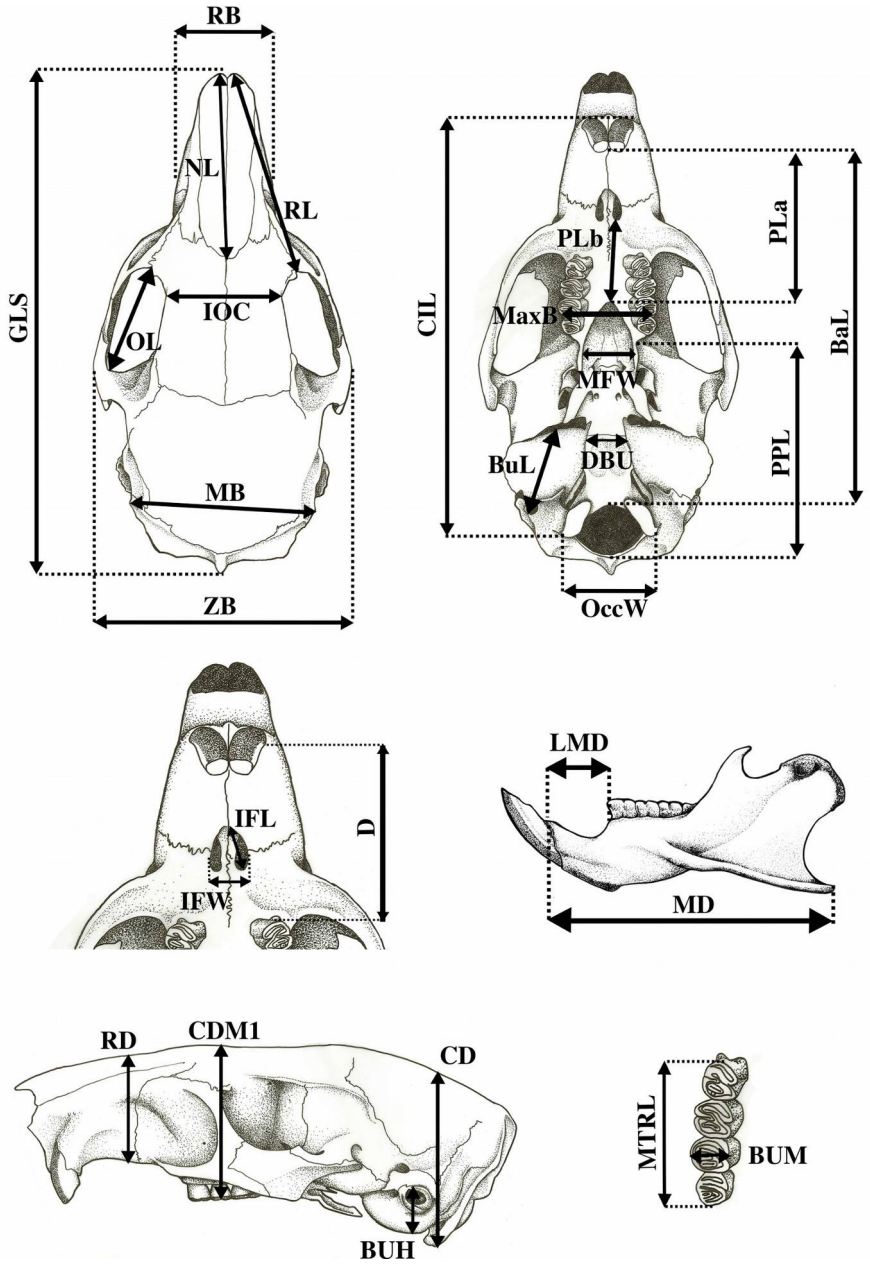


Figure S2: The 29 cranial measurements with their respective acronyms (illustration of a *Trinomys dimidiatus* made by Gustavo S. Libardi). **BaL:** basilar length of Hensel; **BuL:** bullar length; **BUH:** bullar depth; **BUM:** upper second molar width; **CD:** cranial depth; **CDM1:** cranial depth at M1; **CIL:** condyloincisive length; **D:** diastema length; **DBU:** distance between the bullae; **GSL:** greatest length of skull; **IFL:** length of incisive foramen; **IFW**: maximum width of incisive foramen; **IOC**: interorbital constriction; **LMD:** length of mandibular diastema; **MaxB:** maxillary breadth at M2-M3; **MB:** greatest breadth across mastoid; **MD:** mandibular length; **MFW:** mesopterigoid fossa width; **MTRL**: alveolar length of upper molariforms; **NL:** nasal length; **OccW**: occipital condyle width; **OL:** orbit length; **PLa:** palatal length A; **PLb:** palatal length B; **PPL:** post-palatal length; **RB:** rostral breadth; **RD:** rostral depth; **RL:** rostral length; **ZB:** zygomatic arch breadth.

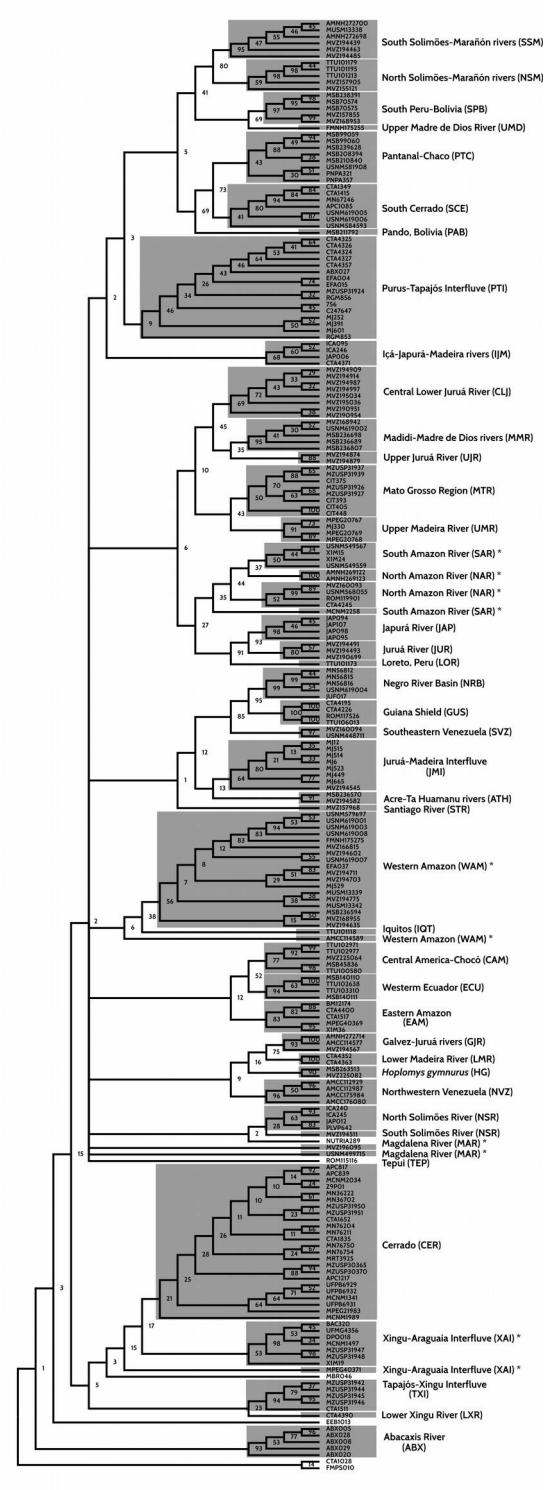


Figure S3: Phylogenetic tree for the genus *Proechimys* with 222 individuals and 88,129 SNPs built under the coalescent model (CM) with SVDquartets. Numbers on the nodes indicate the bootstrap values and the gray rectangles indicate subclades that were also retrieved in the tree with concatenated data and Maximum Likelihood (ML), see Fig. 1 and Fig. 2. Names and acronyms for subclades are next to the rectangles. Only five ML subclades were not monophyletic in MC tree, they are represented with an asterisk.

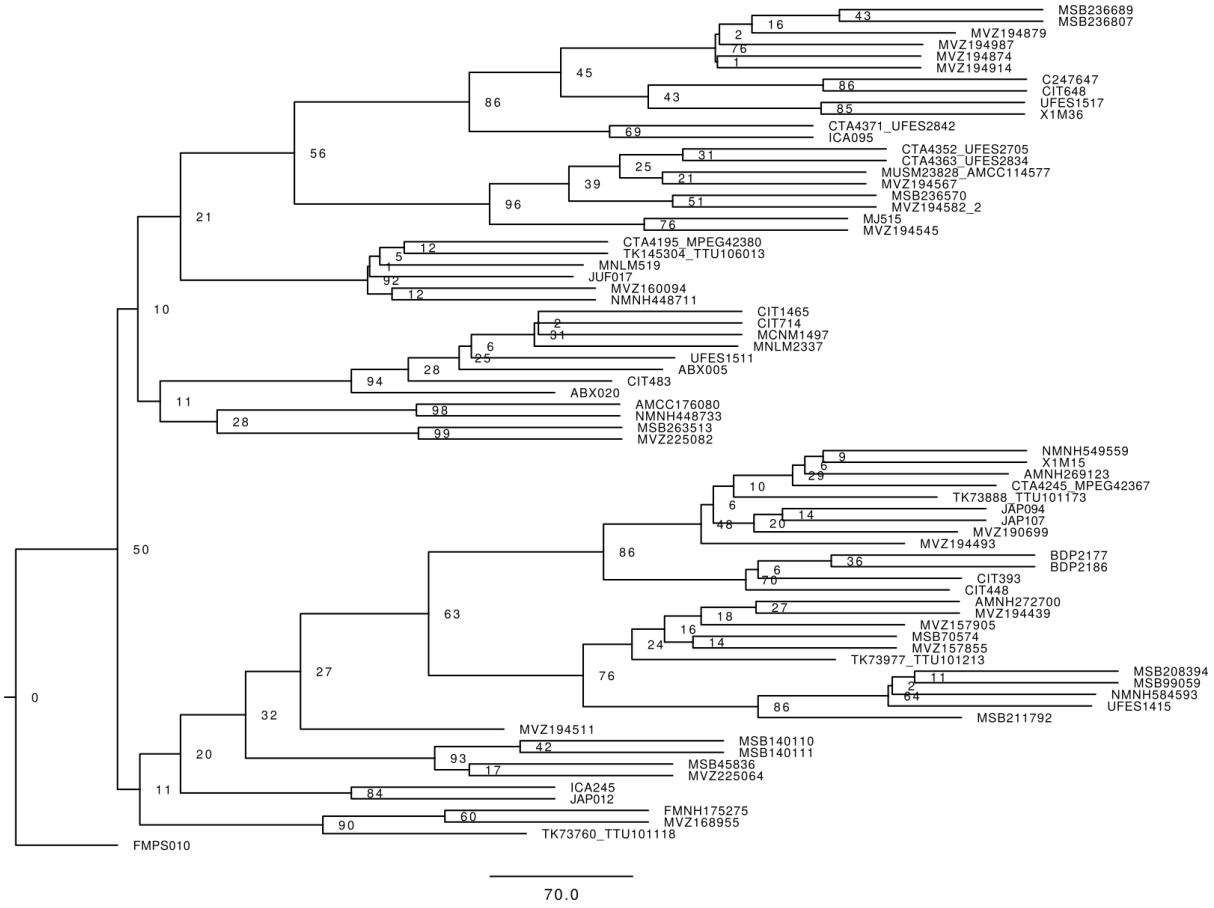


Figure S4: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 741 SNPs, and 30% of missing data inferred under the coalescent model (CM) with SVDquartets. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

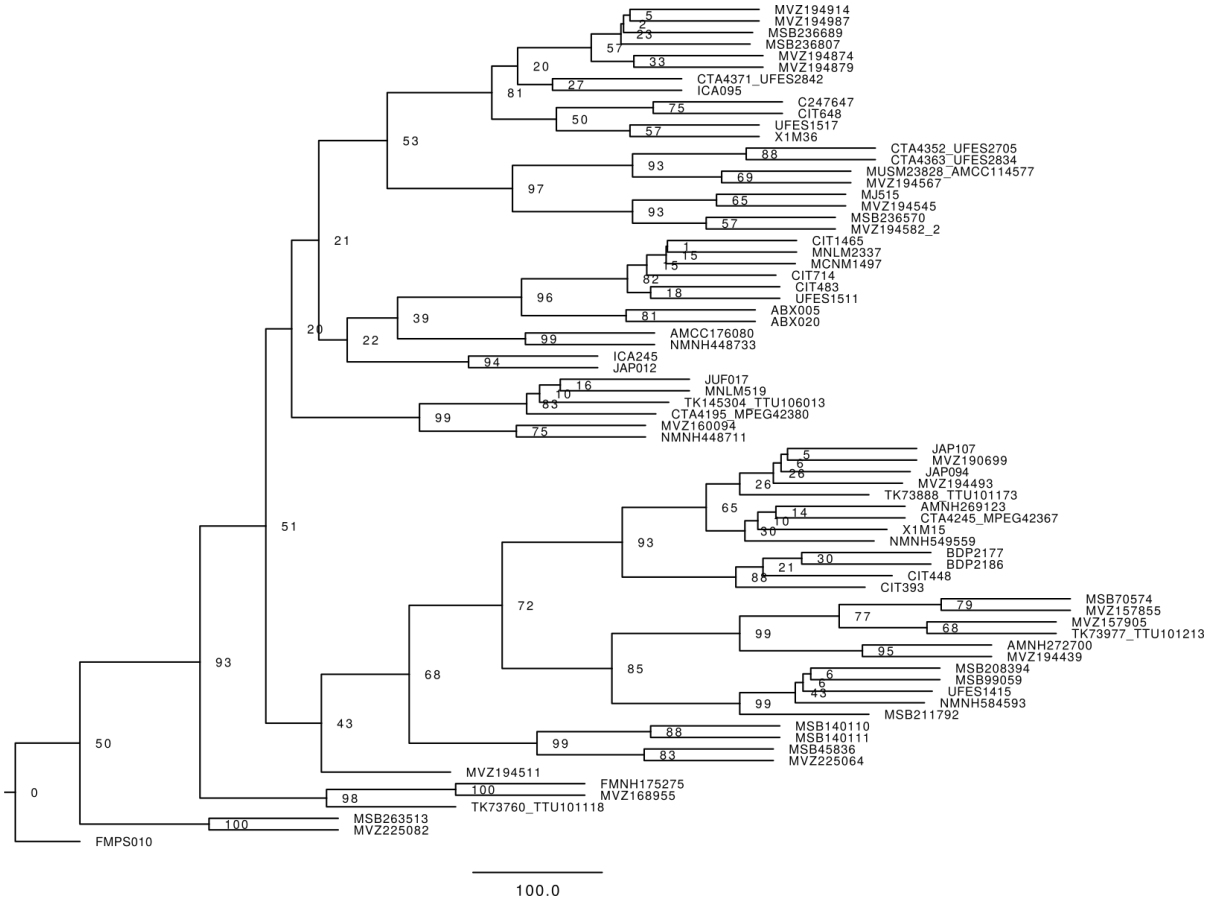


Figure S5: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 1,046 SNPs, and 40% of missing data inferred under the coalescent model (CM) with SVDquartets. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

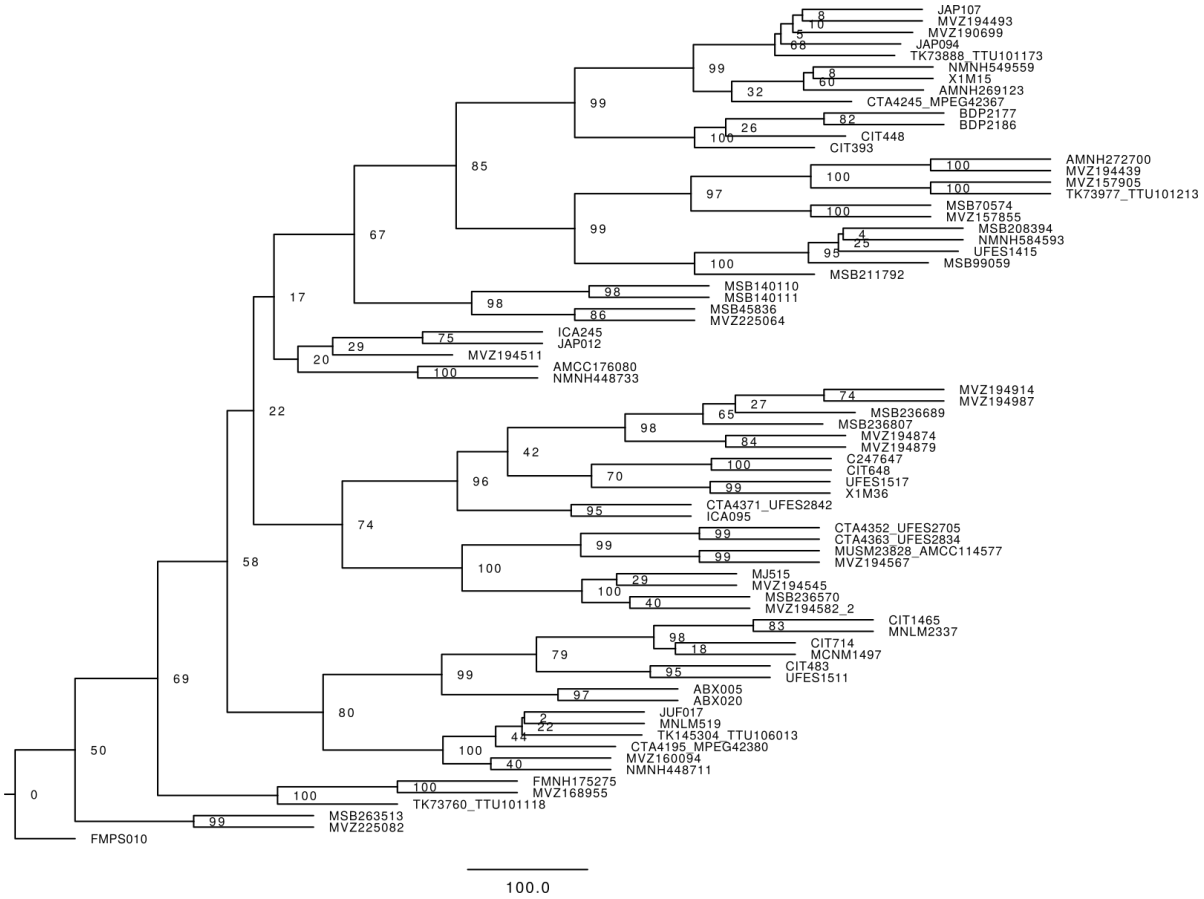


Figure S6: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 1,580 SNPs, and 50% of missing data inferred under the coalescent model (CM) with SVDquartets. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

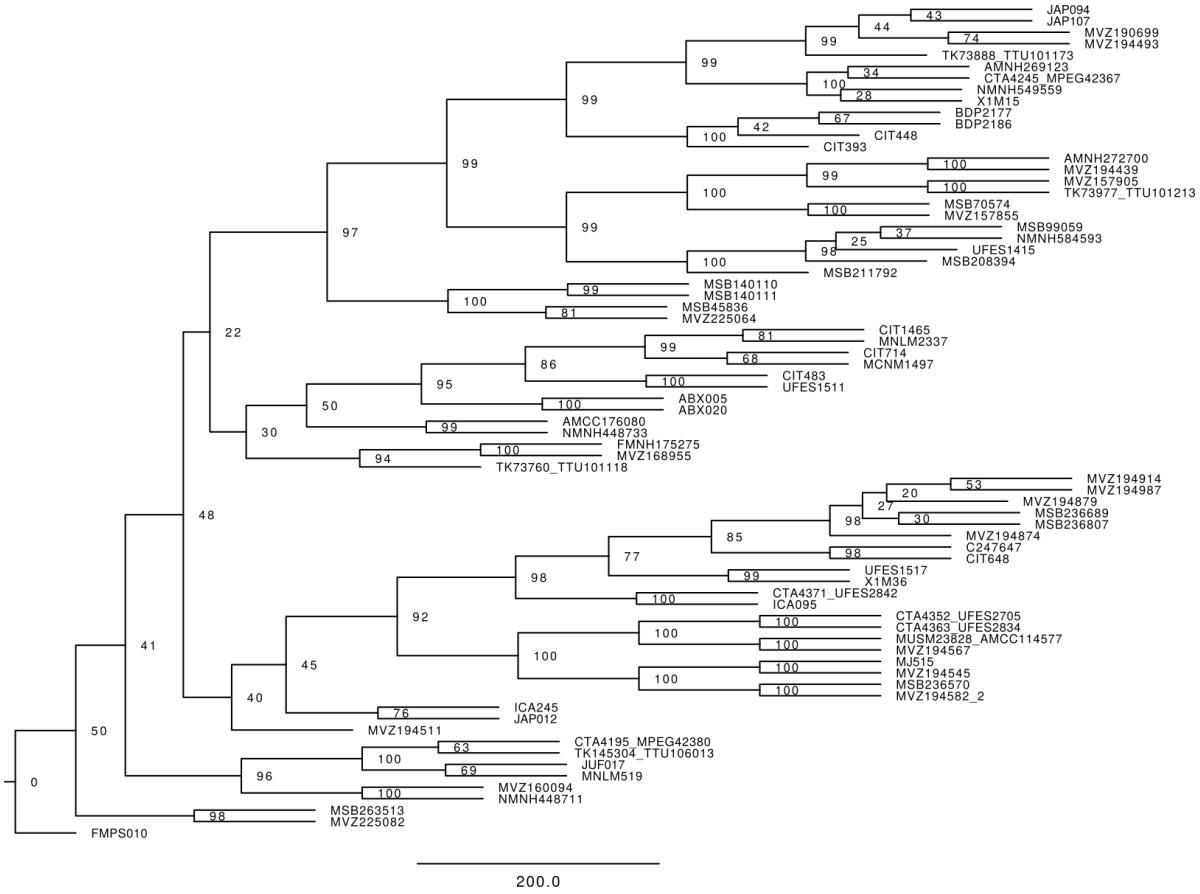


Figure S7: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 2,604 SNPs, and 60% of missing data inferred under the coalescent model (CM) with SVDquartets. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

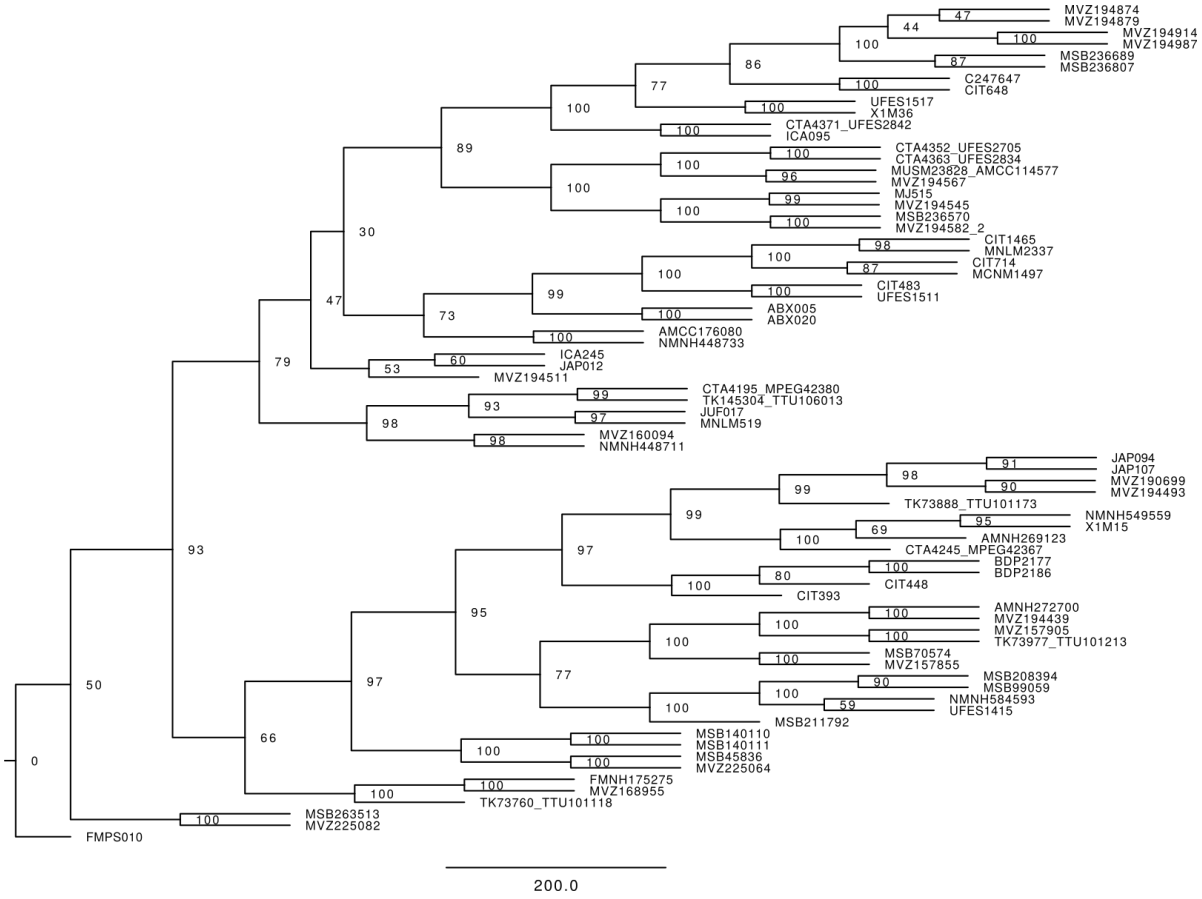


Figure S8: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 6,754 SNPs, and 70% of missing data inferred under the coalescent model (CM) with SVDquartets. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

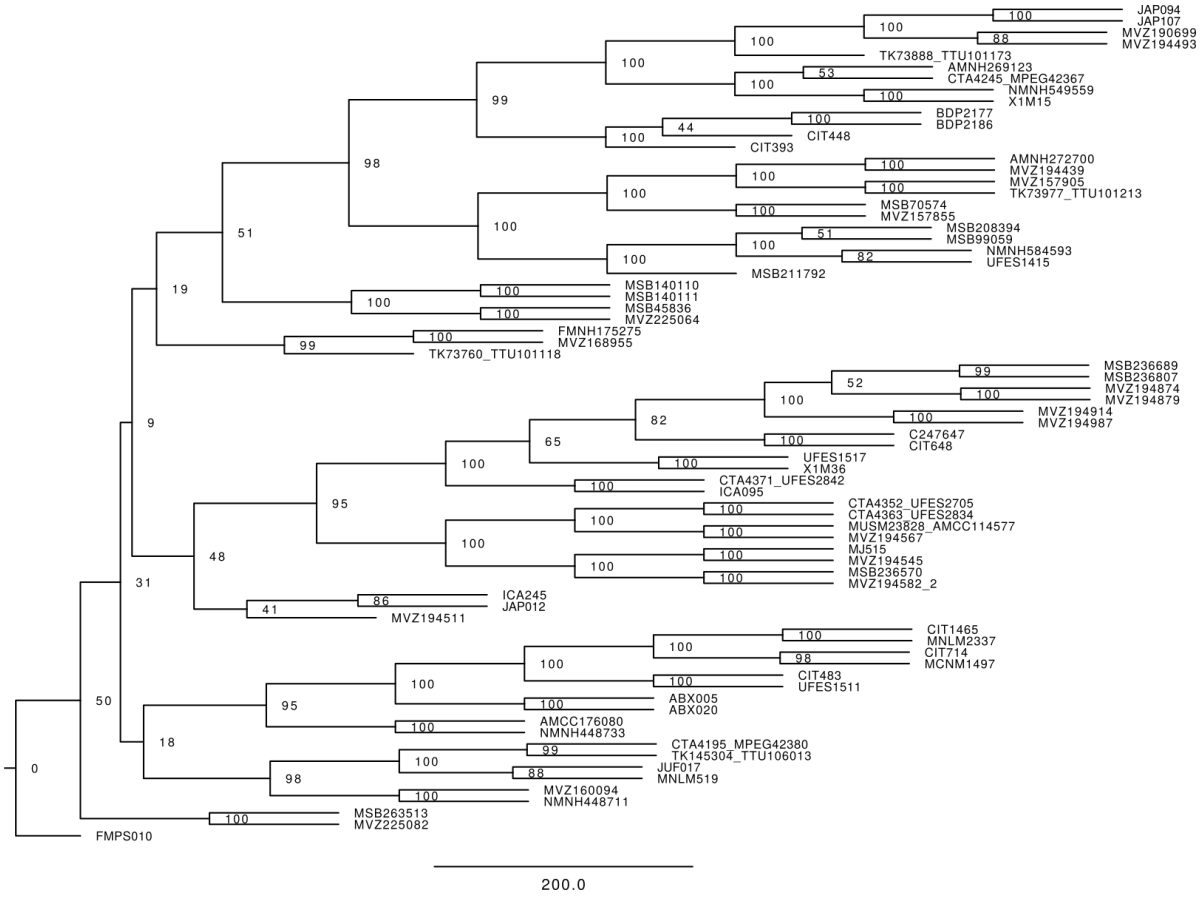


Figure S9: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 15,911 SNPs, and 80% of missing data inferred under the coalescent model (CM) with SVDquartets. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

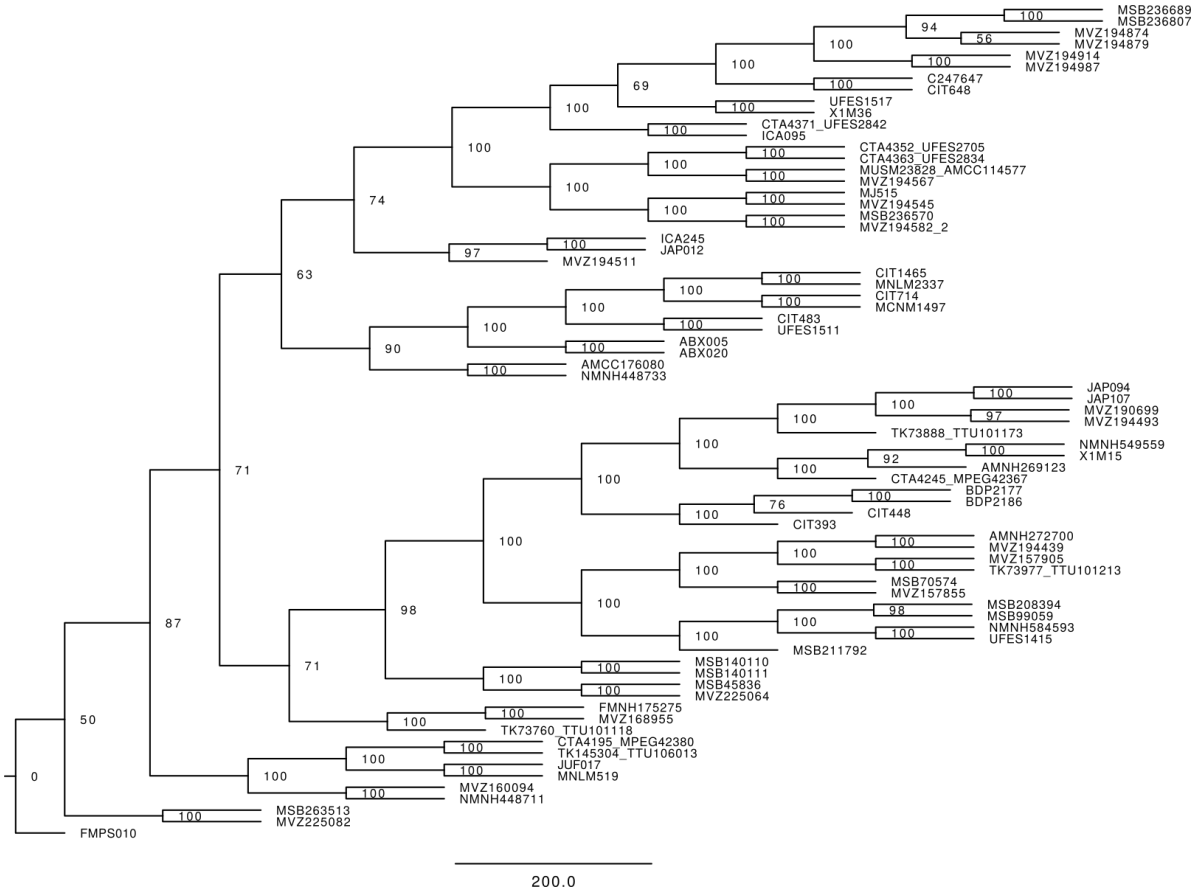


Figure S10: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 59,669 SNPs, and 90% of missing data inferred under the coalescent model (CM) with SVDquartets. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

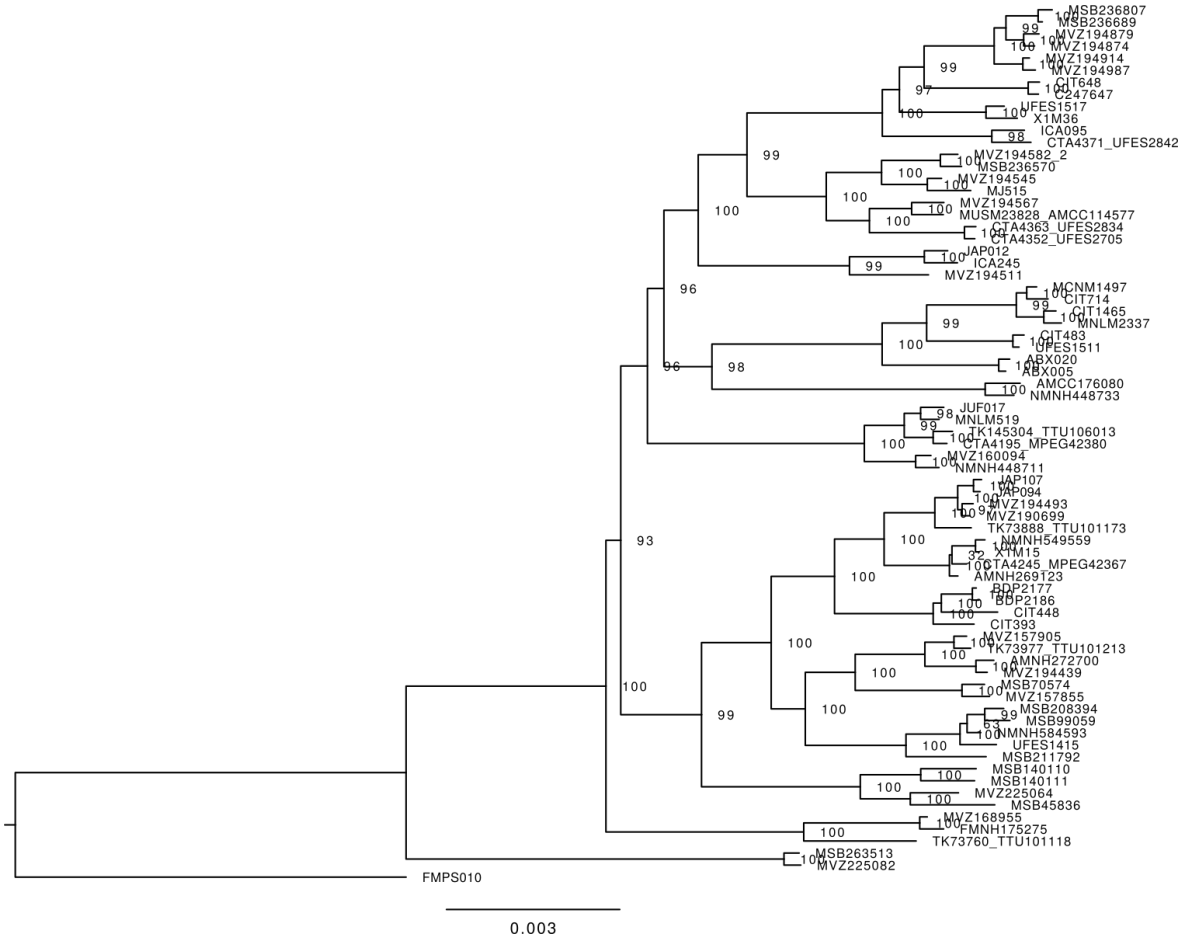


Figure S11: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 741 loci, and 30% of missing data inferred under the Maximum Likelihood (ML) with RaxML. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

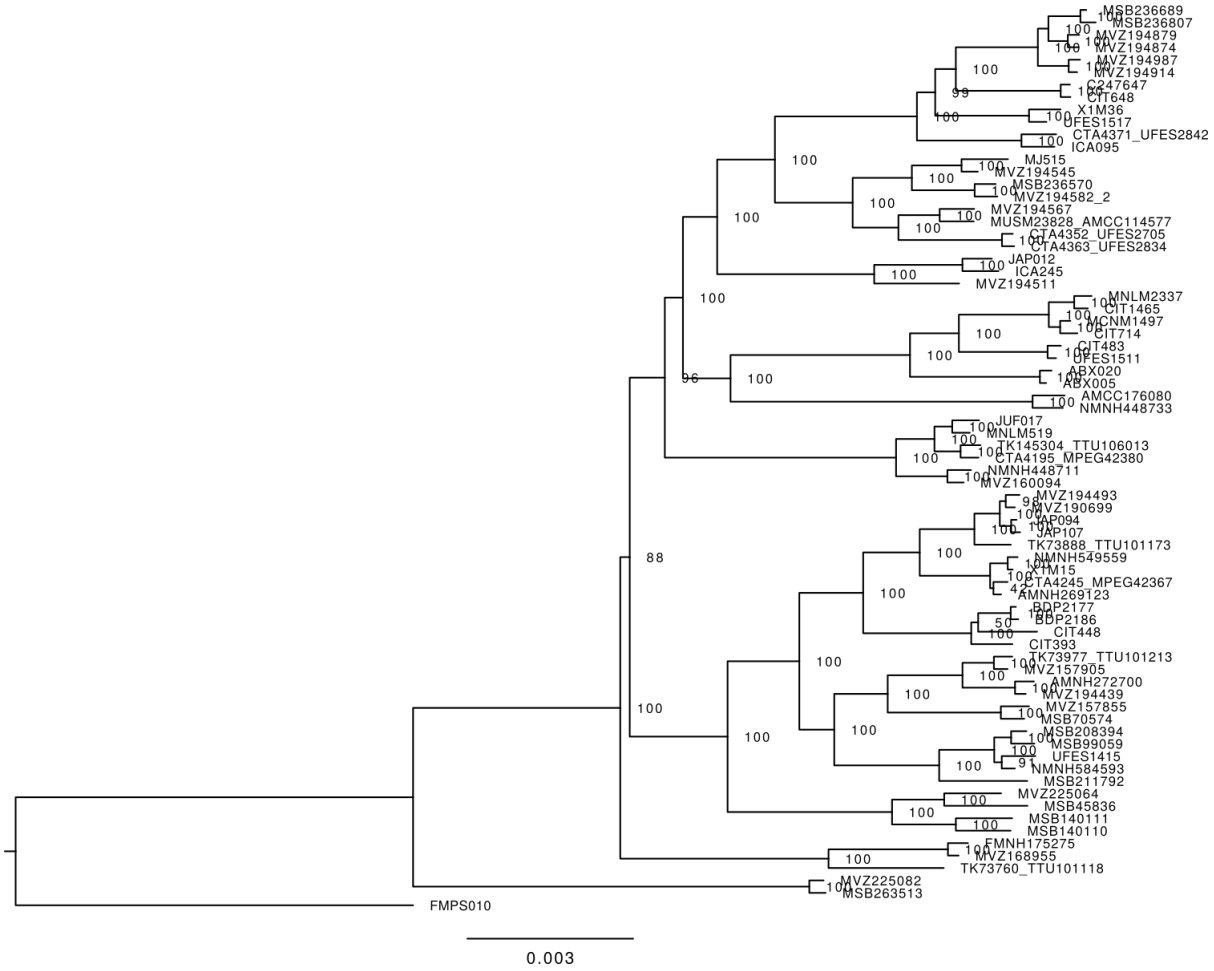


Figure S12: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 1,046 loci, and 40% of missing data inferred under Maximum Likelihood (ML) with RaxML. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

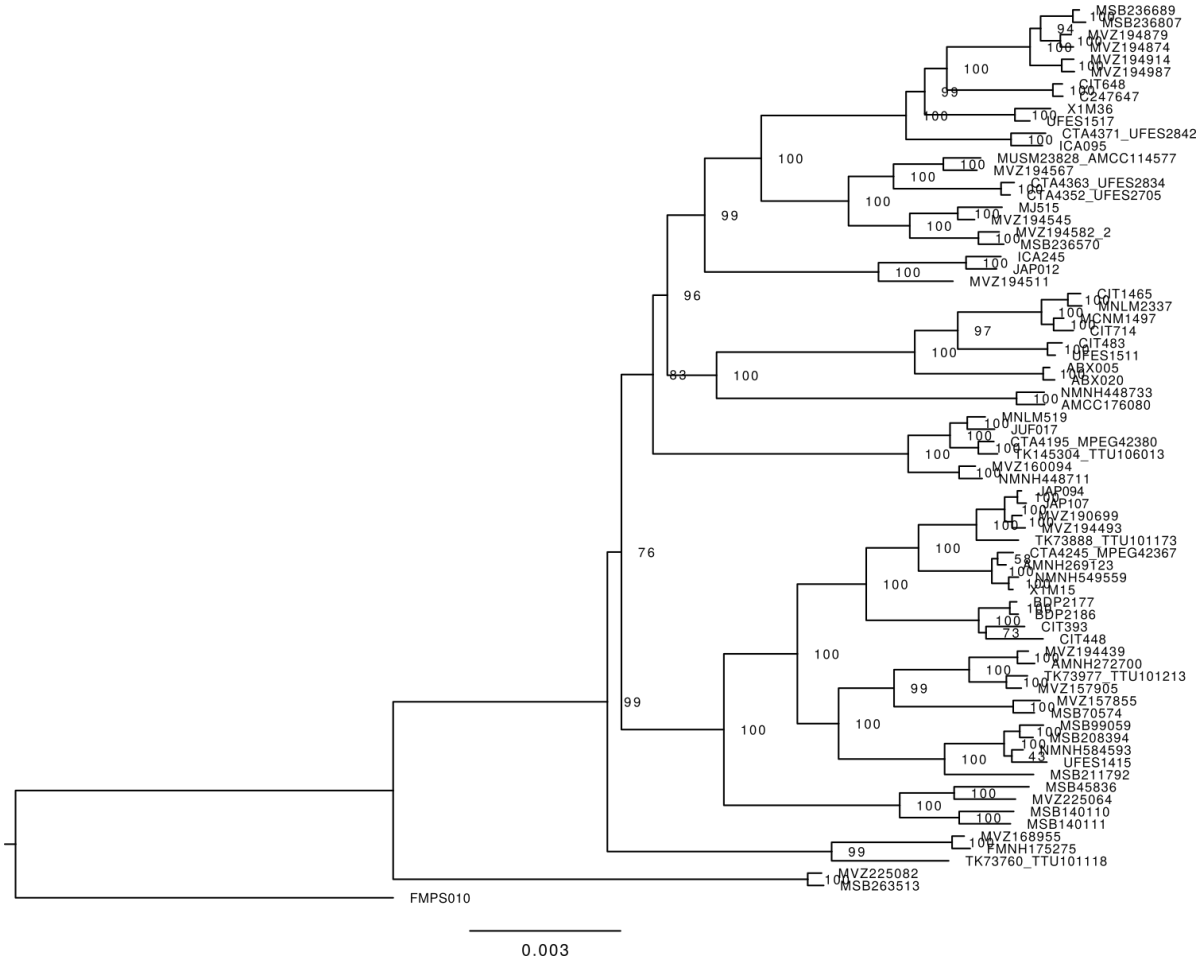


Figure S13: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 1,580 loci, and 50% of missing data inferred under Maximum Likelihood (ML) with RaxML. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

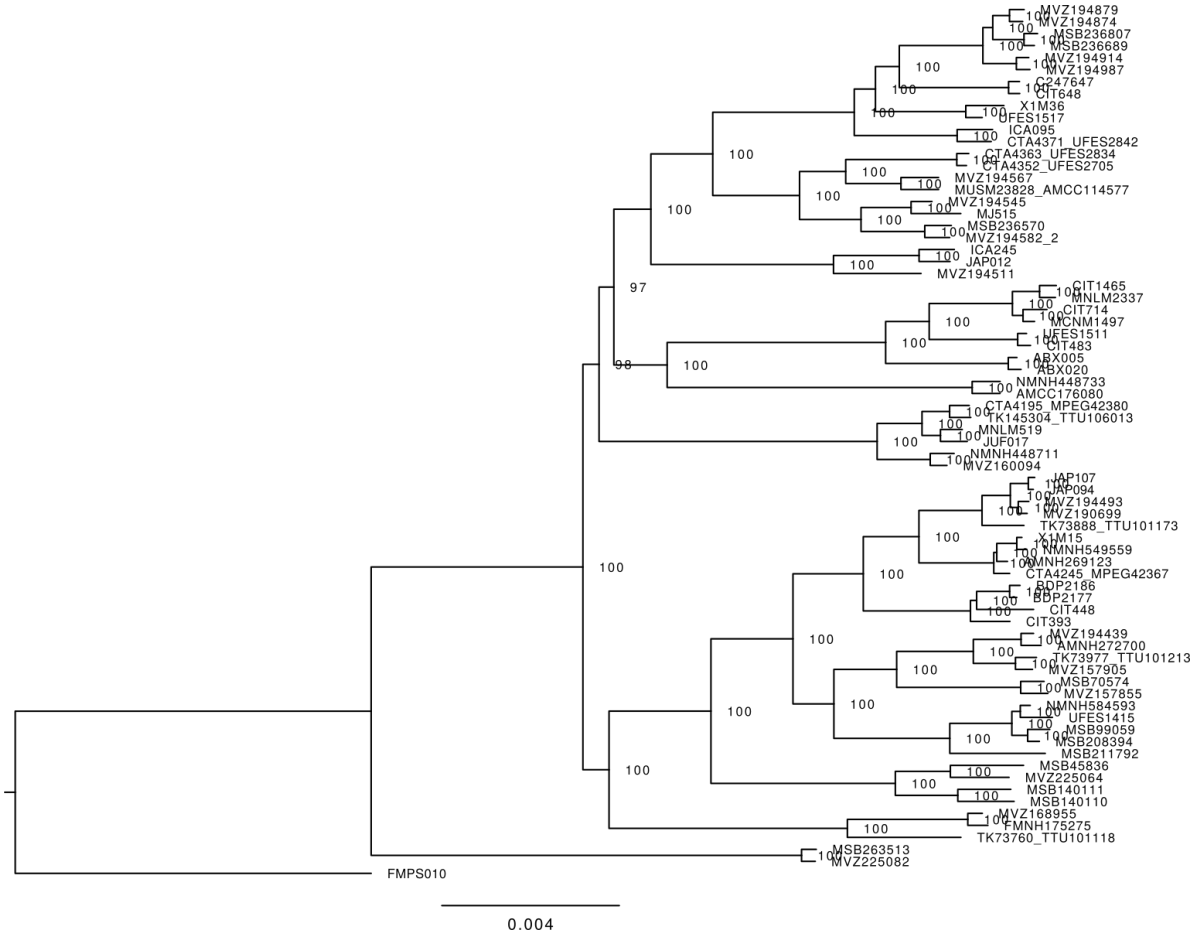


Figure S14: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 2,604 loci, and 60% of missing data inferred under Maximum Likelihood (ML) with RaxML. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

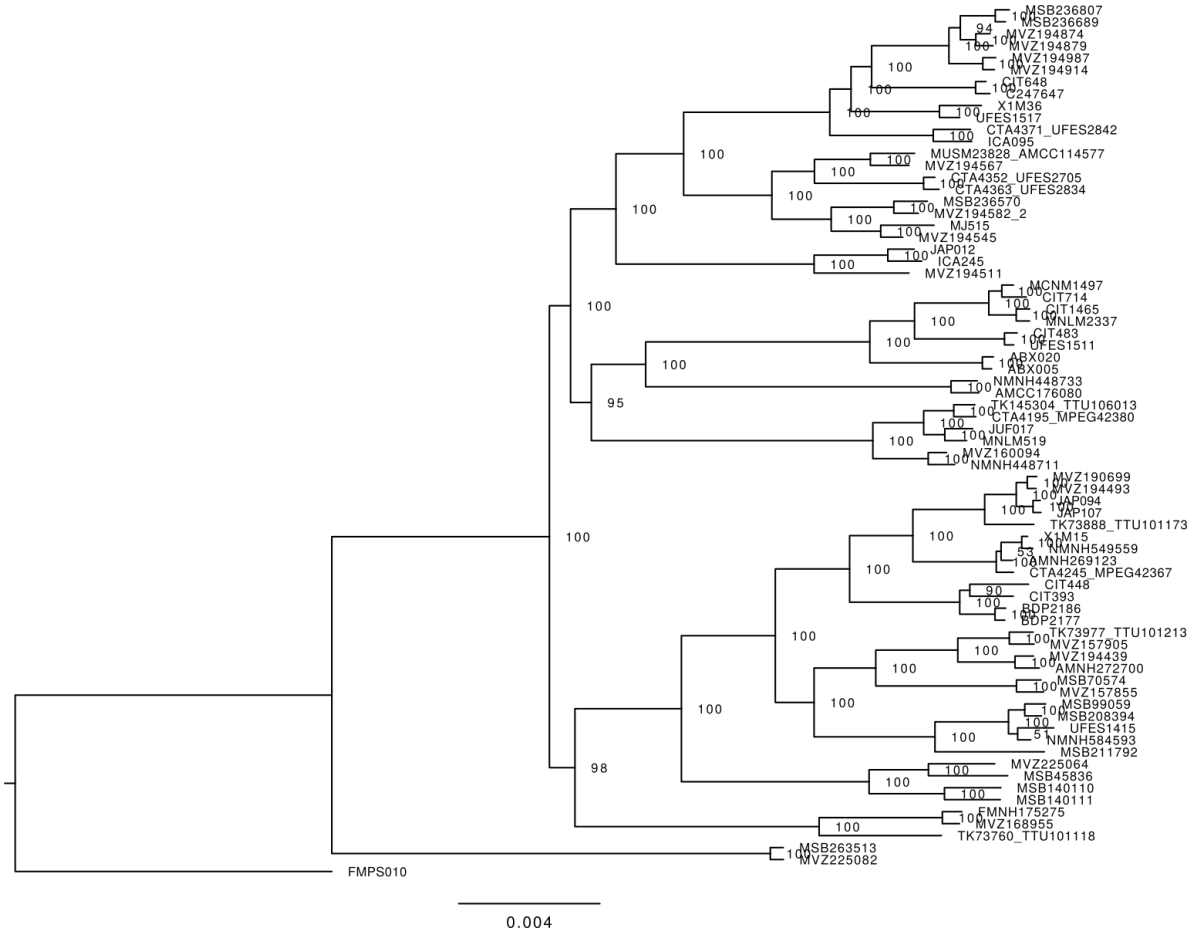


Figure S15: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 6,754 loci, and 70% of missing data inferred under Maximum Likelihood (ML) with RaxML. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

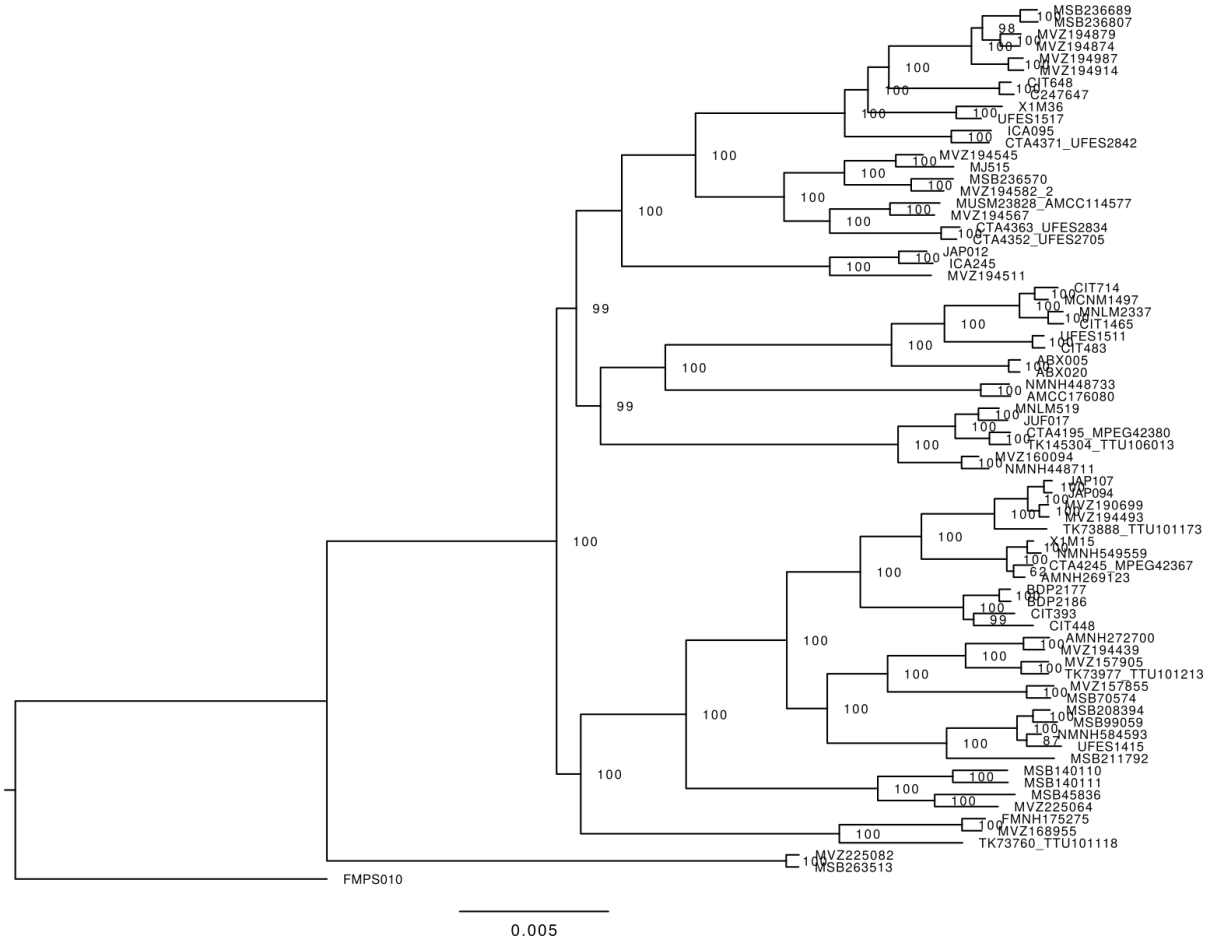


Figure S16: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 15,911 loci, and 80% of missing data inferred under Maximum Likelihood (ML) with RaxML. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

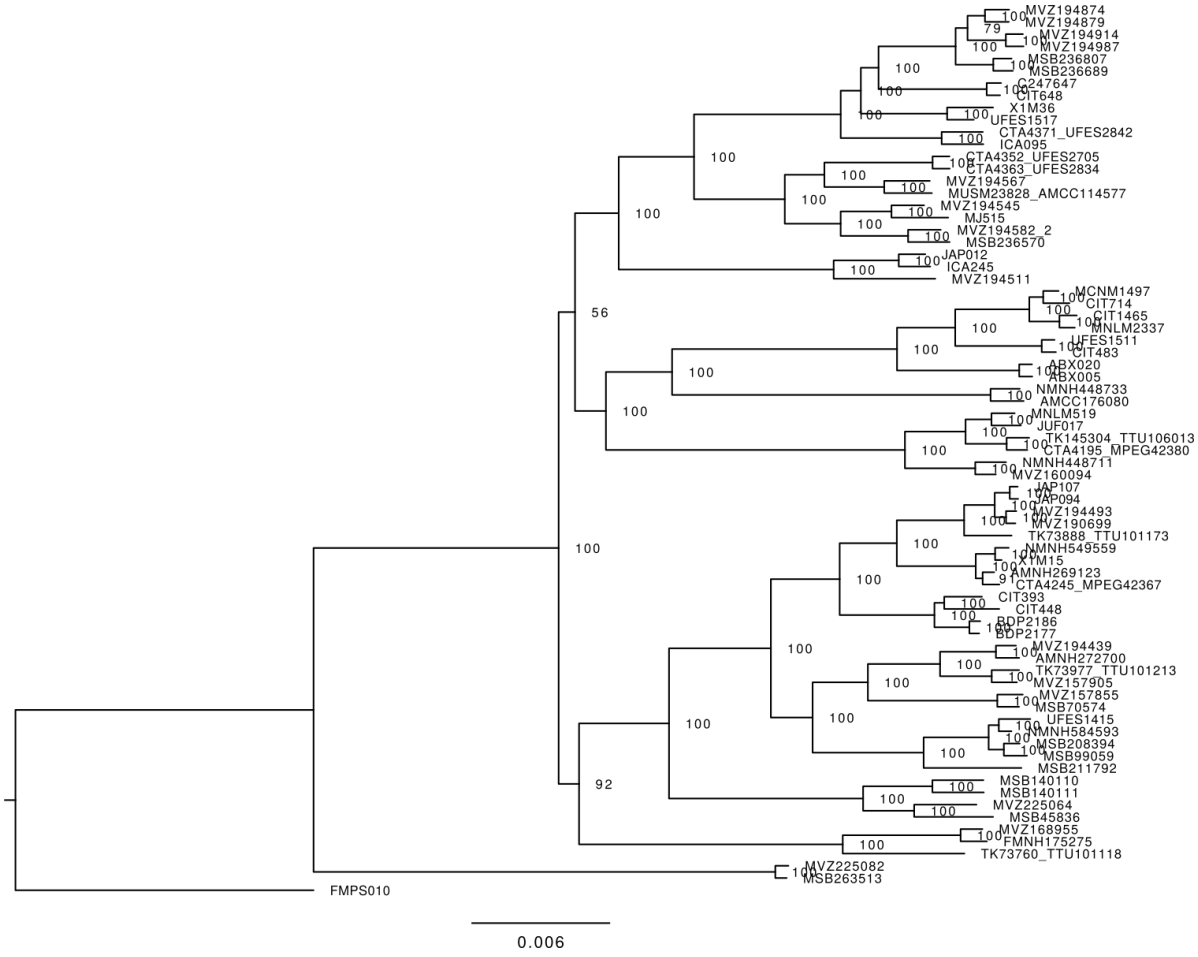


Figure S17: Phylogenetic tree for the genus *Proechimys* based on the strict dataset with 73 individuals, 59,669 loci, and 90% of missing data inferred under Maximum Likelihood (ML) with RaxML. Numbers on the nodes indicate the bootstrap values and information about samples are in the Table S1.

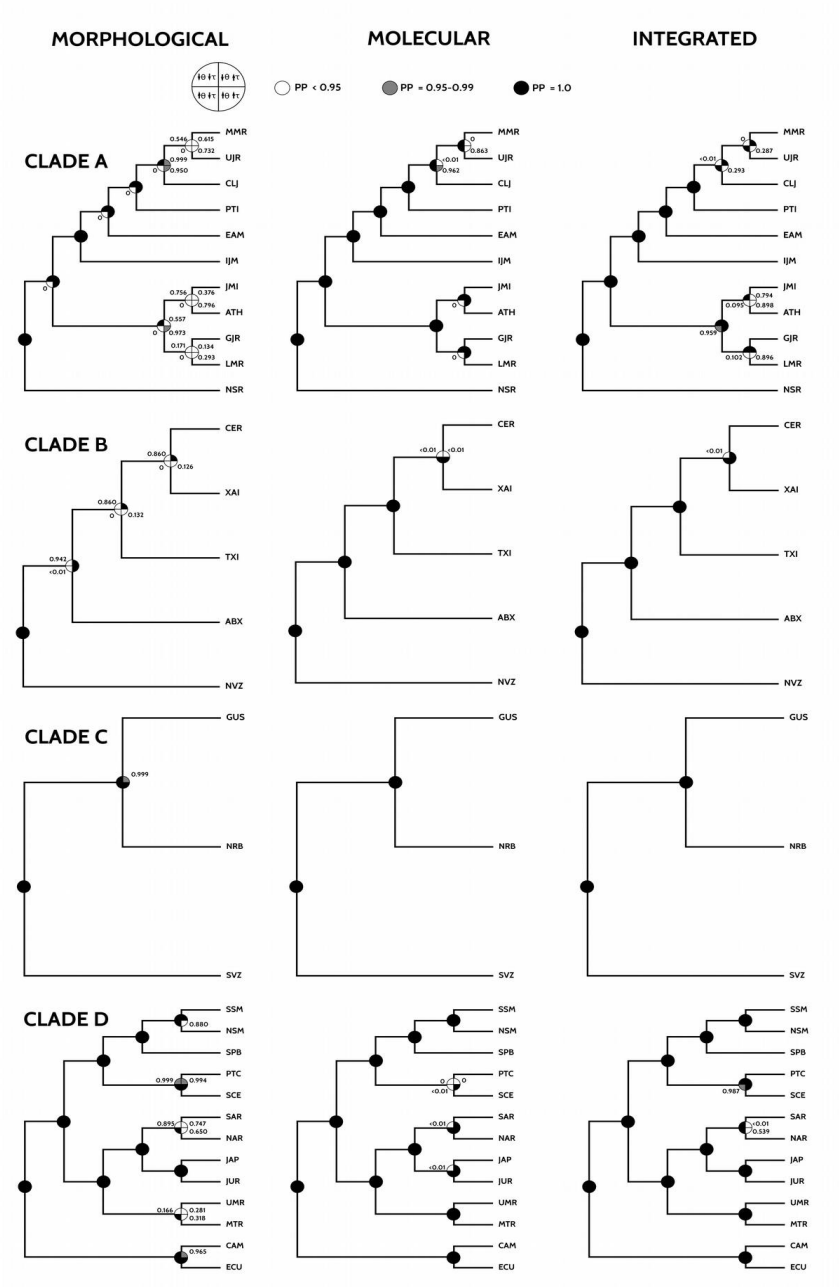


Figure S18: Posterior probabilities (PP) found in simulations for each node of the guide trees used in the Analysis 3 in the iBPP with the four main clades in the genus *Proechimys* (clades A, B, C, and D). Clade E only had one lineage with individuals with more than 500,000 reads, thus it did not enter the analysis (see Material and Methods for details). The acronyms represent lineages found in the conservative tree (see Results and Figures 1 and 2 for the acronym meanings). Circles are divided into four portions, each one representing a combination of θ (ancestral population sizes) and τ (divergence times) values, top left: θ = (1, 10) and τ = (1, 10); top right: θ = G (1, 10) and τ = G (2, 2000); bottom left: θ = G (2, 2000) and τ = G (1, 10); bottom right: θ = G (2, 2000) and τ = G (2, 2000). Columns indicates the datasets used in the simulations.

3. WHEN LANDSCAPE MATTERS: THE ROLE OF THE ANDES UPLIFT AND THE AMAZONIAN LANDSCAPE EVOLUTION IN THE RAPID DIVERSIFICATION OF A SPINY RAT GENUS

Abstract

Landscape evolution has direct implications for the evolutionary history of the local biota. Thus, evolutionary studies of a taxon can also be useful in the study of past geological patterns, bringing more evidences to the evolution of the landscape and to the interactions between environment and biota. *Proechimys* is a spiny rat genus with potential to be a model for studies on the Amazonian landscape evolution, since its 22 valid species currently have a wide distribution throughout the region and also in outside areas, such as, Western Andes, Central America, and Cerrado. Here, I used 82 samples, representing all subclades identified Chapter 2, and 6,000 loci generated from the ddRAD-Seq technique to create a hypothesis on the evolution and biogeography of the species of genus *Proechimys*. I estimated divergence times for the origin of the recovered clades and I tested five models of geographic range evolution to estimate the lineages ancestral areas with two different area sets, representing known areas of endemism and biogeographical dominions. The most recent common ancestor for *Proechimys* was estimated to have originated about 11 Ma (95% CI = 17.01 – 5.98 Ma) in the Miocene, being its ancestral area the Western Amazon. *Proechimys* main clades also presented similar ages for their origins, evidencing an event of rapid diversification in *Proechimys*. Based on these date estimates and clades and species ancestral areas, I hypothesize the biogeographic history for the genus, using the common points in the different biogeographic analyses. The results with the dominions gave more importance to dispersion events while the analyses with areas of endemism, both vicariance and dispersion events were important for the evolution of the species on the genus. The evolutionary history of *Proechimys* agrees with the formation of the Amazon River and the end of the Pebas system, as a consequence of the rise of the Andes Mountains on Miocene. Moreover, main rivers of Amazon basin seem to delineate areas of secondary contact between lineages rather than to have played a role as driver for diversification in the genus.

Keywords: Dispersion; Echimyidae; Miocene; *Proechimys;* Riverine barriers; Vicariance

3.1. Introduction

The Amazon landscape has been influenced by the uplift of the Andes for the last 100 Ma, which has led to changes in river basins, rainfall regimes, climate, and other factors (Sdrolias & Müller, 2006). These changes were accentuated especially during the Miocene,

after the fissure of the Farallon plate into Cocos and Nazca plates around 23 Ma, that increased the plates subduction (Cobbold *et al.*, 2007). Albert et al., (2018) and Hoorn et al., (2010) have made a comprehensive review of the geological changes that led to the current hydrographic patterns in the Amazon. In general the Amazon, especially the Western Amazon, changed from a mega-wetlands system with marine influence during the Miocene, so-called Lake Pebas, to a system with reduced and seasonal flooded areas, the várzea forests (Wesselingh *et al.*, 2001). Another change was in the drainage direction, from north-trending in the Western Amazon water flows during the 20-10 Ma in the Miocene, to east-trending transcontinental Amazon river basin in the present-day (Albert *et al.*, 2018).

In the middle of the Pliocene, the Amazon River basin already presented the current pattern of water flows (Hoorn *et al.*, 2017). Even the Panama land bridge would already be established in the Pliocene (Montes *et al.*, 2015). Thus, the changes in the Amazonian landscape, in general, were due to climatic oscillations in the last 3 Ma (Late Pliocene, Pleistocene, and Holocene). During this period, there was a global cooling trend of the planet with ice age cycles (Zachos *et al.*, 2001). This led mainly to changes in vegetation cover in the region (Haffer, 1969) or to the change in the floristic composition (Colinvaux *et al.*, 2001), affecting more border areas of the biome and the Eastern Amazon than the Western Amazon, especially with regard to precipitation (Cheng *et al.*, 2013).

Evolutionary history, and consequently the taxonomy and classification, of a taxon is intimately connected to the landscape evolution of the region where the taxon inhabits (Hoorn *et al.*, 2010b). Biodiversity studies for an area can bring information and evidence on past geologic and climatic changes, as well as information on interactions between biota and environment (Carnaval *et al.*, 2009; Leite *et al.*, 2016), especially if the studied taxon has a wide distribution throughout the biome.

In this context, the spiny rats of the genus *Proechimys* are a good model species to evaluate the influence of landscape evolution on the Amazonian biota. These rodents of the family Echimyidae are terrestrial, occurring in all Amazonian environments from the seasonal floodplain forests (*várzea* forests) to non-flooded forests (*terra-firme* forests), as well as pristine or more degraded forests (Patton & Leite, 2015). These rodents also occur outside the Amazon basin, such as in northern Venezuela, Central America, and Cerrado (Fabre, Patton, & Leite, 2016). In addition, studies with estimated divergence times showed an old origin of

Proechimys, during the Miocene between 11 – 6 Ma (Fabre *et al.*, 2013, 2017; Álvarez, Arévalo, & Verzi, 2017), when mega-wetlands areas still dominated the landscape (Albert *et al.*, 2018). This taxon encompasses a large geographic area and a lasting geologic history in the continent, favoring it as a candidate for comparisons between biota and landscape evolution in the Amazon.

However, our knowledge on this area is scanty, as there are no published studies for biogeographic patterns for the entire *Proechimys,* but there are some studies dealing with the biogeography of the Family Echimyidae (Fabre *et al.*, 2017), for some species (Da Silva & Patton, 1998), or for some species in specific regions (Patton, Da Silva, & Malcolm, 2000; Van Vuuren *et al.*, 2004; Silva *et al.*, 2018). Fabre et al., (2017) proposed a biogeographic hypothesis where they identified 14 events of dispersion and one vicariance event within the evolutionary history of the family Echimyidae. Following their hypothesis, the genus *Proechimys* would have its diversification after the dispersion of ancestral populations of the sister genus *Hoplomys* to western Andes and consequent isolation by the Northern Andes uplift. In addition, Fabre et al. (2017) also identified the ancestral area for *Proechimys* in the Amazon basin + Guyana Shield, and they suggested that *Proechimys* and another two genera (*Makalata* and *Mesomys*) would represent good models to detail the importance of the orogeny and Andes for the diversification of family Echimyidae.

At the species level, Da Silva & Patton (1998) showed that the main Amazonian rivers limited the distribution of species from *goeldii* group. However, they were not able to hypothesize whether the rivers were the diversification driver or merely the limits of an area of secondary contact. Patton et al., (2000), studying the small mammal communities along the Rio Juruá, indicated that this river was not a barrier to *Proechimys* although it was for other small mammals species. Other studies, with *Proechimys guyannensis* and *Proechimys cuvieri* from Guyana Shield, demonstrated that there were little or no geographical structure within French Guyana (Van Vuuren *et al.*, 2004), and that main rivers in the region did not function as barriers, suggesting that diversification for both taxa was associated to the climatic changes during the Pleistocene and to geological changes in the Rio Amazonas fan (Silva *et al.*, 2018).

A major limitation for such studies is the knowledge on the diversity and distribution of species of the genus, and their phylogenetic relationships (see General Introduction and Chapter 2). Currently *Proechimys* is represented by 22 valid species with several records of

sympatry, especially in the Western Amazon (Patton & Leite, 2015). But recent studies with the genus based on genomic and morphometric data, showed the existence of 28 putative species divided into five main clades, named as clades A, B, C, D and E (Chapter 2). Phylogenetic relationships among the five clades still need further investigation but the authors proposed a phylogeny with good support for them (Chapter 2) and that the instability of the phylogenetic trees in the basal diversification events could be due to a process of rapid diversification that is characteristic of the genera of the Family Echimyidae (Leite & Patton, 2002; Courcelle *et al.*, 2019).

Therefore, based on this new evidence available, coupled with date and ancestral area estimates, I will be able to test the hypothesis that (*i*) dispersal events were the main mechanisms of cladogenesis as suggested by Fabre et al. (2017); (*ii*) Andean orogenesis explained patterns of diversity and distribution for species of the genus; (*iii*) the date estimates for the origin of *Proechimys* and its clades were overlapping, or close together, which could indicate a rapid diversification event; (*iv*) major rivers of Amazon basin played a major role on the diversification of genus.

Thus, my aims were to use the phylogenetic tree available, obtained with the genomic data generated in the Chapter 2 to (*i*) date the diversification events in *Proechimys*, especially at the base of the tree, to verify the hypothesis of rapid diversification; (*ii*) estimate the ancestral areas for the genus and its clades and species, proposing a biogeographic hypothesis for the evolution of *Proechimys*; (*iii*) to compare the biogeographic hypothesis with the landscape evolution and climate change in the Amazon in order to propose the drivers of *Proechimys* diversification.

3.2. Material and Methods

3.2.1. Sampling and phylogenetic tree

I used 82 samples (Table S1), representing the 41 subclades identified in chapter 1, including outgroups that corresponded to five genera of the family Echimyidae: *Clyomys laticeps* (n = 1), *Euryzygomatomys spinosus* (n = 1), *Hoplomys gymnurus* (n = 2) the sister-

genus of *Proechimys, Myocastor coypus* (n = 1), *Thichomys pachyurus* (n = 1), and *Trinomys dimidiatus* (n = 1). Genomic DNA was extracted with DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA), and I built the libraries following Peterson et al. (2012) protocol that were sequenced in three lanes of HiSeq2500 (Illumina, San Diego, CA, USA), according to instructions of the manufacturer to generate 150 base pairs, single-end reads in the Hospital for Sick Children (Toronto, Ontario, Canada). More information on DNA extraction and library preparation is available in the previous chapter (Chapter 2).

I demonstrated previously (Chapter 2) that there are currently five main clades in *Proechimys*, and *Proechimys hoplomyoides* is an independent lineage, probably closer to the genus *Hoplomys* than to other species of *Proechimys*. In addition, I recovered the best tree to represent the phylogenetic relationships among the species and the five clades within the genus through simulations and similarity indexes, using matrices with different amounts of missing data and this tree was employed on the biogeographic analysis of this chapter, on date and ancestral area estimates. More details on simulations, and indices and choice of the best tree are in the Chapter 2.

3.2.2. Divergence times

I randomly selected 6,000 unlinked SNPs, that were present in at least 25% of the samples to date the divergence times among clades, using BEAST 2.5.2 (Bouckaert *et al., 2014),* lognormal relaxed clock, and the Yule model. I followed the best tree topology found in the strict dataset analyses (Chapter 2) to constrain all *Proechimys* clades as monophyletic (not the relationships among the outgroups), because my goal was to date the divergence time and not test the tree topology. I linked the site and clock models across all SNPs, chose GTR+G model with four gamma categories, and estimated the bases rates frequencies and the gamma value. I performed the analysis with 100 million generations, collecting the parameters values every 10,000 generations and trees in every 100,000 generations. I used Tracer 1.7.1 (Rambaut *et al.*, 2018) to verify the MCMC runs convergence upon stable posterior distribution (ESS > 200) for all parameters. I summarized 7,500 stored trees, after 25% of burn in, with the "maximum clade credibility" option, the mean divergence time, and 95% of confidence interval using TreeAnnotator (Bouckaert *et al.*, 2014).

I chose three echimyid rodent fossils for the calibrations, from the list provided by Upham & Patterson (2015). I also followed recommendations from Ho & Phillips (2009), Parham et al. (2012), and Hipsley & Müller (2014) to choose and to set the prior parameters. I selected (*i*) the earliest fossil for the extant genus *Myocastor* (†*Myocastor columnaris*) from the "Mesopotamian" (= "Conglomerado osífero", Late Miocene, 6.8–9.0 Ma; Candela & Noriega, 2004; Upham & Patterson, 2015), setting offset as 6.8 in a lognormal distribution (Mean = 1.0; S = 1.25) for the split between *Myocastor* and other Myocastorini genera (*Thrichomys*, *Hoplomys*, *Proechimys*; Fabre *et al.*, 2017; Courcelle *et al.*, 2019); (*ii*) †*Pampamys,* a sister genus of *Thrichomys* from Late Huayquerian (>6.0 Ma; Verzi, Montalvo, & Deschamps, 2008; Deschamps *et al.*, 2009; Upham & Patterson, 2015), setting the offset as 6.0 for the origin of the *Thrichomys* lineage with a lognormal distribution (Mean = 1.0; S = 1.25); and (*iii*) †*Theridomysops parvulus* the oldest stem taxon to *Euryzygomatomys* + *Clyomys* from Chasicoan-Huayquerian (Late Miocene, 6.8–9.0 Ma; Vucetich, 1995; Upham & Patterson, 2015). Thus, I also set the offset as 6.8 in a lognormal distribution (Mean:1.0; S: 1.25) for the for the origin of the *Euryzygomatomys* + *Clyomys* lineage (Fabre *et al.*, 2017; Courcelle *et al.*, 2019).

3.2.3. Ancestral area estimation

We estimated the ancestral areas for all *Proechimys* clades under different biogeographic scenarios. For this, we run these analysis on the "BioGeoBEARS" package (Matzke, 2013) implemented in the R platform 3.4.4 (R-Development CoreTeam, 2018), employing the tree with date estimates with OTU representing the named clades based in the conservative tree, and a presence/absence matrix with information about the OTUs geographic occurrence. We tested two set of geographical areas, the biogeographic areas defined by Morrone (2014) and the areas of endemism established by Cracraft (1985). Species of genus *Proechimys* occur in seven Neotropical dominions (DOM) from Morrone (2014), namely: (1) MES – Mesoamerican dominion; (2) PAC – Pacific dominion; (3) BBR – Boreal Brazilian dominion; (4) SBR – South Brazilian dominion; (5) SEA – South-eastern Amazonian dominion; (6) CHA – Chaco dominion, and (7) PAR – Parana dominion (Fig. S1). These species also inhabit nine classical areas of endemism (AOE) for the Amazon

region (Cracraft, 1985; da Silva, Novaes, & Oren, 2002; Borges & Da Silva, 2012): (1) BEL – Belem; (2) GUY – Guyana; (3) IME – Imeri; (4) INA – Inambari; (5) JAU – Jau; (6) NAP – Napo; (7) RON – Rondonia; (8) TAP – Tapajos; (9) XGU – Xingu and I categorized all areas outside the Amazon basin into two large areas to the (10) NOR – North, and to the (11) SOU – South (Fig. S2). I performed two analyses with these two areas set, in the first one I maintained all states (ranges) combinations among areas, and in the second one I used only the states combinations that represented adjacent areas. None of the *Proechimys* named clades used as OTU occupied more than 3 areas (in DOM analysis) and 4 areas (in AOE analysis), thus I used these numbers as maximum states allowed in the biogeographic analyses. For the biogeographic hypothesis, I only considered one area as occupied by ancestral populations if it reached at least 20% probability of occupancy.

I compared three range evolution models through Maximum Likelihood (ML): DEC [Dispersion-Extinction-Cladogenesis; (Ree & Smith, 2008)], DIVALIKE [the ML version of Dispersal-Vicariance Analysis based on parsimony, (Ronquist, 1997)], and BAYAREALIKE [based on ML of the BayArea program (Landis *et al.*, 2013)]. These models may evaluate different biogeographic hypotheses because they take into consideration differently the dispersal, vicariance and extinction events. We also tested the DIVALIKE and BAYAREALIKE with founder-event speciation model (+J), which test whether a new species jumped, during cladogenesis, to a different area than the ancestral one (Matzke, 2014). AICwt was used for the statistical comparison between the five models (Wagenmakers & Farrell, 2004), and therefore the DEC-J model was not tested because it presented statistical problems when compared with other models using AIC (Ree & Sanmartín, 2018).

3.3. Results

3.3.1. Divergence times

Information about the processed reads can be assessed in Table S2. I constrained all *Proechimys* clades and subclades as monophyletic, according to the best tree resolution and values of bootstrap found in Chapter 2. Phylogenetic relationships among *Proechimys*

hoplomyoides (TEP), *Hoplomys gymnurus* (HG) and other *Proechimys* specimens (PRO) were not forced, and in this analysis TEP was recovered as the sister group of HG and both groups were sister to PRO (Fig. 1; Node 2).

Proechimys origin probably occurred in the Late Miocene (Mean = 11 Ma; 95% C.I. = 17.01 – 5.98 Ma; Table 1; Fig. 1; Node 4), as well as the origin of the five main clades (Table 1; Fig. 1; Nodes 4, 5, 6 and 10), and their mean (10.53 – 10.14 Ma), median (10.07 – 9.69 Ma) and confidence intervals values of date estimates (16.62 – 5.20 Ma) overlapped almost completely. Divergences between *H. gymnurus*, *P. hoplomyoides* and the other *Proechimys* specimens (PRO) dated from the Middle Miocene (Mean = 12.68 Ma; 95% C.I. = 19.94 – 7.05 Ma; Table 1; Fig. 1; Nodes 1) but the confidence interval of divergence between *H. gymnurus*, and *P. hoplomyoides* is very broad and reached the Pliocene as well (Mean = 8.16 Ma; 95% C.I. = 14.87 – 2.99 Ma; Table 1; Fig. 1; Node 2). During the Pliocene occurred the diversification inside the five main clades in *Proechimys*, and also the origin of the clade formed by specimens from Western Andes (Mean = 5.50 Ma; 95% C.I. = 9.52 – 2.44 Ma; Table 1; Fig. 1; Node 13). We found that most subclades origins dated from the Pliocene-Pleistocene boundary or Pleistocene, as the diversification of the várzea forest subclades (Mean = 1.33 Ma; 95% C.I. = 2.12 – 0.64 Ma; Table 1; Fig. S1; Node 17). The mean values of the origin of the three lineages that currently inhabit the Cerrado biome were different but the confidence interval overlapped, and I cannot rule out that the occupancy of Cerrado biome were concomitant events (Table 1; Fig. 1; Nodes 14-16). For raw date estimates in the phylogeny see Table S3.

3.3.2. Ancestral areas estimation

Analyses using only the adjacent areas have obtained the same results as the regular analyses, thus their results will not be shown here. The best biogeographic model was DIVALIKE + J (Table 2), independently of the set of areas (DOM: LnL = -106.9; AICwt = 0.40; AOE: LnL = -140.2; AICwt = 0.29). Probabilities of occupancy per areas by the ancestral populations, organized by nodes are in the Table S4, for the DOM areas and in the Table S5 for AOE areas.

The range evolution for the genus *Proechimys* using the DOM areas (Fig. 2) presented 64 states, 16 dispersal and 2 vicariance events (Table 3), while with AOE areas (Fig. 3) the analysis showed 330 states, a similar number of dispersal events (n = 17; Table 4) but much more vicariant events (n = 14; Table 4). On DOM analysis, ancestral populations of *Proechimys* probably occupied the SBR dominion (97% of probability of occupancy; Table 3; Fig. 2), the same ancestral area for the clade A (99%), clade B (90%), clade D (97%), and clade E (100%), while the ancestral population of clade C occupied the BBR dominion (100%) (Fig. 2). On the other hand, the AOE analysis showed that the origin of *Proechimys* could have happened in areas to the north of South America (NOR) (36%), or in the area of endemism Guyana (28%) or in Inambari (24%) in Western Amazon (Table 4; Fig. 3). For the five main clades, the ancestral populations of clade A presented a 96% occupancy probability for Inambari area, clade B 59% for the north of South America, clade C 100% for the Guyana, clade D 46% for the north of South America and 35% for Inambari area, and clade E 81% of the occupancy probability for the Napo area (Fig. 3; Table 4).

For DOM analysis, diversification and origin of the five main clades of *Proechimys* occurred in the Western Amazon. After a dispersal event from Western Amazon to the northern region of South America (D8, Fig. 2), Clade C had its entire biogeographical history in the northern bank of Rio Solimões-Amazonas (BBR), an extensive area including coastal Venezuela, Guyana Shield, Oriental Amazon, and coastal northeastern Brazil regions. Clade E lineages remained in the Western Amazon, with no changes in the ancestral range. The biogeographic history of the other three main clades (A, B and D) was more complex. Clade A diversification took place in Western Amazon and only recently there were four dispersal events for new areas: two events to northern South America (BBR area) (D13; D14, Fig. 2), one in the Pliocene and other in the Pleistocene, and two other events for the eastern South America (SEA and CHA areas) (D15; D16, Fig. 2) also in the Pliocene and Pleistocene (Table 3). Clade B ancestral range was also in the Western Amazon, and during its diversification only one lineage remained in the area (ABX), while other lineages, through four dispersal events, occupied the northwestern Venezuela (PAC area; D9), southeastern Amazon (SEA area; D10) and the forested areas of the dry diagonal in South America in the Pliocene and Pleistocene (CHA area; D11 and D12). In the Late Miocene (8.73 – 5.5 Ma) there was a dispersal event from Western Amazon to the Western Andes (D2) in the Clade D, and a posterior event D3 to Central America (D3; 3.96 – 1.64 Ma; Table 3). In addition, there were

three latitudinal dispersions at similar times (D5, D6, and D7), and other one to southeastern South America (D4; 1.02 – 0.64) in the Pleistocene. The only vicariance event in the Clade D was between the SEA and BBR areas in the Pleistocene, around 0.9 Ma (Table 3). The Cerrado biome occupation occurred by three different dispersal events, in three different species: *P.* aff. *steerei* (PTI lineage, D16; Table 3), *P. longicaudatus* (SCE lineage, D4; Table 3) dispersed from Western Amazon, and *P. roberti* (CER lineage, D12; Table 3) from Eastern Amazon.

In the AOE analysis, the ancestral populations of *Proechimys* occupied the Inambari, Guyana and Northern areas of South America (NOR). Then, they dispersed Napo area (D3; Fig. 3; Table 4). Clade E had been originated from populations from Napo and Inambari areas, and a posterior vicariant event (V6; Fig. 3; Table 4) separated the populations into lineages WAN and STQ in the Pleistocene. Ancestral populations of the Clade D occupied the NOR and Inambari areas, and a vicariance event (V3) separated the populations from these two areas in the Miocene (Fig. 3; Table 4). All posterior diversification events in clade D happened in the lineages from Inambari area in the Western Amazon. One branch of these lineages remained within Inambari area, with three latitudinal (north-south direction) dispersal events (D5, D4, and D9), all at similar time period (3.28 – 0.74 Ma; Table 4). The other branch had vicariant (V2, V4, V7, and V8) as well as dispersal events (D7, D8, D9, and D11) in longitudinal (east-west direction), along with the most recent vicariant event also latitudinally. Ancestral populations of *Proechimys* have already occupied the Guyana area, and a vicariance event (V9) isolated the Clade C from NOR and Inambari areas where the ancestral populations of clades A and B probably inhabited, and later clade C populations occupied Imeri area in a recent dispersal event (D12; Fig. 3; Table 4). NOR and Inambari areas were isolated by V10 event in the Miocene (Fig. 3; Table 4), and the populations in the northern South America originated the clade B populations. Some populations remained in the north region, evolving to *P. guairae* and the other populations through dispersal event (D17; Fig. 3; Table 4) reached the south of Amazon River (Rondônia, Tapajós and SOU areas). Dispersal (D10) and vicariance (V12, V13, and V14) events led to the diversification and occupation of areas into southeastern South America, such as, Cerrado biome. Clade A had the same biogeographical history as in the DOM analysis. Ancestral populations originated in the Western Amazon (Inambari area) after the V10 event and with four dispersal events (D13, D14, D15, and D16; Fig. 3; Table 4), the clade A reached the current range distribution in the

eastern and northern regions of South America. The AOE results for the Cerrado occupation was similar to the DOM analysis: two species dispersed from the Western Amazon, *P.* aff. *steerei* (D16; Table 4) and *P. longicaudatus* (D4; Table 4) in an older event tham the one indicated by the DOM analysis. However, the main difference in the occupation was in *P. roberti* from the Eastern Amazon. Ancestral populations of *P. roberti* would have occupy the Cerrado areas (SOU) since the D17 dispersal event (Fig. 3; Table 4), and only by one vicariance event (V14; Fig. 3; Table 4) there was differentiation in different species between Belem + SOU areas and Xingu area.

Besides the differences in number of vicariance events between the two biogeographic scenarios, there are 10 equal events recovered in both analyses: D5, D6, D10, D13, D14, D15, D16, V1, V2, and the D12 event recovered in DOM analysis was equivalent to the V14 in AOE analysis (Fig. 2–3; Table 3–4). Since I considered only areas with probability of occupancy greater than 20%, some ancestral ranges resulting from the AOE analysis were formed by non-adjacent areas (Table 4). These areas that would allow the connection between the non-adjacent range were discarded because they had less than 20% of occupancy probability. In these cases, we discarded these ancestral ranges involving non-adjacent ranges in the biogeographic reconstruction presented in the Discussion.

3.4. Discussion

Divergence time estimates supported the hypothesis of rapid diversification within the genus *Proechimys*, since the origin of the genus and the initial divergence events (Fig. 1; Nodes 4, 5, 6, and 10) of its five main clades presented similar ages with great overlapping of the confidence intervals. Rapid diversification events has already been indicated as the reason for low resolution in the basal phylogenetic relationships among the genera of Echimyidae, leading to difficulties in establishing subfamilies, tribes and others categories (Lara, Patton, & Da Silva, 1996; Leite & Patton, 2002; Fabre *et al.*, 2017) until recently when Courcelle *et al.* (2019) inferred the first well-supported phylogeny for Echimyidae through exon capture technique.

Some vicariant events in both analyses, for example V2, and V6, V12, V13 or V14 (Fig. 3) in AOE analyses, could be explained by presence of some of the main Amazonian

rivers. Amazonas River (V2 between lineages of *P. cuvieri* SAR and NAR), Marañón River (V6 between *P. simonsi* lineages WAM and STQ), Tapajós River (V12 between *P.* aff. *roberti* 2 lineage ABX and TXI+LRX+XAI+CER lineages), Xingu River (V13 between *P. roberti* lineages XAI and CER, and *P.* aff. *roberti* 1 lineages TXI and LRX), and Araguaia River (V14 between *P. roberti* lineages XAI and CER) seem to have been important on the diversification of the species of *Proechimys*, especially in the Clades B and D. However, the vicariant events and origin of above mentioned lineages did not match the origin of the Amazonian rivers (Table 1; Table S3), since the modern Amazon basin was well established since 4 Ma (Hoorn *et al.*, 2010a; Albert *et al.*, 2018), and the groups are younger than this period. The only vicariant event that is old enough to be influenced by the rivers was the V12 by Tapajós Rivers (5.50 – 1.58 Ma; Table 4). However, this and other events, with the exception of the V2, was only recovered in the AOE analyses. Thus, it is more plausible that the rivers worked as secondary contact areas than drivers for diversification events, according to the Da Silva & Patton (1998) hypotheses.

Another possibility, still considering the river as a driver for diversification, is not considering the origin or establishment of the riverbed but rather the dynamics of the river and the amount of water drained by it. The riverbed may have been established millions of years ago, even for meandering rivers (Hoorn *et al.*, 2017) but the water volume and consequently its width and depth may vary greatly during geological time (Clapperton, 1993; Miall, 2002). Thus, even if the estimated dates cladogenesis between two groups on opposite river banks did not match to the river origin, the river could be could be a driver if the estimated dates were coincident with the increase of precipitation and water volume, turning the river an impermeable barrier, not only among different taxa but also at different historical moments for a same taxon.

DOM analysis showed a simpler biogeographical history when compared to the AOE analysis, with dispersion events playing a major role than vicariance, similar to the hypothesis established by Fabre et al. (2017), perhaps due to the smaller number of areas on the former approach, as well as the larger size of these areas. In the AOE analysis, the vicariance events had the same importance as the dispersal ones. In spite of the larger number of areas, the clades A, B and E had similar biogeographical histories in both analysis in number and type of events, whereas in the clade B and D the differences were due to larger number of vicariance events in AOE analysis.

DOM analysis indicated that ancestral populations of *H. gymnunus* (HG), *P. hoplomyoides* (TEP) and strict *Proechimys* (PRO) occupied the Western Amazon (SBR). From there they disperse to the north of South America and Central America, (D1) and then a vicariance event (V1) isolated the populations of the (BBR) in the Tepuis and the populations Western Andes (PAC) that gave origin to *P. hoplomyoides* and *H. gymnunus* respectively. Strict *Proechimys* ancestral populations remained in the Western Amazon. However, the AOE analyzes suggested that ancestral populations of these three lineages occupied in northern South America and Guyana area, and there was a dispersal event to the Western Amazonia (Inambari area) that gave origin to the strict *Proechimys.* The same V1 event occurred for the origin of *P. hoplomyoides* and *H. gymnunus.*

It is possible, however, to rule out this northern hypothesis for the origin of the three lineages by geological data. During the Middle Miocene the Amazon Basin shape was different, assembling in the same region the Pan-Amazon, the current basins of the Orinoco, Magdalena and northern Paraná rivers (Lundberg *et al.*, 1998). Moreover, during this period the Pebas system was the dominant landscape feature, a large wetland area widespread throughout the Western Amazon (Hoorn *et al.*, 2010a). Biological data indicated that there were connections for aquatic biota among through these wetlands on Western Amazon, Tropical Andes, and Guyana Shield but that this route was obstructed to terrestrial organisms (Wesselingh, 2006). Thus, there were no evidences of connections between Western Amazon and the northern region during the origin of *Proechimys*, only between Western Amazon and central region of South America through Paraguay basin.

DOM and AOE analyzes showed the origin of the genus *Proechimys* associated to the Western Amazon (SBR or Inambari areas, respectively) during the Middle Miocene, between 12.68 – 11.00 Ma (95% C. I. around 20 – 6 Ma), similarly to other studies (Upham & Patterson, 2015; Álvarez *et al.*, 2017). The most recent common ancestor (MRCA) for the five main clades of *Proechimys* coincided with the ending of the Middle Miocene Disruption (MMD), a global cooling period around 14 Ma, which led to a peak of extinctions in terrestrial and aquatic life forms (Zachos *et al.*, 2001; Lewis *et al.*, 2008). Considering the probability of vacant niches left after the MMD, the diversification rates could have increased for all organisms, and consequently for the ancestral populations of *Proechimys* (Kawata, 2002; Lekevičius, 2009).

As the current monsoonal climate was already established in the Western Amazon during the Middle Miocene, around 16 Ma (Kaandorp *et al.*, 2005), the main drivers of *Proechimys* diversification during this period may have been the uplift of the Andes and its consequences on the establishment of the drainage systems in the Amazon basin (Hoorn *et al.*, 2010b). In the Middle Miocene, the basins of the Solimões and Amazonas rivers were separated by the Purus Arch, and the waters in Western Amazon were drained to the north, in a paleo-Orinoco basin that flowed into the Caribbean Sea (Albert *et al.*, 2018). Northern Andes did not reach their current altitude, but they were already high enough to be a topographic barrier to the air masses, as indicated by the formation of the basins with Andean sediments in the region (Mora *et al.*, 2010). Although Western Amazon was dominated by wetlands in this period, there were also forests, even fragmented (Hoorn, 1993; Pons & De Franceschi, 2007). Currently, *Proechimys* species inhabit both *várzea* and *terra-firme* forests in the Western Amazon (Patton *et al.*, 2000), so there are no major impediments for its ancestral populations to occupy the Pebas system, even without developing phenotypical adaptations to flooded environments.

The MRCA and subsequent diversification of the five main clades occurred around 10 – 8 Ma (95% C. I. = 15 – 4 Ma) in the Western Amazon, according to the DOM analysis, or in Western Amazon and other adjacent areas in AOE analysis. During this period occurred the major change in the drainage system flow in the Amazon basin, which passes from the south-north (to the Caribbean Sea) to the west-east (to the Atlantic Ocean) pathway, with the current flow of the Amazon River settled down around 7 Ma (Hoorn *et al.*, 2010b; Mora *et al.*, 2010). As a consequence, the Pebas system and wetlands declined, allowing the evolution and diversification of terrestrial biota (Wesselingh & Salo, 2006), including the diversification of forests in Western Amazon and in sub-Andean region (Hoorn et al 2010b), and most likely of the current species of *Proechimys*, as they are typical inhabitants of these habitats (Patton & Leite, 2015). Thus, the diversification of main clades in *Proechimys* could have followed the expansion of non-flooded environments in Western Amazon. Albeit the lineages in clade C have diversified between 4 – 1 Ma, the split among clades A + B and C was in Late Miocene, contemporary to these major changes in Amazon basin, and perhaps these events have isolated clade C from the Western Amazon clades, in the Guyana Shield. The same reasoning can be applied to the clade E, with its diversification occurring in the Pliocene, but its split from clade D may have happened before, in the Late Miocene.

Drainage flow changes in the Amazon basin occurred due to the uplift of the Vaupés Arch or Swell, around 10 – 8 Ma, by tectonic activity (Jaramillo et al., 2010; Albert et al., 2018). Vaupés Arch blocked most of the rivers that flowed to the north, into the Caribbean Sea, dividing the Pan-Amazon in the Orinoco and Western Amazon basins (Hoorn *et al.*, 2010a). Consequently, the Andean sediments were accumulated in the Western Amazon, creating sedimentary basins, silting the region enough to breach Purus Arch, and to connect the Solimões and Amazonas basins (Albert *et al.*, 2018). Hoorn et al. (2017) found Andean sediments in the mouth of the Amazon River from 9.4 Ma, but they stated that the contribution of Andean sediments was similar to the present only around 7.0 Ma, indicating that Amazon River was completely established by that time. The rise of the Fitzcarrald Arch probably by a flat subduction of the Nazca Ridge, around 4 Ma, separated the basins of the Ucayali and Madre de Dios rivers in Peru, leaving Amazon basin configuration very similar to the present-day (Espurt *et al.*, 2010).

Most of diversification events in *Proechimys*, regardless the set of areas, occurred in Late Pliocene and Pleistocene (< 4 Ma), a timing when the geological changes in the Amazon compared to the Miocene were less important (Hoorn *et al.*, 2010b; Mora *et al.*, 2010), in accordance to other studies on diversification in the Amazon (Moritz *et al.*, 2000; Antonelli *et al.*, 2010). However, from that time to the present-day the world witnessed several cycles of climate change (Zachos *et al.*, 2001). In the Pliocene, up to 3.0 Ma, the temperature was on average 3 ºC warmer than the pre-industrial period (Haywood & Valdes, 2004), favoring the development of forests and humid environments. Around 3.0 Ma, the Northern Hemisphere Glaciation intensified, which marks the end of the Pliocene and the transition to the Pleistocene (Bartoli, Hönisch, & Zeebe, 2011). From that period there is a global trend of cooling with several cycles of glacial and interglacial periods (Zachos *et al.*, 2001; Mudelsee & Raymo, 2005). Thus, climate changes may have affected much more the recent diversification of *Proechimys* than geological changes (Silva *et al.*, 2018).

Another event that was different depending on the analyses was the arrival of *Proechimys* populations in the Western Andes, trans-Andean areas, during the Pliocene in a unique event. DOM analysis indicated one dispersal event from Western Amazon (SBR) to Western Andes and another one, in the Pleistocene, to Central America. On the other hand, AOE analysis indicated a broad distribution of ancestral populations in the Western Andes (NOR) and Western Amazon (Inambari) and a posterior vicariant event separating these two

areas. Considering the dispersal hypothesis, during the Pliocene and Pleistocene, the Andes had considerable altitude to be a topographical and ecological barrier to dispersal (Mora *et al.*, 2010). However, dispersal events between these regions in rodents are not uncommon (Percequillo et al., in prep; Upham et al., 2013; Fabre et al., 2017). The Huacabamba Depression in northern Peru, considered a biogeographic barrier between Northern and Central Andes (Weigend, 2002), has a minimum altitude of 2145 m and currently encompass dry forests, and it is commonly believed to represent a connection route between the Western Andes and the Amazon basin (Duellman, 1979; Percequillo et al., in prep; Morrone & Urtubey, 1997; Prado & Percequillo, 2018). *Proechimys* is more common in lowland forests but there are occurrence records of specimens at altitudes of 2,000 m (Patton & Leite, 2015). Furthermore, when the dispersal event took place during the Late Miocene-Pliocene, the climate was warmer (Zachos *et al.*, 2001) and the Huacabamba Depression was lower than nowadays (Mora *et al.*, 2010); therefore, whether this dispersal event happened, it was likely around that region. One alternative dispersal route through the northern South America is doubtful because by that time Northern Andes were already high enough to be a topographic barrier (Mora *et al.*, 2010; Albert *et al.*, 2018), and the ancestral populations occupied the Western Amazon in both analyses. However, the vicariance event hypothesis cannot be ruled out. One ancestral populations widespreadly distributed on both cis- and trans-Andean area may have been connected by Huacabamba Depression until the Andes were high enough becoming a barrier to the dispersion, and specimens of *Proechimys* could no longer cross in the region. Dispersion to Central America from CAM subclade (Mean: 1.64; 2.89 – 0.64 Ma) was posterior to the formation of the Isthmus of Panama (Hoorn & Flantua, 2015; Montes *et al.*, 2015) and concomitantly with the replacement of savannas by forests in the Panama land bridge around 1 Ma (Leigh, O'Dea, & Vermeij, 2014).

Western Amazon presented a stable climate in the Pleistocene (Cheng *et al.*, 2013; Häggi *et al.*, 2017), and the few changes in ranges of lineages that inhabited the region, especially in clade A, corroborates this idea. Pleistocene glacial cycles rather than fragmenting the forest in the Amazon, as predicted by Refugia Hypothesis (Haffer, 1969), caused more change in the floristic composition of forests, especially in the border areas as Eastern Amazon (Colinvaux et al. 1996; 2001; Bush & Oliveira 2006). Most of the changes in distributional ranges in this period were in direction to drier areas such as Cerrado biome and the Eastern Amazon. Thus, it is reasonable to affirm that Pleistocene climatic oscillations

were important to evolutionary history of the genus only in more recent events, as supported by Silva et al. (2018). These climatic oscillations may have affected the occupations of the forested areas of the dry diagonal in South America, such as the Cerrado, which occurred at different geographic points: an originated event from the Eastern Amazon and two from the Western Amazon. It was not possible to affirm that these events occurred at different Pleistocene epochs, because their confidence intervals were overlapped but the distinct means for the events indicate that they may have occurred in different Pleistocene glacial cycles.

Finally, using two sets of areas in ancestral range estimations showed that the number/size of areas in the analyses influenced the vicariance events but did not affect the number of dispersions. In addition, it was possible to recover 10 common events in both scenarios, such as the *Proechimys* origin in the Western Amazon, the relationship between landscape evolution and diversification of the genus, and the differentiation within clades in the Pliocene-Pleistocene under the influence of climatic oscillations. Therefore, different number/size of areas in the analyses had little effect on the reconstruction of the biogeographic history of *Proechimys,* rather they further affect the model ability to identify the presence of vicariant events. Fabre et al. (2017) suggested more dispersion than vicariance events in the biogeographic history of the Family Echimyidae, using six large areas. We also recovered a dominance of dispersion events in the DOM analysis, with fewer areas of larger size. Therefore, despite being a family-scale analysis, the number of areas and their size may have influenced the results about vicariance events.

3.5. Conclusions

The most recent common ancestor (MRCA) for *Proechimys* was estimated to have originated about 11 Ma (95% CI = 17.01 – 5.98 Ma), in the Miocene. The diversification of the five main clades occurred at a similar age (10.53 – 10.54 Ma), in the Late Miocene. Thus, the origin and basal diversification of *Proechimys* (clades A+B+C; clades A+B; and clades D+E) can be associated to a rapid diversification event.

The main differences among biogeographic analyses using different sets of areas were the influence of the vicariance events, which had little effect on the biogeographic hipothesis in DOM analyses (smaller areas number with large size), when compared to the

AOE analyses (larger areas number with small size). Ages for clades origin did not coincide with the formation of the Amazonian rivers that delimit their distributions. Therefore, it is more plausible that the presence of the rivers delimit secondary contact areas between the lineages.

Proechimys diversification is associated with the landscape evolution of the Amazon basin. Origin of the genus probably occurred in the Western Amazon, during the existence of the Pebas System, a mega-wetland area, and after the Middle Miocene Disruption which generated extinction and consequently vacant niches in the region. The diversification of main clades agreed with the change to the current flow of the Amazon River, the decrease of the wetlands and the expansion and diversification of the terrestrial biota and the *terra-firme* forests. Most intense diversification events within the clades dated back to the Late Pliocene and Pleistocene, when a global tendency of cooling has initiated, as well as cyclical ice ages. Thus, more recent diversification events were closely linked to climate change.

References

ALBERT, J. S., P. VAL, C. HOORN, J. S. ALBERT, P. VAL, AND C. HOORN. 2018. The changing course of the Amazon River in the Neogene: center stage for Neotropical diversification. Neotropical Ichthyology 16:e180033.

ÁLVAREZ, A., R. L. M. ARÉVALO, AND D. H. VERZI. 2017. Diversification patterns and size evolution in caviomorph rodents. Biological Journal of the Linnean Society 121:907–922.

ANTONELLI, A., A. QUIJADA-MASCAREÑAS, A. J. CRAWFORD, J. M. BATES, P. M. VELAZCO, AND W. WÜSTER. 2010. Molecular studies and phylogeography of Amazonian Tetrapods and their relation to geological and climatic models. Pp. 386–404 in Amazonia: Landscape and Species Evolution. Wiley-Blackwell Publishing Ltd., Oxford, UK.

BARTOLI, G., B. HÖNISCH, AND R. E. ZEEBE. 2011. Atmospheric CO_2 decline during the Pliocene intensification of Northern Hemisphere glaciations. Paleoceanography 26.

BORGES, S. H., AND J. M. C. DA SILVA. 2012. A new area of endemism for Amazonian birds in the Rio Negro basin. The Wilson Journal of Ornithology 124:15–23.

BOUCKAERT, R. ET AL. 2014. BEAST 2: a software platform for bayesian evolutionary analysis. PLoS Computational Biology 10:e1003537.

BUSH, M. B., AND P. E. DE OLIVEIRA. 2006. The rise and fall of the Refugial Hypothesis of Amazonian speciation: a paleoecological perspective. Biota Neotropica 6:bn0010601.

CANDELA, A. M., AND J. I. NORIEGA. 2004. Los coipos (Rodentia, Caviomorpha, Myocastoridae) del "Mesopotamiense"; (Mioceno tardío; Formación Ituzaingó) de la provincia de Entre Ríos, Argentina. INSUGEO, Miscelánea 12:77–82.

CARNAVAL, A. C., M. J. HICKERSON, C. F. HADDAD, M. T. RODRIGUES, AND C. MORITZ. 2009. Stability predicts genetic diversity in the brazilian atlantic forest hotspot. Science 323:785–789.

CHENG, H. ET AL. 2013. Climate change patterns in Amazonia and biodiversity. Nature Communications 4:1411.

CLAPPERTON, C. M. 1993. Nature of environmental changes in South America at the Last Glacial Maximum. *Palaeogeography, Palaeoclimatology, Palaeoecology*. .

COBBOLD, P. R., E. A. ROSSELLO, P. ROPERCH, C. ARRIAGADA, L. A. GÓMEZ, AND C. LIMA. 2007. Distribution, timing, and causes of Andean deformation across South America. Geological Society, London, Special Publications 272:321–343.

COLINVAUX, P. A., G. IRION, M. E. RÄSÄNEN, M. B. BUSH, AND J. N. DE MELLO. 2001. A paradigm to be discarded: geological and paleoecological data falsify the Haffer & Prance refuge hypothesis of Amazonian speciation. Amazoniana 16:609–646.

COLINVAUX, P. A., P. E. DE OLIVEIRA, J. E. MORENO, M. C. MILLER, AND M. B. BUSH. 1996. A long pollen record from lowland Amazonia: forest and cooling in glacial times. Science 274:85–88.

COURCELLE, M., M. TILAK, Y. L. R. LEITE, E. J. P. DOUZERY, AND P.-H. FABRE. 2019. Digging for the spiny rat and hutia phylogeny using a gene capture approach, with the description of a new mammal subfamily. Molecular Phylogenetics and Evolution 136:241–253.

CRACRAFT, J. 1985. Historical biogeography and patterns of differentiation within the South American avifauna: areas of endemism. Ornithological Monographs:49–84.

DESCHAMPS, C. M., E. C. VIEYTES, A. I. OLIVARES, AND M. G. VUCETICH. 2009. Primer registro de *Cardiatherium chasicoense* (Rodentia, Hydrochoeridae) fuera del área pampeana (Argentina) y su valor bioestratigráfico. Ameghiniana 46:295–305.

DUELLMAN, W. E. 1979. The herpetofauna of the Andes: Patterns of distribution, origin, differentiation, and present communities. Pp. 371–459 in The South American herpetofauna: its origin, evolution, and dispersal (W. E. Duellman, ed.). University of Kansas Natural History Museum, Lawrence.

ESPURT, N., P. BABY, S. BRUSSET, M. RODDAZ, W. HERMOZA, AND J. BARBARAND. 2010. The Nazca Ridge and uplift of the Fitzcarrald Arch: implications for regional geology in northern South America. Pp. 89–100 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

FABRE, P. H. ET AL. 2017. Mitogenomic phylogeny, diversification, and biogeography of South American spiny rats. Molecular Biology and Evolution 34:613–633.

FABRE, P. H., T. GALEWSKI, M. K. TILAK, AND E. J. P. DOUZERY. 2013. Diversification of South American spiny rats (Echimyidae): A multigene phylogenetic approach. Zoologica Scripta 42:117–134.

FABRE, P. H., J. L. PATTON, AND Y. L. R. LEITE. 2016. Family Echimyidae (hutias, South American spiny-rats and coypu). Pp. 552–641 in Handbook of the Mammals of the World. Vol 6. Lagomorphs and Rodents I (D. E. Wilson, T. E. J. Lacher & R. A. Mittermeier, eds.). Lynx Edicions, Barcelona.

HAFFER, J. 1969. Speciation in Amazonian forest birds. Science 165:131–137.

HÄGGI, C. ET AL. 2017. Response of the Amazon rainforest to late Pleistocene climate variability. Earth and Planetary Science Letters 479:50–59.

HAYWOOD, A. M., AND P. J. VALDES. 2004. Modelling Pliocene warmth: contribution of atmosphere, oceans and cryosphere. Earth and Planetary Science Letters 218:363–377.

HIPSLEY, C. A., AND J. MÜLLER. 2014. Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. Frontiers in Genetics 5:138.

HO, S. Y. W., AND M. J. PHILLIPS. 2009. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Systematic Biology 58:367–380.

HOORN, C. 1993. Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia: results of a palynostratigraphic study. Palaeogeography, Palaeoclimatology, Palaeoecology 105:267–309.

HOORN, C. ET AL. 2010a. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330:927–931.

HOORN, C. ET AL. 2017. The Amazon at sea: Onset and stages of the Amazon River from a marine record, with special reference to Neogene plant turnover in the drainage basin. Global and Planetary Change 153:51–65.

HOORN, C., AND S. FLANTUA. 2015. An early start for the Panama land bridge. Science 348:186–187.

HOORN, C., F. P. WESSELINGH, J. HOVIKOSKI, AND J. GUERRERO. 2010b. The development of the Amazonian mega-wetland (Miocene; Brazil, Colombia, Peru, Bolivia). Pp. 123–142 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

JARAMILLO, C. ET AL. 2010. The origin of the modern Amazon rainforest: implications of the palynological and palaeobotanical record. Pp. 317–334 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

KAANDORP, R. J. G., H. B. VONHOF, F. P. WESSELINGH, L. R. PITTMAN, D. KROON, AND J. E. VAN HINTE. 2005. Seasonal Amazonian rainfall variation in the Miocene Climate Optimum. Palaeogeography, Palaeoclimatology, Palaeoecology 221:1–6.

KAWATA, M. 2002. Invasion of vacant niches and subsequent sympatric speciation. Proceedings of the Royal Society of London. Series B: Biological Sciences 269:55–63.

LANDIS, M. J., N. J. MATZKE, B. R. MOORE, AND J. P. HUELSENBECK. 2013. Bayesian analysis of biogeography when the number of areas is large. Systematic Biology 62:789–804.

LARA, M. C., J. L. PATTON, AND M. N. F. DA SILVA. 1996. The simultaneous diversification of South American echimyid rodents (Hystricognathi) based on complete citochrome b sequences. Molecular Phylogenetics and Evolution 5:403–413.

LEIGH, E. G., A. O'DEA, AND G. J. VERMEIJ. 2014. Historical biogeography of the Isthmus of Panama. Biological Reviews 89:148–172.

LEITE, Y. L. R. ET AL. 2016. Neotropical forest expansion during the last glacial period challenges refuge hypothesis. Proceedings of the National Academy of Sciences 113:1008–1013.

LEITE, Y. L. R., AND J. L. PATTON. 2002. Evolution of South American spiny rats (Rodentia, Echimyidae): the star-phylogeny hypothesis revisited. Molecular Phylogenetics and Evolution 25:455–464.

LEKEVIČIUS, E. 2009. Vacant niches in nature, ecology, and evolutionary theory: a mini-review. Ekologija 55:165–174.

LEWIS, A. R. ET AL. 2008. Mid-Miocene cooling and the extinction of tundra in continental Antarctica. Proceedings of the National Academy of Sciences of the United States of America 105:10676–80.

LUNDBERG, J. G., L. G. MARSHALL, J. GUERRERO, B. K. HORTON, M. C. S. L. MALABARBA, AND F. WESSELINGH. 1998. The stage for neotropical fish diversification: A history of tropical South American rivers. Pp. 13–48 in Phylogeny and Classification of Neotropical Fishes (L. R. Malabarba, R. E. Reis, R. P. Vari, Z. M. Lucena & C. A. S. Lucena, eds.). Edipucrs, Porto Alegre.

MATZKE, N. J. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Frontiers of Biogeography 5:242–248.

MATZKE, N. J. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Systematic Biology 63:951–970.

MIALL, A. D. 2002. Architecture and sequence stratigraphy of Pleistocene fluvial systems in the Malay Basin, based on seismic time-slice analysis. AAPG Bulletin 86:1201–1216.

MONTES, C. ET AL. 2015. Middle Miocene closure of the Central American Seaway. Science 348:226–9.

MORA, A. ET AL. 2010. Tectonic history of the Andes and sub-Andean zones: Implications for the development of the Amazon drainage basin. Pp. 38–60 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

MORITZ, C., J. L. PATTON, C. J. SCHNEIDER, AND T. B. SMITH. 2000. Diversification of rainforest faunas: an integrated molecular approach. Annual Review of Ecology and Systematics 31:533–563.

MORRONE, J. J. 2014. Biogeographical regionalisation of the Neotropical region. Zootaxa 3782:110 pp.

MORRONE, J. J., AND E. URTUBEY. 1997. Historical biogeography of the northern Andes: a cladistic analysis based on five genera of Rhytirrhinini (Coleoptera: Curculionidae) and *Barnadesia* (Asteraceae). Biogeographica 73:115–121.

MUDELSEE, M., AND M. E. RAYMO. 2005. Slow dynamics of the Northern Hemisphere glaciation. Paleoceanography 20:n/a-n/a.

PARHAM, J. F. ET AL. 2012. Best practices for justifying fossil calibrations. Systematic Biology 61:346–359.

PATTON, J. L., M. N. DA SILVA, AND J. R. MALCOLM. 2000. Mammals of the rio Juruá and the evolutionary and ecological diversification of Amazonia. Bulletin American Museum of Natural History 244:1–306.

PATTON, J. P., AND R. N. LEITE. 2015. Genus *Proechimys* J. A. Allen, 1889. Pp. 950–989 in Mammals of South America, Volume 2: Rodents (J. L. Patton, U. F. J. Pardiñas & G. D'Elía, eds.). University Of Chicago Press, Chicago.

PERCEQUILLO, A. R. ET AL. Tempo and mode of evolution of Oryzomyine rodents (Rodentia, Cricetidae, Sigmodontinae): a phylogenomic approach. In Prep.

PETERSON, B. K., J. N. WEBER, E. H. KAY, H. S. FISHER, AND H. E. HOEKSTRA. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

PONS, D., AND D. DE FRANCESCHI. 2007. Neogene woods from western Peruvian Amazon and palaeoenvironmental interpretation. Bulletin of Geosciences 82:343–354.

PRADO, J. R. DO, AND A. R. PERCEQUILLO. 2018. Systematic studies of the genus *Aegialomys* Weksler et al., 2006 (Rodentia: Cricetidae: Sigmodontinae): geographic variation, species delimitation, and biogeography. Journal of Mammalian Evolution 25:71–118.

R-DEVELOPMENT CORETEAM. 2018. R Development Core Team. *R: A Language and Environment for Statistical Computing*.

RAMBAUT, A., A. J. DRUMMOND, D. XIE, G. BAELE, AND M. A. SUCHARD. 2018. Posterior summarization in bayesian phylogenetics using Tracer 1.7. Systematic Biology 67:901–904.

REE, R. H., AND I. SANMARTÍN. 2018. Conceptual and statistical problems with the DEC+J model of founder-event speciation and its comparison with DEC via model selection. Journal of Biogeography 45:741–749.

REE, R. H., AND S. A. SMITH. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57:4–14.

RONQUIST, F. 1997. Dispersal-Vicariance Analysis: a new approach to the quantification of historical biogeography. Systematic Biology 46:195–203.

SDROLIAS, M., AND R. D. MÜLLER. 2006. Controls on back-arc basin formation. Geochemistry, Geophysics, Geosystems 7:Q04016.

SILVA, C. R. ET AL. 2018. The role of Pleistocene climate change in the genetic variability, distribution and demography of *Proechimys cuvieri* and P. guyannensis (Rodentia: Echimyidae) in northeastern Amazonia. PLoS One 13:e0206660.

DA SILVA, J. M., F. C. NOVAES, AND D. C. OREN. 2002. Differentiation of *Xiphocolaptes* (Dendrocolaptidae) across the river Xingu, Brazilian Amazonia: recognition of a new phylogenetic species and biogeographic implications. Bulletin of the British Ornithologists' Club 122:185–193.

DA SILVA, M. N., AND J. L. PATTON. 1998. Molecular phylogeography and the evolution and conservation of amazonian mammals. Molecular Ecology 7:475–486.

UPHAM, N. S., R. OJALA-BARBOUR, J. BRITO M, P. M. VELAZCO, AND B. D. PATTERSON. 2013. Transitions between Andean and Amazonian centers of endemism in the radiation of some arboreal rodents. BMC Evolutionary Biology 13:191.

UPHAM, N. S., AND B. D. PATTERSON. 2015. Evolution of caviomorph rodents: a complete phylogeny and timetree for living genera. Pp. 63–120 in Biology of caviomorph rodents: diversity and evolution (A. I. Vassallo & D. Antenucci, eds.). SAREM Series A, Buenos Aires.

VERZI, D. H., C. I. MONTALVO, AND C. M. DESCHAMPS. 2008. Biostratigraphy and biochronology of the Late Miocene of central Argentina: Evidence from rodents and taphonomy. Geobios 41:145–155.

VUCETICH, M. G. 1995. *Theridomysops parvulus* (Rovereto, 1914), un primitivo Eumysopinae (Rodentia, Echimyidae) del Mioceno Tardío de Argentina. Mastozoología Neotropical 2:167–172.

VAN VUUREN, B. J., S. KINET, J. CHOPELET, AND F. CATZEFLIS. 2004. Geographic patterns of genetic variation in four Neotropical rodents: conservation implications for small game mammals in French Guiana. Biological Journal of the Linnean Society 81:203–218.

WAGENMAKERS, E.-J., AND S. FARRELL. 2004. AIC model selection using Akaike weights. Psychonomic Bulletin & Review 11:192–196.

WEIGEND, M. 2002. Observations on the biogeography of the Amotape-Huancabamba Zone in northern Peru. The Botanical Review 68:38–54.

WESSELINGH, F. P. ET AL. 2001. Lake Pebas: a palaeoecological reconstruction of a Miocene, long-lived lake complex in western Amazonia. Cainozoic Research 1:35–81.

WESSELINGH, F. P. 2006. Miocene long-lived lake Pebas as a stage of mollusc radiations, with implications for landscape evolution in western Amazonia. Scripta Geologica 133:1–17.

WESSELINGH, F. P., AND J. A. SALO. 2006. A Miocene perspective on the evolution of the Amazonian biota. Scripta Geologica 133:439–458.

ZACHOS, J., M. PAGANI, L. SLOAN, E. THOMAS, AND K. BILLUPS. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. Science 292:686–93.

Tables

Table 1: Divergence times in million years ago (Ma) for the main clades of the *Proechimys* phylogeny, with mean, median, and 95% confidence interval (C.I.) values. Values for the other clades, see Table S3. Position of the node numbers are shown in Fig. 1. A = clade A; B = clade B; C = clade C; CER= lineage in the clade B with individuals from Cerrado biome; CLJ+MMR+URJ = lineages with individuals from *várzea* (seasonal flooded forest) in the Western Amazon; D = clade D; HG = *Hoplomys gymnurus;* MAR+ECU+CAM = lineages from Western Andes; PRO = *Proechimys* "core" (except for Tepui individual); PTI= lineage in the clade A with individuals from Cerrado biome; SCE = lineage in the clade D with individuals from Cerrado biome; TEP = Tepui individual.

Nodes	Clades	Mean	Median	95% C.I.	
				Lower	Upper
1	HG + TEP +PRO	12.68	12.12	19.94	7.05
2	HG + TEP	8.16	7.68	14.87	2.99
3	HG	0.50	0.45	1.00	0.11
4	PRO	11.00	10.56	17.01	5.98
5	A+B+C	10.34	9.90	15.82	5.26
6	A+B	10.14	9.69	15.46	5.20
7	A	8.20	7.82	12.89	4.48
8	B	8.73	8.39	13.44	4.44
9	C	2.55	2.40	4.41	1.10
10	D+E	10.53	10.07	16.62	5.79
11	D	8.73	8.32	13.52	4.30
12	E	3.53	3.35	5.64	1.47
13	MAR+ECU+CAM	5.50	5.17	9.52	2.44
14	SCE	0.64	0.60	1.16	0.26
15	PTI	0.22	0.19	0.46	0.04
16	CER	0.35	0.32	0.63	0.11
17	CLJ+MMR+URJ	1.33	1.24	2.12	0.64

Table 2: Log-likelihood values (LnL) for each one of five biogeographic models tested using the Morrone's dominions for Neotropics (DOM analysis) and the areas of endemism for Amazon (AOE analysis). Akaike Information Criterion (AIC) 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 informed.

Analyses	Models	LnL	P	d	e	j	AIC	AICwt
DOM	DEC	-114.2	2	0.012	0.0053	0	232.4	0.0008
	DIVALIKE	-111.2	2	0.013	0.0004	0	226.4	0.015
	DIVALIKE+J	-106.9	3	0.0094	1.0e-12	0.021	219.8	0.40
	BAYAREALIKE	-139.1	2	0.016	0.091	0	282.2	1.2e-14
	BAYAREALIKE+J	-113.2	3	0.0063	0.0019	0.04	232.3	0.0008
AOE	DEC	-153.4	2	0.0076	1.0e-12	0	310.8	1.4e-06
	DIVALIKE	-149.9	2	0.0093	1.0e-12	0	303.7	4.8e-05
	DIVALIKE+J	-140.2	3	0.0048	1.0e-12	0.033	286.3	0.29
	BAYAREALIKE	-230.0	2	0.010	0.010	0	464.0	7.5e-40
	BAYAREALIKE+J	-143.4	3	0.0034	1.0e-07	0.044	292.8	0.011

Table 3: Probable events of vicariance (V) and dispersion (D) according to the biogeographic analyzes based on the Morrone's dominions (DOM) for Neotropical region (Fig. 2). Clades where each event occurred (Clades), mean age (Mean) and confidence interval of 95% (95% C.I.) in millions of years ago (Ma), geological Epoch (Epoch), and the direction (Direction) of the events through areas (Movement) are also presented. For Epoch: MIO = Miocene; PLIO = Pliocene; PLEIS = Pleistocene. For areas: MES = Mesoamerican dominion; PAC = Pacific dominion; BBR = Boreal Brazilian dominion; SBR = South Brazilian dominion; SEA = South-eastern Amazonian dominion; CHA = Chaco dominion; and PAR = Parana dominion. For clades: A = clade A; B = clade B; C = clade C; D = clade D; E = clade E; HG = *Hoplomys gymnurus;* PRO = *Proechimys* "core" (except for Tepui individual); TEP = Tepui individual. For Movements: > for new areas occupied, and / for a division in the ancestral range distribution.

Events	Clades	Time (Ma)		Epoch	Direction	Movement
		Mean	95% C.I.			
D1	HG+TEP+PRO	12.68 – 8.16	19.94 – 2.99	MIO – PLIO	latitudinal	SBR > BBR+PAC+MES
D2	D	8.73 – 5.50	13.52 – 1.47	MIO – PLIO – PLEIS	longitudinal	SBR > PAC
D3	D	3.96 – 1.65	6.16 – 0.64	PLIO – PLEIS	latitudinal	PAC > MES
D4	D	1.02 – 0.64	1.66 – 0.26	PLEIS	longitudinal	SBR > CHA
D5	D	2.11 – 0.74	3.36 – 0.30	PLIO – PLEIS	latitudinal	SBR > BBR
D6	D	3.28 – 0.90	5.06 – 0.40	PLIO – PLEIS	both	SBR > BBR+SEA
D7	D	0.79 – 0.30	1.29 – 0.11	PLEIS	latitudinal	SBR > BBR
D8	C	10.34 – 2.55	15.82 – 1.10	MIO – PLIO – PLEIS	latitudinal	SBR > BBR
D9	B	8.73 – 0.88	13.44 – 0.30	MIO – PLIO – PLEIS	longitudinal	SBR > PAC
D10	B	3.71 – 2.27	5.50 – 1.12	PLIO – PLEIS	longitudinal	SBR > SEA
D11	B	1.25 – 0.28	3.77 – 0.08	PLIO – PLEIS	both	SEA > CHA

Table 3: Continuation.

Events	Clades	Time (Ma)		Epoch	Direction	Movement
		Mean	95% C.I.			
D12	B	0.91 – 0.35	1.47 – 0.11	PLEIS	both	SEA > BBR+CHA
D13	A	2.11 – 0.83	3.38 – 0.27	PLIO – PLEIS	latitudinal	SBR > BBR
D14	A	3.96 – 1.19	6.26 – 0.37	MIO – PLIO – PLEIS	latitudinal	SBR > BBR
D15	A	3.34 – 0.81	5.20 – 0.29	PLIO – PLEIS	longitudinal	SBR > SEA
D16	A	2.78 – 0.22	4.41 – 0.04	PLIO – PLEIS	both	SBR > CHA+SEA
V1	HG+TEP	8.16	14.87 – 2.99	MIO – PLIO	longitudinal	BBR / PAC+MES
V2	D	0.9	1.49 – 0.4	PLEIS	latitudinal	BBR / SEA

Table 4: Probable events of vicariance (V) and dispersion (D) according to the biogeographic analyzes based on the areas of endemism (AOE) for the Amazon (Fig. 3). Clades where each event occurred (Clades), mean age (Mean) and confidence interval of 95% (95% C.I.) in millions of years ago (Ma), geological Epoch (Epoch), and the direction (Direction) of the events through areas (Movement) are also presented. For Epoch: MIO = Miocene; PLIO = Pliocene; PLEIS = Pleistocene. For areas: BEL = Belem; GUY = Guyana; IME = Imeri; INA = Inambari; JAU = Jau; NAP = Napo; RON = Rondonia; TAP = Tapajos; XGU = Xingu, NOR = North to the Amazon; SOU = South to the Amazon. For clades: A = clade A; B = clade B; C = clade C; D = clade D; E = clade E; HG = *Hoplomys gymnurus;* PRO = *Proechimys* "core" (except for Tepui individual); TEP = Tepui individual. For Movements: > for new areas occupied, / for a division in the ancestral range distribution, and * for non-adjacent areas using 20% or more of probabilities of occupancy.

Events	Clades	Time (Ma)		Epoch	Direction	Movement
		Mean	95% C.I.			
D1	THR+HG+TEP+PRO	17.14 – 12.68	27.36 – 7.05	MIO – PLIO	latitudinal	SOU > NOR+GUY *
D2	PRO	12.68 – 11.00	19.94 – 5.98	MIO – PLIO	latitudinal	NOR+GUY > INA
D3	D+E	11.00 – 10.53	17.01 – 5.79	MIO – PLIO	both	NOR+GUY+INA > NAP
D4	D	2.94 – 1.02	4.69 – 0.50	PLIO – PLEIS	latitudinal	INA > SOU
D5	D	2.11 – 0.74	3.36 – 0.30	PLIO – PLEIS	latitudinal	INA > NAP
D6	D	3.28 – 0.90	5.06 – 0.40	PLIO – PLEIS	both	NAP+INA+TAP > GUY+XGU
D7	D	6.78 – 4.68	10.49 – 2.36	MIO – PLIO – PLEIS	longitudinal	INA > RON
D8	D	4.68 – 3.28	7.25 – 1.53	MIO – PLIO – PLEIS	both	INA > TAP+NAP *
D9	D	3.61 – 0.78	5.77 – 0.30	PLIO – PLEIS	latitudinal	INA > SOU
D9	D	3.28 – 1.55	5.06 – 0.76	PLIO – PLEIS	both	NAP+INA+TAP > JAU
D10	B	3.51 – 2.27	5.50 – 1.12	MIO – PLIO – PLEIS	both	TAP+SOU > XGU+BEL

Table 4: Continuation.

Events	Clades	Time (Ma)		Epoch	Direction	Movement
		Mean	95% C.I.			
D11	D	1.59 – 1.38	2.63 – 0.65	PLIO – PLEIS	longitudinal	RON > TAP
D12	C	1.22 – 0.38	1.92 – 0.11	PLEIS	longitudinal	GUY > IME
D13	A	2.11 – 0.83	3.38 – 0.27	PLIO – PLEIS	latitudinal	INA > JAU
D14	A	3.96 – 1.19	6.26 – 0.37	MIO – PLIO – PLEIS	latitudinal	INA > JAU
D15	A	3.34 – 0.81	5.20 – 0.29	PLIO – PLEIS	longitudinal	INA > TAP+XGU *
D16	A	2.78 – 0.22	4.41 – 0.04	PLIO – PLEIS	both	INA > RON+TAP+SOU
D17	B	8.73 – 3.51	13.44 – 1.58	MIO – PLIO – PLEIS	latitudinal	NOR > RON+TAP+SOU *
V1	HG+TEP	8.16	14.87 – 2.99	MIO – PLIO	longitudinal	NOR / GUY
V2	D	0.9	1.49 – 0.4	PLEIS	latitudinal	GUY / TAP+XGU
V3	D	8.73	13.52 – 4.30	MIO – PLIO	both	INA / NOR
V4	D	4.68	7.25 – 2.36	MIO – PLIO – PLEIS	longitudinal	INA / RON
V6	E	1.8	3.30 – 0.61	PLIO – PLEIS	latitudinal	NAP / INA
V7	D	1.55	2.54 – 0.76	PLEIS	both	NAP / INA+JAU
V8	D	0.79	1.29 – 0.34	PLEIS	latitudinal	INA / JAU
V9	A+B+C	10.34	15.82 – 5.26	MIO – PLIO	longitudinal	GUY / NOR+INA
V10	A+B	10.14	15.46 – 5.20	MIO – PLIO	latitudinal	INA / NOR
V12	B	3.51	5.50 – 1.58	PLIO – PLEIS	both	RON / TAP+SOU
V13	B	2.27	3.77 – 1.12	PLIO – PLEIS	longitudinal	TAP / XGU+BEL+SOU
V14	B	0.91	1.47 – 0.43	PLEIS	longitudinal	XGU / BEL+SOU

Figures

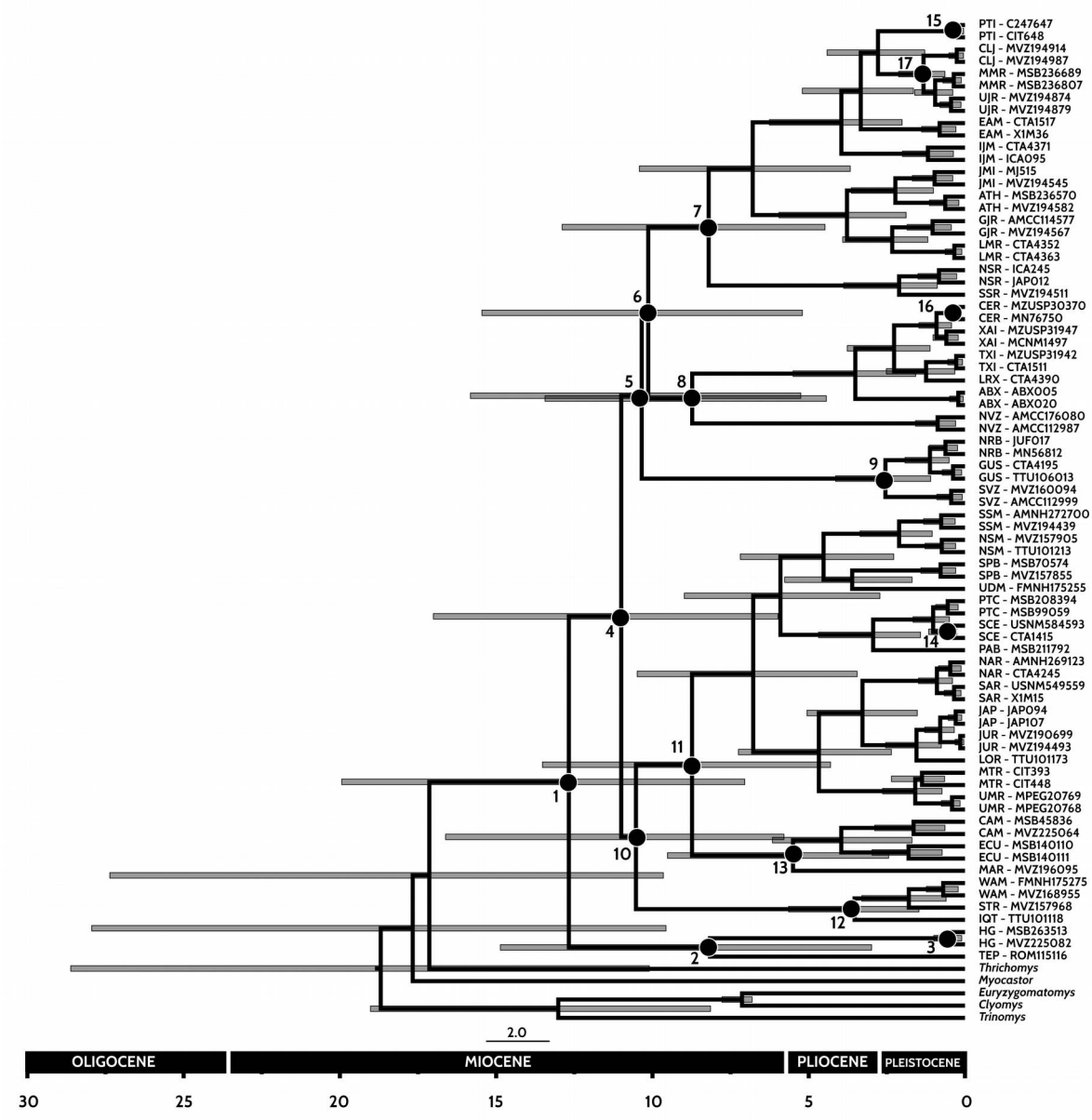


Figure 1: Dated tree for the genus *Proechimys*. Tips indicate the subclades acronyms and the individuals used in the analyzes. For more details on samples, see Table S1 and Table S2. Bars represent the 95% confidence interval for the dates on each node. Dates in absolute values are in the Table S3 and the numbered nodes are also present in the Table 1. Time scale is in millions of years (Ma) and axis is divided by Epochs, and the Holocene is not shown due to scale.

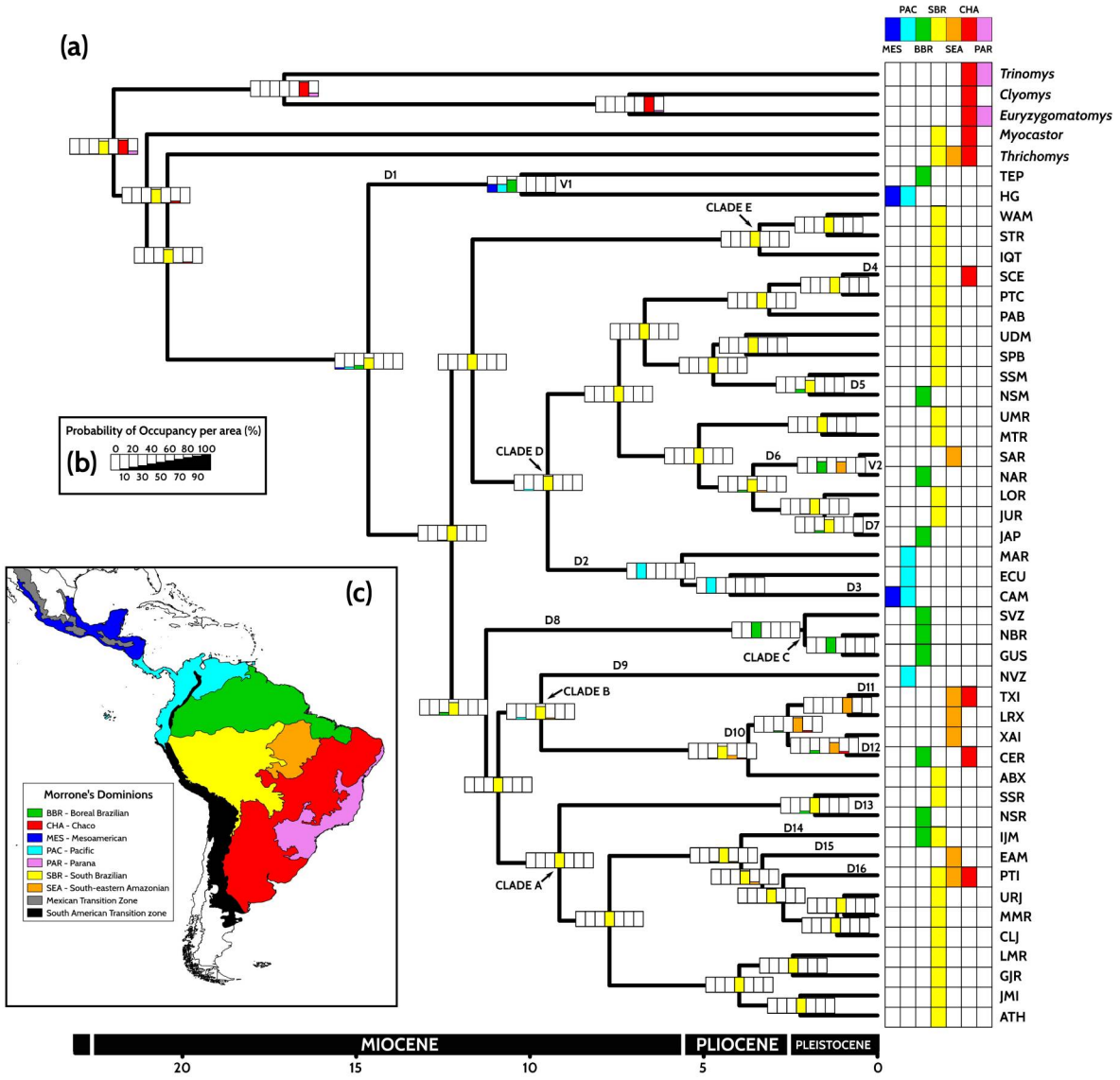


Figure 2: Geographic range evolution in the genus *Proechimys* (a) based on the DIVALIKE + J model and on the seven Morrone's biogeographic dominions (DOM) for Neotropical region (c). Dated tree was built with BEAST, and the tips indicate the subclades acronyms, see Table S1 and Table S2 for details. Each node and tip has a rectangle divided into 7 smaller squares, each of them with different colors, representing the probability of the ancestral populations occupied that area (b). In the tips, all areas where the lineage occurs have a 100% of probability of occupancy. The five main clades are indicated with arrows. Letters indicate the possible of vicariant (V) and dispersal (D) events. Time scale is in millions of years (Ma) and axis is divided by Epochs, and the Holocene is not shown due to scale.

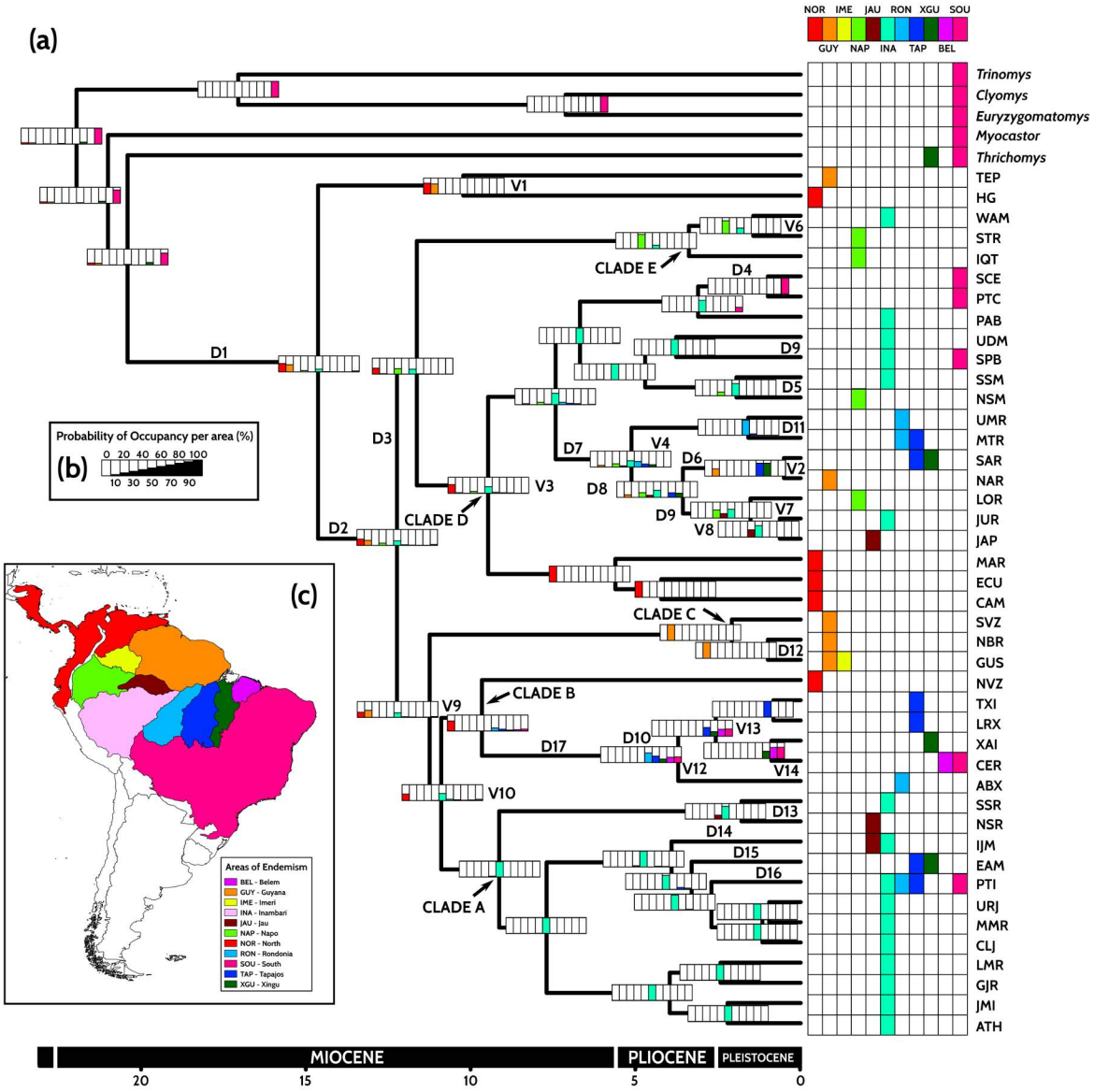


Figure 3: Geographic range evolution in the genus *Proechimys* (a) based on the DIVALIKE + J model and on the areas of endemism (AOE) for Amazon (c). Dated tree was built with BEAST, and the tips indicate the subclades acronyms, see Table S1 and Table S2 for details. Each node and tip has a rectangle divided into 11 smaller squares, each of them with different colors, representing the probability of the ancestral populations occupied that area (b). In the tips, all areas where the lineage occurs have a 100% of probability of occupancy. The five main clades are indicated with arrows. Letters indicate the possible vicariant (V) and dispersal (D) events in the phylogeny. Time scale is in millions of years (Ma) and axis is divided by Epochs, and the Holocene is not shown due to scale.

Supplementary Material

Tables

Table S1: 82 Samples used in the genomics analyzes with information about the locality, species groups, and institution of origin of the samples. **AMNH-AMCC:** Ambrose Monell Cryo Collection, American Museum of Natural History, New York, USA; **CIT:** Coleção de Tecidos Miguel Trefaut Rodrigues, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil; **FMNH:** Field Museum of Natural History, Chicago, Illinois, USA; **LMUSP:** Laboratório de Mamíferos da Escola Superior de Agricultura "Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil; **MCN-M:** Coleção de Mastozoologia do Museu de Ciências Naturais da Pontifícia Universidade Católica de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; **MN:** Museu Nacional da Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; **MPEG**: Museu Paraense Emílio Goeldi, Belém, Pará, Brazil; **MSB:** Museum of Southwestern Biology, Alburqueque, New Mexico, USA; **MVZ:** Museum of Vertebrate Zoology, Berkeley, California, USA; **MZUSP:** Museu de Zoologia da Universidade de São Paulo, São Paulo, Brazil; **USNM:** National Museum of Natural History, Washington, D.C., USA; **ROM:** Royal Ontario Museum, Toronto, Ontario, Canada; **TTU**: Texas Tech University, Lubbock, Texas, USA; **UFES-CTA:** Coleção de Tecido Animal da Universidade Federal do Espírito Santo, Vitória, Espírito Santo; Brazil. Table with more information about locality in ***.csv*** format is available on https://github.com/jdalapicolla.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
ABX005		LMUSP	guyannensis	-58.63314	-4.31203
ABX020		LMUSP	guyannensis	-58.63442	-4.34328
AMCC112987	USNM448733	AMNH-AMCC	trinitatis	-72.8091	9.84363
AMCC112999	USNM448711	AMNH-AMCC	guyannensis	-61.09269	5.07502
AMCC114577	MUSM23828	AMNH-AMCC	gardneri	-73.66744	-4.90625
AMCC176080	JOG4488; EBRG25473	AMNH-AMCC	trinitatis	-69.95012	11.81811
AMNH269123	LHE1163	AMNH-AMCC	longicaudatus	-52.92366	5.27438
AMNH272700	RSV2095	AMNH-AMCC	longicaudatus	-73.16208	-5.2495
C247647	C247647-6738	MZUSP	goeldii	-56.51269	-9.45183
CIT393	PEU960065	CIT	longicaudatus	-59.44718	-10.17484
CIT448	M968498	CIT	longicaudatus	-57.39563	-9.5678
CTA1028	YL53	CTA-UFES	*E. spinosus*	-43.5	-20.083
CTA1415	MVZ197661; LPC462	CTA-UFES	longicaudatus	-52.35583	-15.63333
CTA1511	UFMG3029	CTA-UFES	guyannensis	-55.93028	-9.59694
CTA1517	MVZ197575; LPC564	CTA-UFES	goeldii	-55.93028	-9.59694

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
CTA4195	MPEG42380; SLF225	CTA-UFES	guyannensis	-56.36422	-1.79591
CTA4245	MPEG42367	CTA-UFES	longicaudatus	-56.36422	-1.79591
CTA4352	UFES2705	CTA-UFES	gardneri	-59.1306	-3.57972
CTA4363	UFES2834	CTA-UFES	gardneri	-59.1306	-3.57972
CTA4371	UFES2842	CTA-UFES	goeldii	-59.1306	-3.57972
CTA4390	UFES2945; BM74724	CTA-UFES	guyannensis	-51.77012	-3.12428
EEB1013		LMUSP	*T. dimidiatus*	-44.368	-22.80657
FMNH175255	UPE133	FMNH	longicaudatus	-71.38542	-12.77165
FMNH175275	UPE231	FMNH	simonsi	-71.49185	-13.02362
FMPS010		LMUSP	*C. laticeps*	-57.65317	-18.99968
ICA095		LMUSP	goeldii	-68.33989	-2.90496
ICA245		LMUSP	echinothrix	-68.88251	-3.03233
JAP012		LMUSP	echinothrix	-65.75583	-1.76103
JAP094		LMUSP	longicaudatus	-66.35717	-1.76422
JAP107		LMUSP	longicaudatus	-66.35717	-1.76422
JUF017		LMUSP	guyannensis	-62.15008	-0.94917
MBR046		LMUSP	*T. pachyurus*	-54.61611	-20.46514
MCNM1497	LGV161	MCN-M	guyannensis	-49.72934	-6.43589
MJ515		MZUSP	gardneri	-64.85128	-9.45414
MN56812	MNLM519	MN	guyannensis	-64.78914	1.2086
MN76750	MNLM2337	MN	guyannensis	-48.29032	-10.0782
MPEG20768	BDP2186	FMNH	longicaudatus	-61.93104	-10.84726
MPEG20769	BDP2177	FMNH	longicaudatus	-61.93104	-10.84726
MSB140110		MSB	decumanus	-80.11667	-3.88333
MSB140111		MSB	decumanus	-80.75	-1.38333
MSB208394		MSB	longicaudatus	-65.51084	-14.01672
MSB211792		MSB	longicaudatus	-66.13381	-11.03798
MSB236570		MSB	gardneri	-68.91681	-11.3501
MSB236689		MSB	goeldii	-67.56023	-11.49004
MSB236807		MSB	goeldii	-66.77966	-11.74947
MSB263513		MSB	*H. gymnurus*	-81.15074	8.53204
MSB45836		MSB	semispinosus	-84.7178	15.84003
MSB70574		MSB	longicaudatus	-65.55	-17.1
MSB99059		MSB	longicaudatus	-64.49139	-14.89556
MVZ157855	JLP8271	MVZ	longicaudatus	-69.068	-12.63333

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
MVZ157905		MVZ	longicaudatus	-77.751	-4.022
MVZ157968		MVZ	simonsi	-77.751	-4.022
MVZ160094	JLP9044	MVZ	guyannensis	-61.66667	4.46667
MVZ168955	EY631	MVZ	simonsi	-69.07289	-12.6
MVZ190699	JLP15922	MVZ	longicaudatus	-68.76672	-6.46666
MVZ194439	INPA3442	MVZ	longicaudatus	-72.78304	-8.66663
MVZ194493	MPEG28367; MNFS331	MVZ	longicaudatus	-70.75008	-6.83344
MVZ194511	MPEG25501; JUR297; MPEG28378	MVZ	echinothrix	-66.01666	-3.31666
MVZ194545	MPEG25512; MNFS857; MPEG28381	MVZ	gardneri	-68.9002	-6.58334
MVZ194567	MPEG25505; MNFS541; MPEG28385	MVZ	gardneri	-70.85003	-6.75
MVZ194582	MPEG25509; MNFS1166; MPEG28392	MVZ	gardneri	-72.78304	-8.66663
MVZ194874	MPEG28572; MNFS1507	MVZ	goeldii	-72.81662	-8.36666
MVZ194879	MPEG28575; MNFS1548	MVZ	goeldii	-72.81667	-8.36666
MVZ194909	MPEG28486; JUR68	MVZ	goeldii	-70.73359	-6.80001
MVZ194914	MPEG28491; JUR81	MVZ	goeldii	-70.73359	-6.80001
MVZ194987	MPEG28540; MNFS497	MVZ	goeldii	-70.75008	-6.83344
MVZ196095		MVZ	trinitatis	-73.9511	6.31417
MVZ225064		MVZ	semispinosus	-82.58217	8.46342
MVZ225082		MVZ	*H. gymnurus*	-79.92578	8.68753
MZUSP30370	UUPI412; CIT1465	CIT	guyannensis	-45.20264	-8.86342
MZUSP31924	APC176; CIT648	CIT	goeldii	-58.49231	-10.32276
MZUSP31942	M97032; CIT483	CIT	guyannensis	-54.87472	-11.50302
MZUSP31947	APC273; CIT714	CIT	guyannensis	-51.11932	-10.01433
NUTRIA289		CTA-UFES	*M. coypus*	NA	NA
ROM115116		ROM	trinitatis	-60.76667	5.33333
TTU106013	TK145304	TTU	guyannensis	-54.73945	4.26732
TTU101118	TK73760	TTU	simonsi	-73.26836	-4.02398
TTU101173	TK73888	TTU	longicaudatus	-73.26836	-4.02398
TTU101213	TK73977	TTU	longicaudatus	-73.26836	-4.02398
USNM549559		USNM	longicaudatus	-52.37	-3.65
USNM584593		USNM	longicaudatus	-61.0347	-14.7672

Table S1: Continuation.

Catalog Number	Alternative Number	Source	Species Group	Longitude	Latitude
X1M15		LMUSP	longicaudatus	-49.94611	-4.25168
X1M36		LMUSP	goeldii	-49.09108	-5.34573

Table S2: RAD processing results of the 82 samples used in the analyses, showing in the initial number and post-filters of reads, clusters, values of heterozygosity (H) and its error (HE), the final number of clusters after correction by heterozygosity, the number of loci retained In addition, there is information about the five main clades and 41 subclasses where the sample was recovered in the conservative tree from Chapter 1. Table in ***.csv*** format available on https://github.com/jdalapicolla.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained
ABX005	B	ABX	4,250,491	4,246,405	365,088	79,939	0.008	0.002	76,308	13,542
ABX020	B	ABX	2,723,645	2,721,069	310,633	74,213	0.007	0.002	71,460	13,422
AMCC112987	B	NVZ	3,039,552	3,036,575	387,817	70,208	0.006	0.002	67,263	6,496
AMCC112999	C	SVZ	1,493,735	1,492,180	188,042	47,498	0.008	0.001	45,740	4,426
AMCC114577	A	GJR	2,609,974	2,607,328	367,853	69,937	0.011	0.002	66,450	7,378
AMCC176080	B	NVZ	2,627,971	2,625,323	420,811	70,249	0.006	0.002	67,026	6,308
AMNH269123	D	NAR	4,709,606	4,704,845	440,479	81,567	0.009	0.002	76,914	13,064
AMNH272700	D	SSM	1,608,196	1,606,633	316,822	62,693	0.008	0.003	59,402	8,654
C247647	A	PTI	3,778,844	3,775,227	378,714	85,432	0.008	0.002	81,523	14,038
CIT393	D	MTR	2,300,707	2,298,561	374,711	80,194	0.007	0.002	76,732	13,581
CIT448	D	MTR	2,766,677	2,764,014	314,952	72,701	0.006	0.002	69,875	12,888
CTA1028	OUT	OUT	3,141,959	3,138,855	425,166	73,175	0.007	0.002	69,803	719
CTA1415	D	SCE	2,175,679	2,173,754	259,666	71,385	0.005	0.002	68,929	11,964
CTA1511	B	TXI	2,672,598	2,669,989	311,918	78,351	0.007	0.002	75,603	14,372
CTA1517	A	EAM	2,568,226	2,565,803	267,401	77,763	0.009	0.002	74,721	12,902
CTA4195	C	GUS	2,021,054	2,018,923	304,134	63,369	0.007	0.002	60,838	6,203

Table S2: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained
CTA4245	D	NAR	2,214,988	2,212,738	215,530	53,681	0.007	0.002	51,365	8,999
CTA4352	A	LMR	3,599,480	3,595,923	402,105	75,248	0.009	0.002	71,702	8,359
CTA4363	A	LMR	2,440,084	2,437,512	323,691	66,449	0.009	0.002	63,268	7,349
CTA4371	A	IJM	1,124,626	1,123,499	213,095	52,486	0.007	0.002	50,231	7,200
CTA4390	B	TXI	198,766	198,558	124,352	3,750	0.013	0.004	3,291	510
EEB1013	OUT	OUT	1,245,322	1,244,143	201,547	55,354	0.006	0.002	53,400	531
FMNH175255	D	UMD	408,058	407,613	65,103	17,918	0.006	0.002	17,249	2,187
FMNH175275	E	WAM	3,271,144	3,268,074	456,885	78,046	0.007	0.003	73,782	6,549
FMPS010	OUT	OUT	3,561,188	3,557,750	377,580	71,386	0.009	0.002	66,710	656
ICA095	A	IJM	2,819,142	2,816,301	399,671	75,502	0.010	0.002	71,120	10,149
ICA245	A	NSR	2,414,180	2,411,946	390,167	67,468	0.008	0.003	63,220	5,829
JAP012	A	NSR	1,507,100	1,505,662	240,879	57,940	0.007	0.002	55,455	5,117
JAP094	D	JAP	1,350,168	1,348,909	290,864	58,209	0.006	0.003	55,186	9,993
JAP107	D	JAP	1,476,001	1,474,595	352,015	61,519	0.006	0.003	57,817	10,001
JUF017	C	NRB	4,061,708	4,057,600	480,217	90,502	0.009	0.002	85,678	9,437
MBR046	OUT	OUT	1,674,854	1,673,033	238,582	58,226	0.006	0.002	55,716	1,069
MCNM1497	B	XAI	2,400,373	2,398,159	291,650	72,106	0.007	0.002	69,525	13,455
MJ515	A	JMI	2,144,733	2,142,758	319,340	65,331	0.009	0.003	61,958	6,845

Table S2: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained
MN56812	C	NRB	3,830,255	3,826,355	514,924	82,841	0.009	0.002	78,076	8,025
MN76750	B	CER	3,399,309	3,396,093	309,899	71,700	0.006	0.002	69,064	13,283
MPEG20768	D	UMR	3,871,135	3,867,308	471,710	81,980	0.010	0.002	77,284	11,730
MPEG20769	D	UMR	3,596,375	3,592,865	443,107	79,054	0.009	0.002	74,875	11,866
MSB140110	D	ECU	3,710,546	3,706,843	421,001	75,722	0.007	0.002	72,141	7,999
MSB140111	D	ECU	2,771,091	2,768,271	354,251	67,979	0.007	0.002	65,013	6,984
MSB208394	D	PTC	1,681,886	1,680,194	233,896	59,112	0.006	0.002	56,910	8,547
MSB211792	D	PAB	999,198	998,289	158,650	51,920	0.005	0.003	49,660	6,962
MSB236570	A	ATH	2,558,413	2,556,048	382,635	70,321	0.010	0.003	66,417	7,206
MSB236689	A	MMR	806,223	805,423	153,452	48,627	0.007	0.003	46,109	7,298
MSB236807	A	MMR	1,131,093	1,130,084	247,090	53,394	0.007	0.003	50,398	7,987
MSB263513	HG	HG	1,774,930	1,773,026	372,885	65,644	0.007	0.002	61,894	4,156
MSB45836	D	SEM	1,495,950	1,494,764	239,216	54,149	0.004	0.002	52,318	5,462
MSB70574	D	SPB	3,163,614	3,160,405	411,678	75,158	0.008	0.002	71,413	10,435
MSB99059	D	PTC	1,574,661	1,573,182	321,540	59,536	0.006	0.003	56,569	8,347
MVZ157855	D	SPB	3,537,288	3,533,940	385,874	74,723	0.007	0.003	70,699	10,693
MVZ157905	D	NSM	2,862,935	2,860,277	406,759	74,530	0.009	0.003	70,490	10,404
MVZ157968	E	STR	184,713	184,529	90,165	7,409	0.007	0.004	6,530	515

Table S2: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained
MVZ160094	C	SVZ	3,725,061	3,721,306	411,353	78,198	0.010	0.001	74,412	7,588
MVZ168955	E	WAM	3,325,897	3,322,862	419,611	77,869	0.008	0.003	73,791	6,494
MVZ190699	D	JUR	2,215,186	2,213,146	394,036	70,819	0.007	0.003	66,846	12,234
MVZ194439	D	SSM	1,898,970	1,897,183	327,530	68,419	0.008	0.003	64,717	9,991
MVZ194493	D	JUR	941,957	941,104	240,458	50,590	0.005	0.003	47,861	8,922
MVZ194511	A	SSR	760,766	760,024	97,646	21,155	0.007	0.002	20,120	1,413
MVZ194545	A	JMI	2,183,678	2,181,345	229,409	62,652	0.011	0.002	59,810	6,804
MVZ194567	A	GJR	1,884,446	1,882,495	325,434	66,022	0.011	0.002	62,725	6,899
MVZ194582	A	ATH	839,115	838,414	159,305	48,548	0.010	0.002	46,251	4,916
MVZ194874	A	UJR	2,626,650	2,624,245	474,978	79,458	0.011	0.003	73,690	11,210
MVZ194879	A	UJR	1,759,530	1,757,966	312,783	63,610	0.010	0.003	59,708	9,350
MVZ194914	A	CLJ	3,147,232	3,144,201	473,314	80,987	0.009	0.003	75,571	11,270
MVZ194987	A	CLJ	2,649,630	2,647,221	439,430	78,430	0.009	0.003	73,465	12,136
MVZ196095	D	MAR	151,445	151,231	72,935	5,230	0.006	0.004	4,835	376
MVZ225064	D	SEM	3,403,305	3,399,963	476,625	80,606	0.007	0.002	76,597	8,414
MVZ225082	HG	HG	3,673,839	3,670,151	455,668	84,417	0.008	0.002	80,050	4,975
MZUSP30370	B	CER	3,192,286	3,189,447	301,742	73,339	0.005	0.002	70,824	13,895
MZUSP31924	A	PTI	2,895,615	2,892,827	260,567	72,153	0.008	0.002	69,010	12,102

Table S2: Continuation.

Samples	Clades	Subclades	Initial Reads	Filtered Reads	Initial Clusters	Filtered Clusters	H	HE	Final Clusters	Loci Retained
MZUSP31942	B	TXI	2,980,107	2,977,260	300,271	74,560	0.007	0.002	71,840	13,763
MZUSP31947	B	TXI	2,852,271	2,849,567	309,353	73,524	0.006	0.002	71,000	13,688
NUTRIA289	OUT	OUT	2,242,635	2,240,340	328,132	68,584	0.005	0.002	65,852	1,382
ROM115116	TEP	TEP	238,801	238,591	70,571	10,508	0.008	0.002	9,885	394
TTU101118	E	IQT	1,836,157	1,834,449	286,697	63,188	0.007	0.003	60,405	5,171
TTU101173	D	LOR	1,636,788	1,635,185	352,978	60,685	0.005	0.003	57,456	10,330
TTU101213	D	NSM	1,915,489	1,913,686	378,740	68,192	0.008	0.003	64,245	9,941
TTU106013	C	GUS	2,039,625	2,037,504	294,435	62,661	0.006	0.002	59,997	6,211
USNM549559	D	SAR	1,413,419	1,412,033	279,142	54,892	0.005	0.003	52,105	9,319
USNM584593	D	SCE	2,479,653	2,477,161	337,134	66,992	0.008	0.001	64,120	9,499
X1M15	D	SAR	1,357,606	1,356,316	198,154	62,148	0.005	0.002	60,088	11,848
X1M36	A	EAM	2,394,628	2,392,385	319,232	71,479	0.007	0.002	68,657	11,687

Table S3: Divergence times in millions years ago (Ma) for the clades of the *Proechimys* phylogeny, with mean, median, and 95% confidence interval (C.I.) values. Dated tree is shown in Fig. S20. A = clade A; B = clade B; C = clade C; D = clade D; E = clade E; HG = *Hoplomys gymnurus;* PRO = *Proechimys* "core" (except for Tepui individual); TEP = Tepui individual.

Clades	Mean	Median	95% C.I.	
			Lower	Upper
Clyomys+Euryzygomatomys	7.14	7.02	7.78	6.81
Clyomys+Euryzygomatomys+Trinomys	13.01	12.41	19.02	8.14
Root	18.69	17.83	28.61	10.10
Myocastor+Thrichomys+HG+TEP+PRO	17.67	16.88	27.94	9.57
Thrichomys+HG+TEP+PRO	17.14	16.49	27.36	9.65
HG+TEP+PRO	12.68	12.12	19.94	7.05
HG+TEP	8.16	7.68	14.87	2.99
HG	0.50	0.45	1.00	0.11
PRO	11.00	10.56	17.01	5.98
A+B+C	10.34	9.90	15.82	5.26
A+B	10.14	9.69	15.46	5.20
A	8.20	7.82	12.89	4.48
B	8.73	8.39	13.44	4.44
C	2.55	2.40	4.41	1.10
D+E	10.53	10.07	16.62	5.79
D	8.73	8.32	13.52	4.30
E	3.53	3.35	5.64	1.47
STR+WAN	1.80	1.65	3.30	0.61
WAN	0.70	0.64	1.24	0.22
MAR+ECU+CAM	5.50	5.17	9.52	2.44
ECU+CAM	3.96	3.79	6.16	1.70
ECU	1.81	1.68	2.97	0.73
CAM	1.65	1.51	2.89	0.64

Table S3: Continuation.

Clades	Mean	Median	95% C.I.	
			Lower	Upper
SSM+NSM+SPB+UDM+PTC+SCE+PAB+NAR+SAR+ JAP+JUR+LOR+MTR+UMR	6.78	6.43	10.49	3.45
SSM+NSM+SPB+UDM+PTC+SCE+PAB	5.91	5.56	8.98	2.72
NAR+SAR+JAP+JUR+LOR+MTR+UMR	4.68	4.46	7.25	2.36
SSM+NSM+SPB+UDM	4.53	4.29	7.19	2.28
SSM+NSM	2.11	2.01	3.36	1.05
SSM	0.76	0.71	1.32	0.31
NSM	0.74	0.69	1.28	0.30
SPB+UDM	3.61	3.44	5.77	1.69
SPB	0.78	0.73	1.41	0.30
NAR+SAR+JAP+JUR+LOR	3.28	3.12	5.06	1.53
MTR+UMR	1.59	1.50	2.63	0.74
MTR	1.38	1.30	2.35	0.65
UMR	0.42	0.39	0.76	0.16
NAR+SAR	0.90	0.85	1.49	0.40
NAR	0.47	0.43	0.84	0.12
SAR	0.35	0.32	0.66	0.13
JAP+JUR+LOR	1.55	1.47	2.54	0.76
JAP+JUR	0.79	0.74	1.29	0.34
JAP	0.30	0.28	0.54	0.11
JUR	0.15	0.14	0.32	0.04
PTC+SCE+PAB	2.94	2.78	4.69	1.42
PTC+SCE	1.02	0.96	1.66	0.50
PTC	0.55	0.51	0.94	0.22
SCE	0.64	0.60	1.16	0.26
NBR+GUS	1.22	1.05	1.92	0.51
NBR	0.63	0.58	1.14	0.24
GUS	0.38	0.35	0.73	0.11
SVZ	0.44	0.39	0.90	0.09
NVZ	0.88	0.80	1.59	0.30
CER+XAI+TXI+LRX+ABX	3.51	3.32	5.50	1.58

Table S3: Continuation.

Clades	Mean	Median	95% C.I.	
			Lower	Upper
CER+XAI+TXI+LRX	2.27	2.12	3.77	1.12
TXI+LRX	1.25	1.17	2.43	0.24
CER+XAI	0.91	0.86	1.47	0.43
CER	0.35	0.32	0.63	0.11
XAI	0.60	0.56	1.02	0.22
TXI	0.28	0.26	0.55	0.08
ABX	0.22	0.19	0.49	0.06
NSR+SSR	2.11	1.92	3.38	0.89
NSR	0.83	0.78	1.50	0.27
PTI+CLJ+MMR+UJR+EAM+IJM+JMI+ATH+GJR+LMR	6.79	6.49	10.42	3.67
JMI+ATH+GJR+LMR	3.78	3.58	5.95	1.89
JMI+ATH	2.23	2.11	3.64	1.01
JMI	0.97	0.91	1.68	0.39
ATH	0.63	0.59	1.14	0.20
GJR+LMR	2.33	2.21	3.91	1.19
GJR	1.04	0.96	1.85	0.44
LMR	0.34	0.31	0.63	0.10
PTI+CLJ+MMR+UJR+EAM+IJM	3.96	3.72	6.26	2.02
PTI+CLJ+MMR+UJR+EAM	3.34	3.13	5.20	1.65
PTI+CLJ+MMR+UJR	2.78	2.61	4.41	1.29
CLJ+MMR+UJR	1.33	1.24	2.12	0.64
MMR+UJR	0.96	0.89	1.61	0.39
MMR	0.36	0.32	0.71	0.12
UJR	0.45	0.41	0.80	0.13
CLJ	0.26	0.23	0.52	0.05
PTI	0.22	0.19	0.46	0.04
EAM	0.81	0.77	1.39	0.29
IJM	1.19	1.12	2.00	0.37

Table S4: Probability of occupancy per area for each clade in the *Proechimys* phylogeny, using the Morrone biogeographic dominions (DOM) for the Neotropical region. For areas: MES = Mesoamerican dominion; PAC = Pacific dominion; BBR = Boreal Brazilian dominion; SBR = South Brazilian dominion; SEA = South-eastern Amazonian dominion; CHA = Chaco dominion; and PAR = Parana dominion. For clades: A = clade A; B = clade B; C = clade C; D = clade D; E = clade E; CLY = *Clyomys*; EUR = *Euryzygomatomys*; HG = *Hoplomys gymnurus;* MYO = *Myocastor*; PRO = *Proechimys* "core" (except for Tepui individual); TEP = Tepui individual; TRI = *Trinomys*; THR = *Thrichomys*. Bold numbers indicate probabilities greater than 20%.

Clades	MES	PAC	BBR	SBR	SEA	CHA	PAR
MYO+THR+HG+TEP+PRO	0.01	0.01	0.02	**0.87**	0.03	0.12	0.00
THR+HG+TEP+PRO	0.01	0.01	0.03	**0.86**	0.03	0.08	0.00
HG+TEP+PRO	0.03	0.04	0.08	**0.86**	0.00	0.00	0.00
PRO	0.00	0.00	0.04	**0.97**	0.00	0.00	0.00
A+B+C	0.00	0.01	0.09	**0.91**	0.00	0.00	0.00
A+B	0.00	0.01	0.02	**0.97**	0.01	0.00	0.00
A	0.00	0.00	0.01	**0.99**	0.00	0.00	0.00
ATH+JMI+GJR+LMR+CLJ+MMR+UJR+PTI+EAM+IJM	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
ATH+JMI+GJR+LMR	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
ATH+JMI	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
GJR+LMR	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
CLJ+MMR+UJR+PTI+EAM+IJM	0.00	0.00	0.01	**0.99**	0.01	0.00	0.00
CLJ+MMR+UJR+PTI+EAM	0.00	0.00	0.00	**0.88**	0.14	0.00	0.00
CLJ+MMR+UJR+PTI	0.00	0.00	0.00	**0.93**	0.06	0.01	0.00
CLJ+MMR+UJR	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
MMR+UJR	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
NSR+SSR	0.00	0.00	0.04	**0.96**	0.00	0.00	0.00
B	0.00	0.06	0.01	**0.90**	0.03	0.00	0.00
CER+XAI+TXI+LRX+ABX	0.00	0.00	0.02	**0.93**	0.11	0.01	0.00
CER+XAI+TXI+LRX	0.00	0.00	0.07	0.00	**0.91**	0.08	0.00

Table S4: Continuation.

Clades	MES	PAC	BBR	SBR	SEA	CHA	PAR
CER+XAI	0.00	0.00	0.10	0.00	**0.83**	0.11	0.00
TXI+LRX	0.00	0.00	0.00	0.00	**0.97**	0.03	0.00
C	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
NRB+GUS	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
D+E	0.00	0.01	0.00	**1.00**	0.00	0.00	0.00
D	0.00	0.03	0.00	**0.97**	0.00	0.00	0.00
CAM+ECU+MAR	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
CAM+ECU	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
JAP+JUR+LOR+SAR+NAR+MTR+UMR+SSM+ NSM+SPB+UDM+PAB+PTC+SCE	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
JAP+JUR+LOR+SAR+NAR+MTR+UMR	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
JAP+JUR+LOR+SAR+NAR	0.00	0.00	0.04	**0.95**	0.02	0.00	0.00
JAP+JUR+LOR	0.00	0.00	0.02	**0.98**	0.00	0.00	0.00
JAP+JUR	0.00	0.00	0.05	**0.95**	0.00	0.00	0.00
SAR+NAR	0.00	0.00	**0.52**	0.00	**0.51**	0.00	0.00
MTR+UMR	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
SSM+NSM+SPB+UDM+PAB+PTC+SCE	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
SSM+NSM+SPB+UDM	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
SSM+NSM	0.00	0.00	0.03	**0.97**	0.00	0.00	0.00
SPB+UDM	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
PAB+PTC+SCE	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
PTC+SCE	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
E	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
WAM+STR	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
TEP+HG	**0.32**	**0.32**	**0.54**	0.00	0.00	0.00	0.00
CLY+EUR+TRI	0.00	0.00	0.00	0.00	0.00	**0.90**	**0.21**
CLY+EUR	0.00	0.00	0.00	0.00	0.00	**0.93**	0.07

Table S5: Probability of occupancy per area for each clade in the *Proechimys* phylogeny, using the areas of endemism (AOE) for the Amazon. For areas: BEL = Belem; GUY = Guyana; IME = Imeri; INA = Inambari; JAU = Jau; NAP = Napo; RON = Rondonia; TAP = Tapajos; XGU = Xingu, NOR = North to the Amazon; SOU = South to the Amazon. For clades: A = clade A; B = clade B; C = clade C; D = clade D; E = clade E; CLY = *Clyomys*; EUR = *Euryzygomatomys*; HG = *Hoplomys gymnurus;* MYO = *Myocastor*; PRO = *Proechimys* "core" (except for Tepui individual); TEP = Tepui individual; TRI = *Trinomys*; THR = *Thrichomys*. Bold numbers indicate probabilities greater than 20%.

Clades	NOR	GUY	IME	NAP	JAU	INA	RON	TAP	XGU	BEL	SOU
MYO+THR+HG+TEP+PRO	0.08	0.07	0.00	0.02	0.00	0.03	0.00	0.00	0.12	0.00	**0.82**
THR+HG+TEP+PRO	0.11	0.09	0.00	0.02	0.00	0.04	0.01	0.00	0.15	0.00	**0.74**
HG+TEP+PRO	**0.45**	**0.37**	0.00	0.04	0.00	0.11	0.01	0.01	0.01	0.00	0.05
PRO	**0.36**	**0.28**	0.00	0.09	0.00	**0.24**	0.02	0.01	0.01	0.01	0.05
A+B+C	**0.30**	**0.39**	0.00	0.00	0.01	**0.25**	0.03	0.02	0.01	0.02	0.06
A+B	**0.39**	0.00	0.00	0.00	0.02	**0.43**	0.07	0.05	0.02	0.03	0.07
A	0.00	0.00	0.00	0.00	0.04	**0.96**	0.00	0.00	0.00	0.00	0.00
ATH+JMI+GJR+LMR+CLJ+MMR+UJR+PTI+EAM+IJM	0.00	0.00	0.00	0.00	0.01	**0.99**	0.00	0.00	0.00	0.00	0.00
ATH+JMI+GJR+LMR	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
ATH+JMI	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
GJR+LMR	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
CLJ+MMR+UJR+PTI+EAM+IJM	0.00	0.00	0.00	0.00	0.01	**0.98**	0.00	0.01	0.00	0.00	0.00
CLJ+MMR+UJR+PTI+EAM	0.00	0.00	0.00	0.00	0.00	**0.86**	0.00	0.09	0.06	0.00	0.00
CLJ+MMR+UJR+PTI	0.00	0.00	0.00	0.00	0.00	**0.94**	0.01	0.04	0.00	0.00	0.01
CLJ+MMR+UJR	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
MMR+UJR	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
NSR+SSR	0.00	0.00	0.00	0.00	0.08	**0.92**	0.00	0.00	0.00	0.00	0.00

Table S5: Continuation.

Clades	NOR	GUY	IME	NAP	JAU	INA	RON	TAP	XGU	BEL	SOU
B	**0.59**	0.00	0.00	0.00	0.00	0.00	0.15	0.09	0.05	0.07	0.11
CER+XAI+TXI+LRX+ABX	0.00	0.00	0.00	0.00	0.00	0.00	**0.38**	**0.22**	0.11	0.16	**0.20**
CER+XAI+TXI+LRX	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.49**	**0.25**	**0.37**	**0.39**
CER+XAI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.36**	**0.51**	**0.53**
TXI+LRX	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00
C	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
NRB+GUS	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
D+E	**0.36**	0.03	0.00	**0.31**	0.00	**0.30**	0.02	0.01	0.01	0.00	0.01
D	**0.46**	0.03	0.00	0.11	0.01	**0.35**	0.04	0.02	0.01	0.00	0.01
CAM+ECU+MAR	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CAM+ECU	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
JAP+JUR+LOR+SAR+NAR+MTR+UMR+SSM+NSM+SPB+UDM+PAB+PTC+SCE	0.00	0.05	0.00	0.13	0.02	**0.66**	0.11	0.06	0.04	0.00	0.02
JAP+JUR+LOR+SAR+NAR+MTR+UMR	0.00	0.07	0.00	0.16	0.04	**0.38**	**0.23**	0.13	0.08	0.00	0.00
JAP+JUR+LOR+SAR+NAR	0.00	0.12	0.00	**0.24**	0.09	**0.42**	0.00	**0.21**	0.18	0.00	0.00
JAP+JUR+LOR	0.00	0.00	0.00	**0.42**	**0.20**	**0.51**	0.00	0.00	0.00	0.00	0.00
JAP+JUR	0.00	0.00	0.00	0.00	**0.41**	**0.69**	0.00	0.00	0.00	0.00	0.00
SAR+NAR	0.00	**0.28**	0.00	0.00	0.00	0.00	0.00	**0.47**	**0.44**	0.00	0.00
MTR+UMR	0.00	0.00	0.00	0.00	0.00	0.00	**0.93**	0.07	0.00	0.00	0.00
SSM+NSM+SPB+UDM+PAB+PTC+SCE	0.00	0.00	0.00	0.01	0.00	**0.97**	0.00	0.00	0.00	0.00	0.03
SSM+NSM+SPB+UDM	0.00	0.00	0.00	0.01	0.00	**0.99**	0.00	0.00	0.00	0.00	0.00
SSM+NSM	0.00	0.00	0.00	0.05	0.00	**0.95**	0.00	0.00	0.00	0.00	0.00
SPB+UDM	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00
PAB+PTC+SCE	0.00	0.00	0.00	0.00	0.00	**0.93**	0.00	0.00	0.00	0.00	0.07
PTC+SCE	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**
E	0.00	0.00	0.00	**0.80**	0.00	**0.20**	0.00	0.00	0.00	0.00	0.00
WAM+STR	0.00	0.00	0.00	**0.77**	0.00	**0.23**	0.00	0.00	0.00	0.00	0.00
TEP+HG	**0.56**	**0.48**	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CLY+EUR+TRI	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**
CLY+EUR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**1.00**

Figures

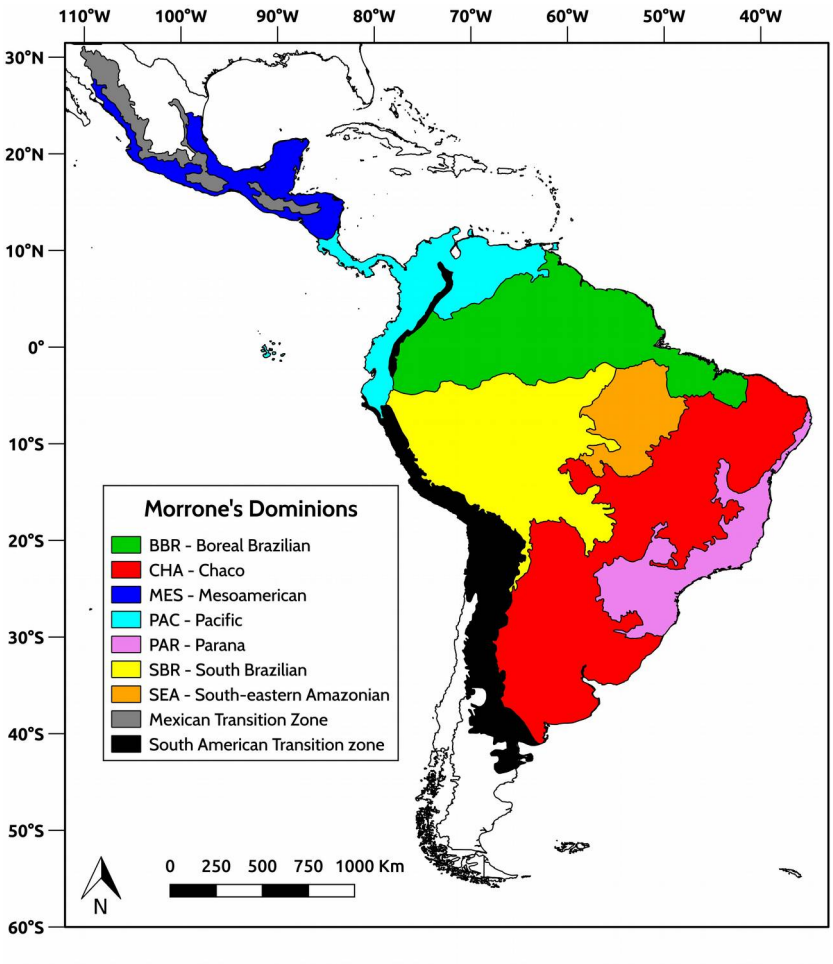


Figure S1: Seven dominions used in the DOM analysis of ancestral areas estimation and two transition zones for the Neotropical region according to Morrone's classification.

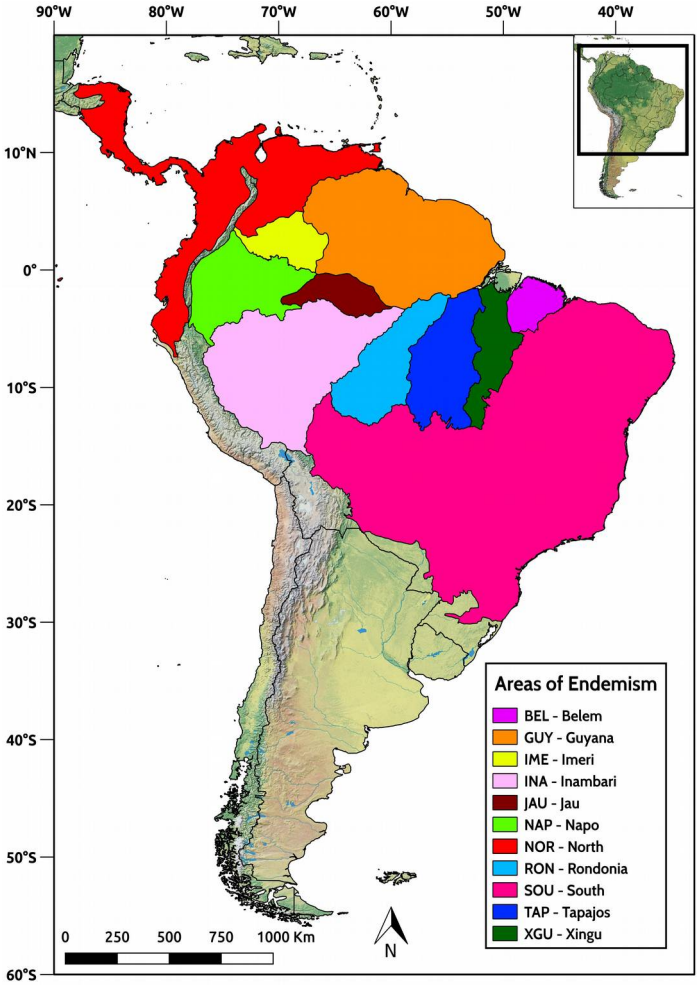


Figure S2: Eleven areas representing the AOE areas areas for ancestral areas estimation, according with nine classical areas of endemism.

4. SAME ORIGINS, DIFFERENT HISTORIES: CONTRASTING THE PHYLOGEOGRAPHIC PATTERNS OF THREE SYMPATRIC SPECIES OF SPINY RATS FROM WESTERN AMAZON

Abstract

Western Amazon is one of the least studied and most diverse areas of the Neotropical region, leading the evolutionary history of its elements in most cases unknown. In order to contribute to the understanding of this history, I studied how geographic space and microhabitat segregation influenced the genetic structure of three sympatric species of the genus *Proechimys* in this region: *P. brevicauda*, *P. simonsi*, and *P. steerei*. These species are morphologically similar but occupy different habitats (*P. brevicauda* and *P. simonsi* inhabit non-flooded forest – *terra-firme* forests – and *P. steerei* seasonal floodplain forests – *várzea* forests), thus it is possible that the existence of segregation at the micro-habitat level led to divergent genetic patterns between them. In this way, I test the hypothesis that sympatric species would have the same phylogeographic patterns due to the common history of geographic space, or alternatively, that micro-habitat segregation and local adaptation would have shaped different genetic structure patterns. I sequenced and identified unlinked SNP (single nucleotide polymorphism) employing ddRAD-Sequencing of 52 individuals (17 of *P. brevicauda*, with 5,050 SNP identified; *P. simonsi*, n = 20, SNP = 4,629; *P. steerei* n = 15, SNPs = 5.819). I estimated the diversity and genetic structure for each species. I tested isolation models by distance, by environmental heterogeneity, and by barriers (areas of endemism, structural arches, ecoregions and pleistocenic refuges by ecological niche models) to verify which model would best explain the pattern of genetic diversity and whether it was the same among sympatric species. In addition, I calculated the overlap of climatic and morphological hypervolumes to evaluate whether there would be any microhabitat segregation. Each species presented a different genetic structure pattern, and different models and barriers explained these patterns, with little importance for isolation by distance. Climatic hypervolumes showed a large overlap and similarity while the opposite was recovered by morphological hypervolumes. These results may indicate that local adaptation, microhabitat differences, and species-traits may be more important than the common geographic space to explain genetic variation in sympatric species.

Keywords: Ecological niche models; Isolation by distance; Isolation by barriers; Isolation by Environmental; Hypervolume

4.1. Introduction

Phylogeography proposes to understand the geographical patterns of the genetic diversity of a species over time and space (Avise, 2009; Hickerson *et al.*, 2010). Genetic diversity is usually associated with reduced gene flow between populations and their consequent isolation (Hartl & Clark, 1997), that can occur due to (*i*) geographic distance that limits the connection between distant populations and prioritizes connections between neighboring populations, a phenomenon known as Isolation by Distance (Wright, 1943); (*ii*) historical processes, such as vicariance and dispersion events that could have connected or separated these populations in the past, also known as Isolation by Barriers (Carnaval & Moritz, 2008; Ribas *et al.*, 2012); (*iii*) ecological and behavioral traits such as mating patterns, migration capacity, habitat use, environmental heterogeneity that may lead to Isolation by Environment and local adaptation of populations (Lee & Mitchell-Olds, 2011; Wang & Bradburd, 2014).

Since variation in the genetic structure is related to geographic features and to environmental changes, phylogeographic studies, in addition to providing hypothesis regarding the evolutionary history of target species, they also allow to advance hypothesis on the environmental changes that have occurred in the areas where these species inhabit (Carnaval *et al.*, 2009; Thome *et al.*, 2010; Costa & Leite, 2012; Ribas *et al.*, 2012). In addition, understanding the evolutionary history of a group as well as the processes of diversification of the region to which these organisms belong are essential for the sustainable use of the region and for the conservation strategies of the species (Moritz & Faith, 1998; Moritz, 2002).

One of the least studied and most diverse areas of the Neotropical region (Finer *et al.*, 2008), with major geological changes in the last millions of years is the Western Amazon (Hoorn *et al.*, 2010). This area is bordered by the Andes on the west and by the rivers Negro and Madeira on the east (Leite & Rogers, 2013), and showed a dynamic geological history with many sedimentary deposits during the Cenozoic, influenced by the Andes uplift (Vonhof, Wesselingh, & Ganssen, 1998; Hoorn *et al.*, 2010; Mora *et al.*, 2010). Western Amazon also presents differences on climatic characteristics, such as greater precipitations than southern and eastern Amazonia (Costa & Foley, 1998). Several phylogeographic hypotheses try to explain the patterns of diversification in the Amazon (Bonvicino & Weksler,

2012; Leite & Rogers, 2013), and also in the Western Amazon, and they typically invoke historical and ecological processes associated to environmental changes. The hypotheses most commonly used to explain the diversification in the Western Amazon are rivers as barriers (Lougheed *et al.*, 1999; Gascon *et al.*, 2000; Patton, da Silva, & Malcolm, 2000), Pleistocene climatic oscillations (Haffer, 1969, 1997), and structural archs (Da Silva & Patton, 1998; Patton, Da Silva, & Malcolm, 2000; Hubert & Renno, 2006).

One of the groups that contribute to this high biodiversity in Western Amazon is the genus *Proechimys* J. A. ALLEN, a terrestrial spiny rat of family Echimyidae. It has a wide distribution in the Neotropics, with its known 22 species occurring from Central America to the Brazilian Cerrado, covering the entire Amazon region (Woods & Kilpatrick, 2005; Patton & Leite, 2015). In the Western Amazon, nine species of *Proechimys* are known, namely *P. brevicauda*, *P. cuvieri*, *P. echinothrix, P. gardneri*, *P. kulinae*, *P. pattoni*, *P. quadruplicatus*, *P. simonsi,* and *P. steerei* (Patton & Leite, 2015; Fabre, Patton, & Leite, 2016) with records of sympatry and syntopy among two to five species of the genus (Malcolm, 1992; Patton *et al.*, 2000; Steiner, Sourrouille, & Catzeflis, 2000).

There are few specific studies on how these sympatric species share the habitat. Emmons (1982) suggested that the high population densities of two sympatric species of *Proechimys* from southern Peru, *P. simonsi* and *P. steerei*, may be associated with the high density of food in mature forests with probable female territoriality, while a third sympatric species in the study area, *Proechimys brevicauda*, presented low-densities populations, with females showing smaller home ranges than males with exclusive-use, similar to a polygynous mating system (Patton & Leite, 2015). Patton et al. (2000) and Voss, Lunde, & Simmons (2001) indicated that sympatric *Proechimys* species are segregated at the micro-habitat level due to differences in body size, occupied area within the biome, or habitat use. Matocq, Patton, & da Silva (2000) used two sympatric species from the Western Amazon: *Proechimys steerei*, which occurs in the seasonal floodplain forests (*várzea* forests) with a more r-selected strategy for reproduction, and *Proechimys simonsi* that inhabits non-flooded forests (*terra-firme* forests) and present a k-selection strategy, on a study designed to evaluate the their population genetics along the Rio Juruá. They concluded that both species have similar genetic diversity, but *P. simonsi* showed a lower gene flow and more genetic structure among the populations, even on the same river bank, the opposite than they expected. Seasonal flooded habitats are more unstable, and a small genetic diversity and population sizes in *P.*

steerei was expected because recurrent events of extinction and colonization should lead to the reduction of genetic diversity and bottleneck effects. This may indicate that another factor, besides the occupied area, would have shaped the genetic structure of the sympatric species in the Western Amazon.

Here I intend to study three sympatric species of *Proechimys* from Western Amazon, namely *Proechimys brevicauda*, *P. simonsi*, and *P. steerei,* as models to assess how geographic space and environmental variation may have influenced genetic structure in a broader scale than Matocq, Patton, & da Silva (2000). I selected these three species because in the previous chapters I have been able to recover some phylogenetic and biogeographical information about these taxa. They were recovered in a different clades from *Proechimys* phylogeny (*P. brevicauda* in clade D, *P. simonsi* in clade E, and *P. steerei* in clade A; Chapter 2), and present, at some level, distinct evolutionary histories. They share a similar biogeographic history without major shifts in geographic distribution, as the range evolution models showed that the Western Amazon was the most probable ancestral area for the three species, and the divergence times was similar between *P. simonsi* and *P. steerei* and a little older for *P. brevicauda* (Chapter 3). Moreover, *Proechimys brevicauda*, *P. simonsi*, and *P. steerei* are morphologically similar and endemic to the Western Amazon, but occupy different habitats, as while *P. brevicauda* and *P. simonsi* inhabit non-flooded forests, and *P. steerei* occurs at seasonal floodplain forests (Patton & Leite, 2015).

Therefore, I employed SNP from genomic data to test the hypothesis that sympatric species would have the same phylogeographic patterns due to the sharing of the same geographic space during their evolutionary history. If the history of the geographic space in which the species inhabit is more important for genetic diversity than local adaption and microhabitat segregation, I expect similar phylogeographic patterns with same diversification models to explain species genetic structure. On the other hand, if the species-traits, such as, the differential habitat use, population densities or phylogenetic distance, *i.e.* microhabitat segregation, have played a main role on genetic diversity, I expect different phylogeographic patterns with different models to explain them among the species.

Thus, my objectives were (*i*) to calculate and compare genetic diversity and structure among sympatric species; (*ii*) to test which model could better explain the pattern of genetic variation in species: isolation by distance, isolation by barriers, or isolation by environment

(*iii*) to compare if the same models may explain the genetic patterns in different species; and (*iv*) to estimate the similarity and overlap between climatic and morphological hypervolumes among the sympatric species to verify the existence of segregation in these microhabitats as suggested by Emmons (1982), Patton et al. (2000) and Voss et al. (2001).

4.2. Material and Methods

4.2.1. Genomic data

I used SNPs of 52 individuals for three species of *Proechimys* co-distributed in the Western Amazon, even with sympatry records, and with different ecological requirements: *P. brevicauda* (n = 17), and *P. simonsi* (n = 20) inhabit non-flooded forests (terra-firme forest) and *P. steerei* (n = 15) occupies the seasonal floodplain forests (várzea forests) (Fig. 1; Supplementary Material: Table S01). Specimens was identified by morphology following Patton & Leite, 2015), and they covered the known geographical distribution of the species, and populations were defined according to the locality, with two individuals as minimum sample size per population (Fig. 1).

I extracted the DNA using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA), and quantified it by Qubit fluorometer (Life Technologies, Grand Island, NY, USA). I prepared one library with all the samples, using Peterson et al. (2012) protocol and 300 ng of initial DNA. DNA was fragmented by two restriction enzymes (*Eco*-RI and *Mse*-I) and unique barcodes for each sample were ligated together with Illumina adapters for sequencing. Then, I selected DNA fragments automatically with a range of 350-450 bp in the Pippin-Prep (Sage Science, Beverly, MA, USA) and we amplified them in a PCR. I sequenced the 150 pb single-end reads on one lane in HiSeq2500 (Illumina, San Diego, CA, USA), following the manufacturers' instructions in the Hospital for Sick Children (Toronto, Ontario, Canada). More details about DNA extraction and library preparation is available in Appendix A, and at: https://github.com/jdalapicolla.

I used STACKS 1.45 pipeline (Catchen *et al.*, 2013) to process raw sequenced reads for each species separately. In this pipeline samples were demultiplexed, following the

barcodes, reads with low quality, more than 2 mismatches per barcode, or with uncalled nucleotides (Ns) were removed. The adapters and barcodes were also removed and the 140 bp reads were stacked and aligned without a reference genome (*i.e.*, *de novo* alignment). We allowed only stacks with 6 or more reads (parameter *-m*) and 3 nucleotides of distance between stacks (*-M*) for the creation of putative loci (Paris, Stevens, & Catchen, 2017). A catalog with all loci of all individuals was made with up to 3 fixed differences expected between individuals (*-n*) (Paris *et al.*, 2017). Populations of each species were defined (Fig. 1; Table S1) and only the loci present in at least two populations were considered for the output. After this step the data were filtered, removing the segregating sites and individuals and loci with no missing data. For the analyses I use one SNP per locus randomly chosen (*--write_random_snp*). Scripts of the analyses and more details about the sequencing analysis is in Appendix B and available at: https://github.com/jdalapicolla.

4.2.2. Distributional data

Occurrence points for the target species were obtained in visits to Brazilian and American collections, using only specimens that I identified based on morphology following Patton & Leite (2015). I also used some localities based on genetic data provided by Schetino (2008). A database with 468 occurrence points was created (*P. brevicauda*, n = 182; *P. simonsi*, n = 166; *P. steerei*, n = 120). I verified the localities using Google Maps (http://maps.google.com.br), two tools from SpeciesLink (http://splink.cria.org.br/): "geoLoc" and "infoXY", and ornithological gazetteers of the Neotropics (Paynter, 1982, 1992, 1993, 1997; Stephens & Traylor, 1983, 1985; Paynter & Traylor, 1991a,b).

4.2.3. Environmental data

I used environmental variables representing the past: Last Interglacial [LIG; about 120-140 thousands years ago (Ka)], Last Glacial Maximum (LGM; about 22 Ka), Middle Holocene (about 6 Ka), and the present-day. I used 19 environmental variables based in precipitation and temperature during 1960-1990, available in WorldClim 1.4

(www.worldclim.com) (Hijmans *et al.*, 2005) with resolution of 2.5 arc-seconds, around 5 Km^2 each cell (Table S2). For the environmental variables from the past I used the MPI-ESM-P simulations for the Coupled Model Intercomparison Project phase 5 (CMIP5) for LGM and Middle Holocene, and I used the environmental variables from (Otto-Bliesner *et al.*, 2006) for LIG, all available in WorldClim 1.4 (www.worldclim.com) (Hijmans et al., 2005).

4.2.4. Morphological data

I used 135 adults specimens, age classes 8, 9, and 10 from Patton & Rogers (1983) to estimate morphological similarities among sympatric species (*P. brevicauda,* n = 36; *P. simonsi,* n = 59; and *P. steerei,* n = 40; see Table S3). All individuals were identified at the specific level following Patton & Leite (2015). I took 22 cranial measurements (Fig. S1) with a digital caliper (precision: 0.01 mm), with no missing data. I eliminated outliers, transformed to logarithm, and standardized the variables to perform the analyses.

4.2.5. Genetic structure

To understand how the populations are genetically structured I calculate nucleotide diversity (π), Wright's inbreeding coefficient (F_{IS}), observed heterozygosity (H_{obs}), percentage of polymorphic sites (%pol), and F_{ST} (SNP-based F statistics) in the *populations* function STACKS 1.45 pipeline (Catchen *et al.*, 2013). In addition, I performed a Principal Component Analysis (PCA) with SNPs using *dudi.pca* function in the R package "adegenet" and (Jombart & Ahmed, 2011) for each species to identify possible clusters. I also calculated the fixation indices of each species by pairwise populations and globally: Nei's G_{ST} (Nei & Chesser, 1983), Hedrick's G''_{ST} (Hedrick, 2005), and Jost's *D* (Jost, 2008), and only globally the Meirmans' Φ_{ST} *(Meirmans, 2006)*, using "mmod" (Winter, 2012) and the functions: *diff_stats*, *pairwise_Gst_Nei*, *pairwise_Gst_Hedrick*, and *pairwise_D.* All analyses were performed in R 3.4.4 (R-Development CoreTeam, 2018).

4.2.6. Isolation by distance (IBD)

I carried out a Procrustes analysis (Wang *et al.*, 2010; Knowles *et al.*, 2016) per species to verify whether there is some consistent pattern between the geography and genetic variation using the package "vegan" (Oksanen *et al.*, 2015), function *protest*, and 10,000 permutations in R 3.4.4 (R-Development CoreTeam, 2018). Procrustes evaluated whether genomic data are in accordance with the expected gene flow under isolation by distance (IBD) model (Prado *et al.*, 2019), maximizing the similarities of two maps (set of coordinates), in other words, it minimizes the sum of the squares of Euclidean distances between them, rotating one of the maps, and the rotation amount is measured in degrees (Wang, Zöllner, & Rosenberg, 2012). One of the maps (coordinate set) represented the genetic diversity using the first two main components (PC1 and PC2) of the PCA analysis, and the other map corresponded to the geographical coordinates of the samples. The association between the two maps was calculated by the t_0 statistic that ranges from 0 (no association) to 1 (perfect association) (Knowles *et al.*, 2016). To verify the robustness of t_0, I excluded one population at a time and recalculated the PCA and Procrustes analyses. Thus, it was possible to calculate (i) if the association between geography and genetics was strongly dependent on a population and (ii) the maximum and minimum association value which could be recovered in each species (t''). I also compared the samples coordinates in the original PCA and in the with modified PCA (without one population by turn) and calculate the similarity between them (t') (Wang *et al.*, 2012; Knowles *et al.*, 2016; Prado *et al.*, 2019).

I also tested geographic distance effect on genomic variation through a Mantel test (Mantel, 1967) with the logarithmic geographic distance matrix (D matrix) and $F_{ST}/(1-F_{ST})$ values for pairwise populations (G matrix), using "ade4" (Dray & Dufour, 2007) and *mantel.rtest* function with 10,000 permutations. To determine the geographic coordinates of populations formed by several individuals, I calculated the centroid for the occurrence points for populations with three or more localities, using *centroid* function in "geosphere" R package (Hijmans, 2017), and the mean when there were only two different localities. I also eliminated one population at a time per species, and recalculated the Mantel test to verify if correlation and significance values were dependent on any specific population.

4.2.7. Isolation by barriers (IBB)

I grouped populations in three different barrier groups or strata (i) structural paleoarchs, (ii) ecoregions, and (iii) areas of endemism to test whether the same barrier presented similar effects on the genomic variation of sympatric species. Populations division into three different strata (Fig. S2; Table S1) relied on the studies on structural archs (Hoorn *et al.*, 1995; Lundberg *et al.*, 1998; Hubert & Renno, 2006; Espurt *et al.*, 2010; Mora *et al.*, 2010). I followed the proposal of Dinerstein et al. (2017) for ecoregions boundaries, and to delimit Napo and Inambari areas of endemism I followed Oliveira, Vasconcelos, & Santos (2017). I grouped the samples to the south of Madre de Dios River as "South Inambari" in the area of endemism stratum. Omernik (2004) considered the ecoregion division as areas where there are coincidences between characteristics of geological phenomena such as geology, physiography, vegetation, climate, soil, and others factors. In this way, it can be understood as an ecological or historical barrier for the studied species. Areas of endemism are delimited by the great Amazonian rivers, so when using them, clusters may be considered as interfluvial groups and the test would be about the role of rivers as barriers (Wallace, 1854; Nazareno, Dick, & Lohmann, 2017; Oliveira *et al.*, 2017). With these clusters, I performed analyses of molecular variance (AMOVA) (Excoffier, Smouse, & Quattro, 1992) with 10,000 permutations per stratum in each species, using *poppr.amova* and *randtest* from "poppr" (Kamvar, Tabima, & Grünwald, 2014) and "ade4" (Dray & Dufour, 2007) packages, respectively. All R scripts are available at: https://github.com/jdalapicolla.

To test the effect of past climatic oscillations, forming Pleistocene refuges (a historical barrier) I have created ecological niche models (ENM) with distributional and environmental variables for 4 time scales: present-day, Middle Holocene (about 6 Ka); LGM (22 Ka), and LIG (120-140 Ka). I calculated the overlapping areas among ENM for each species to identify if there were stable areas over time, and the suitable area in Km^2 for the species occurrence through time, and I identified the most important environmental variables for the species distribution. Thus, it was possible to test if the climatic oscillation of the Pleistocene affected similarly the sympatric species distributions, and whether the overlapping areas recovered small stable areas, such as refuges. For ENM, the environmental autocorrelation and the sample bias were reduced in the models, filtering the occurrence

points using a 10 Km buffer. A total of 138 unbiased occurrence points (*P. brevicauda*, n = 44; *P. simonsi*, n = 48; *P. steerei*, n = 46) were used to build the models (Table S3).

I used linear regressions and the variance inflation factor (VIF) to avoid collinearity (Zuur, Ieno, & Elphick, 2010; Dormann *et al.*, 2013; Zuur & Ieno, 2016) and select the environmental variables (Lin, Foster, & Ungar, 2011; Dupuis & Victoria-Feser, 2013). For each species, I performed a linear regression (*lm* function) with the 19 environmental variables and the unbiased occurrence points, using "car" package and *vif* function in 3.4.4 (R-Development CoreTeam, 2018). I calculated the VIF for all variables in the linear model. I eliminated the variable with the highest VIF, and redid the linear regression and calculated the VIF again with the new variable set. This approach of eliminating the variable with higher VIF and redoing the regression was repeated until all remaining variables had a VIF ≤ 2. In the end, the models used five variables to all target species (*P. brevicauda* and *P. steerei*= Bio2, Bio4, Bio8, Bio13, and Bio18; *P. simonsi* = Bio2, Bio4, Bio8, Bio16, and Bio18).

Models were created in the program MaxEnt (Phillips, Anderson, & Schapire, 2006), with 30 replicates, using bootstrap and 30% of occurrence points as test samples. Study area were selected using convex hulls on the occurrence points plus 100 Km^2 buffer. I eliminated areas of the Andes Mountains in the Western Amazon, where the genus *Proechimys* did not occur. Complexity and the feature class (FC) for modeling were chosen in "ENMeval" package (Muscarella *et al.*, 2014), with *ENMevaluate* function in R 3.4.4 (R-Development CoreTeam, 2018). I tested eight FC combinations: L, LQ, LQP, H, T, LQH, LQHP, and LQHPT (L = linear, Q = quadratic, H = hinge, T = threshold, and P = product); and regularization multiplier (RM) values from 0.5 to 3.0 with increments of 0.5, totaling 48 models. Since Velasco & González-Salazar (2019) showed that accuracy in geographical distribution predictions is not related with model complexity chosen by AIC in MaxEnt ENM, I used a lower calibrated Akaike information criterion value (*i.e.* ΔAICc) and other parameters to select the best models: higher Area Under the Curve (AUC) values, lower standard deviation in replicates, and lower difference between train and test AUC. Best model for *P. brevicauda* was FC = H and RM = 3.0; for *P. simonsi* was FC = LQH and RM = 2.0; and for *P. steerei* was FC = LQP, and RM = 3.0. I calculated the Area Under the Curve (AUC), True Skill Statistics (TSS), Omission Error to validate the models under a threshold of 10%. All R scripts are available at: https://github.com/jdalapicolla.

4.2.8. Isolation by environment (IBE)

I carried out a partial Mantel test (Smouse, Long, & Sokal, 1986) to evaluate how the environmental distance is correlated with the genetic distance, controlling the effect of geographic distances. D and G matrices were the same used in the Mantel test for IBD analyses, and the environmental matrix (E matrix) was a resistance matrix estimated from present-day ENM with raw values for suitability for each species. I inverted their suitability values to transform them into a friction layer matrix (1-ENM) (Chan, Brown, & Yoder, 2011). Posteriorly, I used the functions: *transition*, *geoCorrection* and *costDistance* for "gdistance" package (van Etten, 2018) to calculate the least-cost distance between populations, using the *mean* of 16 directions for cells connections to calculate the transition values of the grids. I used the *partial.mantel* function of the "vegan" (Oksanen *et al.*, 2015) to test the dissimilarity of the matrices with 1,000 permutations.

There are several critiques of using the partial Mantel test to evaluate the role of the E matrix in G (Raufaste & Rousset, 2001; Castellano & Balletto, 2002; Legendre & Fortin, 2010; Diniz-Filho *et al.*, 2013). Thus, I also did the partial distance-based Redundancy Analysis (dbRDA) for each species, to test the dissimilarity between the G and E, isolating the effect of D. I used the "vegan" package (Oksanen *et al.*, 2015) and the *capscale* function with *mahalanobis* distance for analyzes and 1,000 permutations. In dbRDA, the input information for D and E must be continuous data. Therefore, I transformed the D matrix into Principal Coordinates (PCNM) by function *pcnm* of the package "vegan" (Oksanen *et al.*, 2015) and to represent the environmental data I created a principal component analysis (PCA) with all 19 present climatic variables for the study area used for the ENM. I used the *rasterPCA* function with standardized data from the "RStoolbox" package (Leutner, Horning, & Schwalb-Willmann, 2018), and extracted the values from the first component (PC1) of the populations in each species.

4.2.9. Hypervolume overlapping and similarities

I estimated morphological and climatic hypervolumes for the three sympatric species of *Proechimys* to verify if they differed in the occupation of the *n*-dimensional hypervolume

(Hutchinson, 1957; Blonder *et al.*, 2014), which may indicate a niche overlap between the species (Blonder, 2016). For the morphological hypervolumes, I used the clean, logarithmized and standardized cranial measurements for the analyzes. I performed Principal Components Analyses (PCA) using "ade4" package (Dray & Dufour, 2007), and the *dudi.pca* function in the platform R 3.4.4 (R-Development CoreTeam, 2018). I used principal components instead of raw variables to reduce the analyses dimensionality.

I used the first nine principal components, representing 95% of variation, to quantify the morphological hypervolume via Gaussian kernel density estimation with "hypervolume" R package (Blonder & Harris, 2018) and the *hypervolume_gaussian* function. Kernel bandwidth was estimated with *estimate_bandwidth* function under Silverman method (Blonder & Harris, 2018; Blonder *et al.*, 2018), and I used "probability" quantile of 95%. Volumes values (*get_volume* function) and Sørensen similarity indexes (*hypervolume_set* and *hypervolume_overlap_statistics* functions) among morphological hypervolumes of the sympatric species was calculated. For similarities and overlapping analyses of hypervolumes among species, the sample size must be the same. As three sympatric species had different sample sizes, I used 36 observations of each species, and for *P. simonsi* and *P. steerei* that had more samples I randomly selected 36 individuals for hypervolumes estimation.

Climatic hypervolumes were estimated using the same package and functions above. I generated 1,000 random points within the present-day ENM to represent the fundamental niche for each of the target species. I extracted environmental information from these random points based on the 19 environmental variables from WorldClim (www.worldclim.com) (Hijmans *et al.*, 2005). With this dataset I performed a Principal Components Analyses (PCA), in the same way as in the analyses of morphological hypervolume, and used the five first principal components that represented 95% of the variation in the climatic hypervolume analyses.

4.3. Results

After I processed the raw reads, eliminated positions and loci with high polymorphism (Fig. S3), and removed the missing data, the number of samples and unlinked SNP were reduced in all target species (*P. brevicauda:* n = 16, SNP = 5,050; *P. simonsi:* n =

15, SNP = 4,629, and *P. steerei:* n = 14, SNP = 5,819) (Table S5). The coverage depth was greater than 20 reads/locus in all species (Table S6).

4.3.1. Genetic structure

The three sympatric species presented different patterns of genomic structure. In the PCA, *P. brevicauda* and *P. steerei* showed well-defined clusters in the multivariate space (Fig. S4), and the first two PC explained much of the variation in these two species (40.93% in *P. brevicauda* and 32.06% in *P. steerei;* Fig. S4). For *P. brevicauda* three clusters were recovered with two populations each: (*i*) IQUITOS + AMAZONAS in the north of range distribution; (*ii*) MADRE+ BOLIVIA in the south; (*iii*) and the central clusters ACRE+ GALVEZ (Fig. S4a). For *P. steerei* three clusters also were recovered in the south of species distribution (MADRE-BENI+PANDO), in the central area (ACRE), and in the east of species distribution (JAINU+JURUA+SOLIMOES). However, samples of *P. simonsi* were scattered throughout the multivariate space and first two PC had little genomic variation explained (19.11%; Fig. S4b), only GALVEZ population appeared more separated in multivariate space.

Global F-statistic and genetic diversity results also corroborated the different patterns in the three sympatric species (Table 1). *Proechimys brevicauda* has higher genomic structure (0.163 – 0.891) and diversity indexes ($\pi = 0.209$), followed by *P. steerei* (0.076 – 0.592; $\pi = 0.191$), with *P. simonsi* showing lower diversity ($\pi = 0.154$) and genomic structure (0.025 – 0.178). When pairwise populations were analyzed, this pattern was repeated (Table S7, S8, and S9): *P. brevicauda* showed more genomic structure between populations, *P. simonsi* the smaller, and *P. steerei* presented intermediate values. F_{ST} showed higher values than D for *P. brevicauda* ($F_{ST} = 0.465 – 0.019$; D = 0.261 – 0.009), while for *P. simonsi* ($F_{ST} = 0.064 – 0.002$; D = 0.049 – 0.008) and *P. steerei* ($F_{ST} = 0.139 – 0.006$; D = 0.117 – 0.006) the values were similar (Table S7). Hedrick's G''_{ST} presented higher values, usually twice as much as the Nei's G_{ST} for all sympatric species (*P. brevicauda:* $G''_{ST} = 0.839 – 0.091$ and $G_{ST} = 0.647 – 0.043$; *P. simonsi:* $G''_{ST} = 0.306 – 0.063$ and $G_{ST} = 0.156 – 0.033$; *P. steerei:* $G''_{ST} = 0.523 – 0.046$ and $G_{ST} = 0.298 – 0.021$;) (Table S8). Inbreeding coefficient index in the populations were similar among the three species with $F_{IS}<0.086$ in all populations (Table S9), while *P.*

brevicauda presented lower intrapopulation values of polymorphic sites, π, and H_{obs} when compared to the other two species (Table S9).

4.3.2. Isolation by distance (IBD)

Only *Proechimys steerei* could have its genomic variation explained by geographic distance in the Mantel test (Fig. 2; Table 2) with a moderate correlation (r = 0.512; p-value = 0.037). However, when I repeated the analysis by eliminating one population at time, the results of *P. steerei* (all tests became non-significant) and *P. simonsi* (three tests became significant) were inconstant (Table 2), only *P. brevicauda* remained with constant results regardless of the population removed.

Procrustes analyses considering the principal components showed significant associations between geography and genomic variation (Fig. 3), with *P. simonsi* presenting the lowest correlation (t_0 = 0.635; p-value = <0.001; Fig. 3b) when compared to *P. brevicauda* (t_0 = 0.730; p-value = <0.001; Fig. 3a) and *P. steerei* (t_0 = 0.718; p-value = <0.001; Fig. 3c). *Proechimys brevicauda* had less variation in the correlation between genomic and geography matrices (t'' = 0.677 – 0.824; Fig. 4; Table S10), regardless the excluded population in the permutation of procrustes analyses. *Proechimys simonsi* (t'' = 0.597 – 0.868; Fig. 4; Table S10) and *P. steerei* (t'' = 0.787 – 0.945; Fig. 4; Table S10) were more dependent of the some populations to reach the observed t_0 values, and the absence of some populations such as YUNGAS in *P. simonsi* and ACRE and JURUA in *P. steerei* may increase the association between genomic and geography matrices (t_0) by more than 0.2 (Fig. 4; Table S10). Larger deviations between geography and genetics (lines length in Fig. 3) were in populations at the extremes of the species range in all taxa, such as AMAZONAS and BOLIVIA populations in *P. brevicauda* (Fig. 3a); YUNGAS and SOLIMOES in *P. simonsi* (Fig. 3b), and SOLIMOES in *P. steerei* (Fig. 3c).

4.3.3. Isolation by barriers (IBB)

AMOVA results also indicated different patterns of genetic differentiation among the sympatric species, with differences between the barriers contribution to explain the genetic variation for each species (Table 3). The genomic structure pattern of *P. brevicauda* is consistent with the structural arches hypothesis (Φ_{ST} = 0.784; p-value = 0.036) and with the areas of endemism (Φ_{ST} = 0.773; p-value = <0.001), the later hypothesis may explain 69.9% of the species genomic variation, while the structural arches may explain 57.1%. *Proechimys steerei* had a similar pattern, with the structural arches and areas of endemism being significant, but in this case, the structural arches may explain more genomic variation (46.04%; Φ_{ST} = 0.470; p-value = <0.001) than areas of endemism (12.75%; Φ_{ST} = 0.482; p-value = <0.001). For *P. simonsi* only the areas of endemism had significant values, but the amount of variation explained by them and Φ_{ST} were small (12.92%; Φ_{ST} = 0.174; p-value = <0.001). Ecoregions presented no significant effect as a barrier to none species (Table 3).

In general, there was little variation in areas with high environmental suitability in the ENM, regardless the species or the time period (Fig. 5). All models presented reasonable or good validation indexes with omission error below 20% and AUC_{TEST} between 0.711 and 0.762, with low standard deviation (AUC_{SD} = 0.029 to 0.05) (Table S11). The most important variables for the models were those related to temperature (Table S12): Bio2 (mean diurnal range) for *P. brevicauda* (44.4% of contribution, and 55.8% of importance) and *P. simonsi* (64.8% of contribution, and 51.2% of importance), and for *P. steerei* were Bio8 (mean temperature of wettest quarter) (59.0% of contribution, and 49.7% of importance), and Bio2 (38.6% of contribution, and 45.4% of importance).

The largest variation in the ENM was in the Middle Holocene model for *P. brevicauda* that showed an evident fragmentation of suitable areas, in the other species and in the other time periods, the suitable areas remained practically the same with small variations in the edges of the distributions (Fig. 5). Even suitable area size for the three sympatric species were similar (Table S13). This stability was also reflected in the estimation of the stable areas over time in the region, that remained practically identical (Fig. S5a-c), which does not seem to be directly related to the rivers or structural arches of Western Amazon (Fig. S5d). I associated environmental stable areas for each species to the Procrustes results (dark gray areas in Fig. 3), and more than one genetic clusters (circles in Fig. 3) matched with a

single stable area, even when there was fragmentation as in the case of *P. brevicauda* (Fig. 3a) or when there was no fragmentation of these environmentally stable areas as in *P. simonsi* and *P. steerei* (Fig. 3b and Fig. 3c), showing little importance of the Pleistocene refugia hypothesis as a historical barrier to explain the genetic variation of these sympatric species.

4.3.4. Isolation by environment (IBE)

Partial Mantel tests for *P. brevicauda* and *P. simonsi* showed no effect of environmental distance on genetic variation (Table 2), while for *P. steerei*, there was significant effect only when all populations were included in the analyses (r = 0.663; p-value = 0.01; Table 2). dbRDA presented similar results, no effect for geographical or environmental distances for the species even when I isolated the effect of geographical distance through conditional tests (Table S14).

4.3.5. Hypervolume overlapping and similarities

The three sympatric species of *Proechimys* from Western Amazon showed greatest environmental hypervolumes similarities (Sørensen = 0.757 – 0.461; Table 4) when compared with the morphological hypervolumes (0.022 – 0.001; Table 4). Highest similarity values in environmental hypervolumes (0.757; Table 4) were between the two species that inhabit non-flooded forest, *terra-firme* forests (*P. brevicauda* and *P. simonsi*), and the lowest similarities values were same when comparing the species that occurs in the seasonal flooded lowland forest, *várzea* forests, *P. steerei*, with non-flooded forest species (0.521 – 0.461; Table 4). *Proechimys steerei* also presented the lowest environmental volume (751.46; Table 4) when compared to the other species of non-flooded forests (*P. brevicauda* = 1772.44; *P. simonsi* = 1929.36; Table 4). Environmental hypervolumes of the three sympatric species indicated a large overlap (Fig. S6a), and the centroids for these hypervolumes were more separated in PC1 that summed 44.52% of the variation (Table S15). Temperature Annual Range (Bio7) is the variable that contributed most to PC1 (Table S15), followed by Bio17 (Precipitation of Driest Quarter) and Bio12 (Annual Precipitation).

Morphological hypervolumes presented almost no similarities among the three species (Table 4) and considering the hypervolume sizes, *P. simonsi* showed the largest volume (1133.24; Table 4), followed by *P. steerei* (892.65) and *P. brevicauda* (585.12), the opposite of that I found in genetic structure. PC1, with 73.69% of the variation, and PC2 with 5.32% exhibited the greatest distances between the hypervolume centroids and the small overlaps between the morphological hypervolumes (Fig. S6b). The variables related to the cranial size (BaL, CIL, and GSL) contributed the most to PC1, and in PC2 were variables related to the rostral region (IFW and NL) and to the teeth (MTRL) (Table S16).

4.4. Discussion

4.4.1. Genetic structure patterns

The three sympatric species of *Proechimys* from the Western Amazon presented different levels of genetic structure (Table S7 and 5). F-statistics and PCA presented congruent results showing greater genetic similarity between the same populations in *P. brevicauda* and *P. steerei*. The clusters showed a geographical orientation, north-south in *P. brevicauda*, and south-east in *P. steerei*. *P. simonsi* did not show a clear clustering pattern. F_{ST} values among *Proechimys brevicauda* populations were higher than other rodents based on SNP data (Fischer *et al.*, 2014; Vega *et al.*, 2017), while *P. simonsi* and *P. steerei* values were comparable to them, and to other mammalian species as cetaceans (Li *et al.*, 2013; Lah *et al.*, 2016) and canids (Cronin *et al.*, 2015). These differences in F_{ST} values may be explained by the within-population diversity in *P. brevicauda,* which influences the index calculation (Meirmans, 2006). Numerous fixation indexes were suggested to replace F_{ST} in genetic structure comparisons to avoid this within-population diversity dependence (Hedrick, 2005; Jost, 2008). Jost's *D* was the genetic differentiation coefficient with the smallest range of values among population of the three species. However, some studies indicated the classical F_{ST} as a good tool to measure structure when using data from SNPs and other biallelic markers (Meirmans & Hedrick, 2011). Thus, I calculated the genetic distances matrix based on F_{ST} for the other analyses. Taking this into account, the high F_{ST} values may be an indication of cryptic diversity in *Proechimys*, especially in the case of *P. brevicauda* which showed a

deeper structure, similar to results I found in Chapter 2. All samples in each species are mutually monophyletic (Chapter 2), and the three sympatric species used here as model were considered valid in the last revisions for the genus, and with the same range distribution used in this study (Patton & Leite, 2015; Fabre *et al.*, 2016). Therefore, the possibility of potential different species within the current established species do not invalidated the results.

Emmons (1982), studying the populations of these three sympatric species in southern Peru, found differences in their population density with *P. simonsi* and *P. steerei* (named in that study as *P. hendeei* and *P. brevicauda*) being more abundant than *P. brevicauda* (named as *P. longicaudatus*). This difference in population density may be another factor associated with the genetic structure. Density-dependent dispersal is common in mammals, where the dispersal capacity of individuals is linked to population density (Matthysen, 2005). In the case of positive density-dependence, individuals from species with high populations density (such as *P. simoni* and *P. steerei*) are forced to disperse more to avoid competition (Waser, 1985; Porter & Dooley, 1993), which would lead to less genetic structure. Species with low density (such as *P. brevicauda*), would disperse less (Waser, 1985; Matthysen, 2005), increasing genetic differentiation. However, there is no data from other localities to verify if the low density population in *P. brevicauda* is a local or a repeated pattern across the landscape, or even studies if *Proechimys* species respond with differential dispersion to the different population densities.

4.4.2. Phylogeographic patterns

Although Mantel tests have presented mostly non-significant correlations and Procrustes analyses significant association between genetic and geography (Table 5), these isolation by distance analyses (IBD) did not show inconsistent results to describe genetic variation. However, these analyses have different goals, as Mantel test verifies if the genetic differences between populations varies linearly with the geographic distance (Diniz-Filho *et al.*, 2013), while the Procrustes shows the deviations from IBD model in the samples, considering genetic and geographic distances as coordinates, and rotating them to maximize their similarities, and presenting a value of association (t_0), which does not have the same meaning as Pearson's r (Wang *et al.*, 2010, 2012). There are some critiques to the power of

Mantel tests and its variations in population genetics, especially in permutation methods (Legendre & Fortin, 2010; Diniz-Filho *et al.*, 2013), while the Procrustes would be a more sensitive analysis to compare deviation patterns from IBD model among species and populations (Wang *et al.*, 2012; Knowles *et al.*, 2016). Significant t_0 values indicated that there was association between geography and genetic in *Proechimys* species, however those values were low when compared to some Alaskan alpine mammals (Knowles *et al.*, 2016) and similar to South American marsh rats (Prado *et al.*, 2019). For Alaskan species, this pattern was explained by the geographic location of the multiple Pleistocene refuges from which populations had their origin, in the case, most of the populations did not disperse very far from the refuges geographic location (Knowles *et al.*, 2016). For marsh rats this pattern is explained by the variation in the historical stability of the South American wetlands, and by their connectivity pattern among localities that would allow larger deviations of the IDB model and lower values of t_0. The ENM results showed historical stability in all three species, it may be the reason for the low t_0 values. *P. brevicauda* and *P. steerei* presented similar t_0 values but the deviations patterns between geography and genetics were different. *P. brevicauda* presented more latitudinal deviations, whereas *P. steerei* more longitudinal, the same pattern than *P. simonsi*. These differences may indicate that the origin of the deviations is not common for all species, and local processes may have affected their genetic structure differently. Moreover, deviations between geography and genetics were higher in *Prochimys* populations at the periphery of the distribution in the three species. It may suggest that the same process led to deviations, such as the presence of refuges and a recent colonization of the more distant areas (Leite & Rogers, 2013). However, different processes can lead to the same patterns and results of other analyses (see below), point out that this is a possibility (Knowles, 2009). Thus, both Mantel and Procrustes results showed that the three sympatric species deviated from the expected by IBD model.

Isolation by barriers models (IBB tests) showed that different barriers had different importance to explain genetic variation in *Proechimys* species (Table 5). Geographic barriers such as the main Amazonian rivers (equivalent to the boundaries of the areas of endemism), and the structural archs were more important than ecological barriers such as ecoregions and environmentally stable areas in the Pleistocene. Using the Procrustes results, the Marañón, Solimões, Madre de Dios, and Madeira rivers may be considered important for the genetic structure, and the last two rivers are geographically close to two structural arches, Fitzcarrald

and Jutaí, which together with Iquitos Arch presented greater power to explain the genetic structure in these *Proechimys* species. This is consistent with other studies in the region (Da Silva & Patton, 1998; Lougheed *et al.*, 1999; Gascon *et al.*, 2000; Patton *et al.*, 2000; Ribas *et al.*, 2012). However, these analyses did not allow me to affirm whether these barriers were the drivers of genetic differentiation. Considering the divergence times of these species (Late Pliocene and Pleistocene, Chapter 3) these barriers are likely areas of secondary contact (Oliveira et al., 2017; see Chapter 3 discussion). However, this model using the barriers as a factor that maintained populations isolated, rather than has been the primary driver for isolation, was the best model to explain the genetic structure in *P. brevicauda* by areas of endemism (= main rivers), and in *P. steerei* by structural archs. In *P. simonsi*, the barriers did not showed much contribution to explain the phylogenetic pattern. Studies with Neotropical birds have already identified that barriers and landscape changes were not necessarily the primary drivers for diversification; rather other species-traits, such as, the dispersal capacity in the landscape and the lineage time that species persists in the landscape better predictors of genetic structure (Smith et al. 2014).

Isolation by environment model (IBE) was not relevant to explain the genetic diversity pattern in these sympatric species (Table 5), either by the environmental distance of the current distribution of the populations (partial Mantel tests) or by the ecoregions (AMOVA), or climatic variables (dbRDA and hypervolume). The partial Mantel tests and climatic hypervolume indicated a slight separation of *P. steerei*, the seasonal flooded forest (*várzea*) species, in relation to the non-flooded forest (*terra-firme*) species (*P. brevicauda* and *P. simonsi*). These results may be explained by the power issue in the partial Mantel test that may increase the error of type I as the significance for *P. steerei* (Raufaste & Rousset, 2001; Legendre & Fortin, 2010). In addition, non-significant results for IBE model may be due to the poor resolution the climatic variables (Peterson & Nakazawa, 2007; Vale, Tarroso, & Brito, 2014). I was dealing with sympatric species, and variables with 1 Km² resolution could be inefficient to demonstrate the existence of environmental differences between the localities. On the other hand, with the same climatic variables, I was able to: (*i*) identify differences in volume and level of overlap in the climatic hypervolumes among species; and to (*ii*) recover differences in the ENM (especially in *P. brevicauda*). This suggest that these climatic variables would have enough resolution power to show differences in the IBE analyses, if the environmental distance had played a role on the diversification of these

species. In addition studies with other rodents from South America employed the same resolution for current and historical climatic variables, and they found significant correlations between genetic diversity pattern, environmental and species-traits (Prado *et al.*, 2019).

Variables related to temperature were the most important for the ENM. Their environmental stability through time could suggest a model overfitting, but the complexity was controlled (see Material and Methods), and the validation metrics did not show large differences between the AUC values in the training and the test steps (Table S10). Studies have shown little variation in precipitation (Cheng *et al.*, 2013), temperature (Colinvaux *et al.*, 1996) and in vegetation (Häggi *et al.*, 2017) during the Pleistocene in Western Amazon, and the most of intense changes in period occurred on the border of the Amazon biome and in the Eastern Amazon (Haffer, 1969; Colinvaux *et al.*, 2001; Bush & Oliveira, 2006), this seems to be a more plausible explanation for ENM stability. Pleistocene climatic oscillations were important factors for the inter- and intraspecific differentiation in *Proechimys* species but for taxa restricted to the Eastern Amazon and Guyana Shield (Leite, 2013; Silva et al., 2018; Chapter 3).

Although the environmental variables do not indicated segregation by microhabitat among populations and species, the results of the morphological hypervolumes with almost no overlap between the three species may indicate important differences in how the species use the habitat. Most important cranial variables for hypervolumes were related to the skull, rostrum, and teeth size. Thus, sympatric species may segregate at the microhabitat level in the Western Amazon, not by climatic differences or by the occupied habitat (seasonal floodplain or non-flooded forests) but rather they may use these habitats differently as resource partition, diet, foraging, mating system, and others. Emmons (1982) suggested that the high density populations of *P. simoni* and *P. steerei* in sympatry was due to the availability of food resources but she found no differences among the species diet. Unfortunately, there are no other published data on the diet or any other ecological aspect of these species to support or refute differences in this microhabitat related with allometry of those variables. Studies have shown that morphological data in hypervolume and functional ecology analyses do not indicate niche overlap but rather these analyses are related to the fitness (Shipley *et al.*, 2016; D'Andrea & Ostling, 2016; Blonder, 2018). *P. simonsi,* with larger hypervolume value for the morphology, also presented a shallow genetic structure without evidence of large inbreeding values (Table 1; Table S9), may indicate a higher gene flow among populations, and a large

morphological hypervolume could help in fitness and local adaptation in different areas of Western Amazon (Ridley, 2004).

Matocq, Patton, & da Silva (2000) also find unexpected results studying *P. simonsi* and *P. steerei.* Cyt b results showed similar number of haplotypes between species, no significant differences in the haplotype diversity, no isolation by distance patterns, stable population sizes for both species, and Rio Juruá was not considered a barrier, similar to my results. The results that are contradictory to mine are lower gene flow and more structured populations in *P. simonsi* in Matcoq et al., (2000) even in populations from the same river bank. This difference may be explained by the genetic markers and analyses used to infer population structure, and of course the study scale. The habitat used by the species, non-flooded (*terra-firme)* or seasonal floodplain (*várzea)* forests, did not predict the patterns of genetic structure, as indicated by Matocq, Patton, & da Silva (2000), because the two *terra-firme* species presented disparate genetic patterns, whereas the *várzea* species that should suffer bottleneck events due to seasonal floods, did not show evidence of recent population growth or inbreeding rates different from other species.

In short, phylogeographic patterns among the three sympatric species of *Proechimys* were different, with little relevance to the common geographic space (climatic variation and geographic distance) to explain the genetic diversity. Physical barriers, such as, rivers and structural arches seem to be more relevant to explain these patterns. Although the environmental distance, based on climatic variables, is not significant to explain the genetic diversity, local adaptation by other species-traits may be important due to differences in morpholological hypervolumes, indicating possible segregation in the habitat use.

4.4.3. Implications for conservation

Conservation biology has the challenge of dealing with various trade-offs to achieve its objectives, one of the most notable is how to choose to preserve certain areas over others, especially in biomes of colossal sizes as Amazon (Myers *et al.*, 2000; Avise, 2010). One option is to use diversity data and prioritize areas (Mittermeier *et al.*, 1998). Genetic diversity data are rarely used to select priority areas for conservation (Resende-Moreira *et al.*, 2019). More than preserving diversity patterns, conservation biology also aims to preserve the

processes that generate and/or maintain the diversity as the gene flow routes (Moritz, 2002), and that reduce the risk of inbreeding depression (Lande, 1988; Frankham, 2010). In this perspective, studies of comparative phylogeography, as the present one, are important to provide data for this decision-making in conservation (Da Silva & Patton 1998; Avise 2010; Resende-Moreira *et al.*, 2019).

Genetic data such as SNP can infer genetic structure patterns at different levels, which are useful for conservation biology (Morin, Martien, & Taylor, 2009; Angeloni *et al.*, 2012). In this study, for example, inbreeding rates among populations, which could indicate depressed levels of diversity (Frankham & Ralls, 1998; Hedrick & Kalinowski, 2000), were proportional and indicate a good population size for all species. This result is reasonable because Western Amazon suffers less with deforestation and dams than the Eastern Amazon (Barona *et al.*, 2010; Fearnside, 2015), and *Proechimys* is a genus known to inhabit both pristine and altered forests (Patton & Leite, 2015).

Although Western Amazon has suffered less with habitat loss lately, when compared to the Cerrado or Eastern Amazon, it already undergoes environmental degradation, especially by heavy metals contamination of rivers due to mining and dams (Malhi *et al.*, 2008) and the pressure for Oil and mining exploration (Finer *et al.*, 2008). In addition, the agricultural and cattle raising frontier is already close to the southern Western Amazon, being Rondônia one of the Brazilian states with the highest rates of deforestation (de Barros Ferraz *et al.*, 2005; Rosa, Souza, & Ewers, 2012; Piontekowski *et al.*, 2014). Therefore, studies aimed at understanding the diversity patterns of the region, based on genetic or not, are crucial and strategic.

In this study, I evaluated the case of three sympatric species belonging to the same genus that, contrary to what was expected, presented distinct patterns of genetic variation, indicating different degrees of association between geographical and environmental barriers and genetic diversity. In addition, the species that inhabit the seasonally flooded forests presented greater genetic diversity than one of the species from non-flooded forest. These environments are more vulnerable to degradation because they are more easily accessible, more used for agriculture and cattle raising and are more affected by heavy-metal contamination than non-flooded forest (Castello *et al.*, 2013). It may indicate that important portions of genetic diversity are more threatened than other by environmental degradation.

These results showed that the evolutionary and geological history of a region is not simple, and that proposing conservation strategies based on few data may underestimate the importance of certain areas and processes. Thus, encouraging studies on genetic diversity and processes that generate and maintain this diversity are also crucial and priority for conservation biology (Allendorf, Hohenlohe, & Luikart, 2010), and to protect genetic and ecosystem resources for future generations.

4.5. Conclusions

Genetic structure patterns of the three sympatric species of *Proechimys* from the Western Amazon was different, with different models of diversification in the region showing different levels of importance to explain genetic diversity in each species.

Proechimys brevicauda, a species inhabiting non-flooded forests (*terra-firme* forests), presented the highest genetic structure values. In addition, geographic and environmental distances were not significant to explain its genetic variation, but the Procrustes analysis identified a moderate association between geography and genetic variation with AMOVA tests recovering that areas of endemism and consequently the main Amazonian rivers as the best hypothesis to explain the genetic structure in this species. *Proechimys simonsi*, another species that also occurs in non-flooded forests (*terra-firme* forests), presented shallower genetic structure among the three sympatric species; geographic and environmental distance were also not significant; and Procrustes recovered the lowest association value. Only the areas of endemism as a barrier was significant but it explained a little portion of the genetic variation. *Proechimys steerei*, a species that inhabits the seasonal flooded forests (*várzea* forest), presented intermediate values of genetic structure. Similarly to *P. brevicauda,* it showed moderate value of association between geography and genetic data in Procrustes and the structural archs are better to explain the genetic variation. The geographic distance was significant but when the physical distance effect was isolated, the environmental distance could explain the genetic variation in the partial Mantel test but not in the bdRDA. Also, the climatic oscillations of the Pleistocene did not seem to be important to explain phylogeographic patterns in the three species.

In addition, morphological hypervolumes indicated almost no overlap among the three species, regardless of the habitat occupied by the species. These results may indicate that local adaptation, microhabitat differences, phylogenetic history, and species-traits may be more important than the common geographic space to explain genetic variation in sympatric species. Moreover, I presented information on population genetics, such as absence of inbreeding depression and loss of genetic diversity, which has implications for the conservation biology of *Proechimys* species and also for the conservation of the diversity of the entire Amazon biome.

References

ALLENDORF, F. W., P. A. HOHENLOHE, AND G. LUIKART. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics 11:697–709.

ANGELONI, F., N. WAGEMAKER, P. VERGEER, AND J. OUBORG. 2012. Genomic toolboxes for conservation biologists. Evolutionary Applications 5:130–143.

AVISE, J. C. 2009. Phylogeography: retrospect and prospect. Journal of Biogeography 36:3–15.

AVISE, J. C. 2010. Perspective: conservation genetics enters the genomics era. Conservation Genetics 11:665–669.

BARONA, E., N. RAMANKUTTY, G. HYMAN, AND O. T. COOMES. 2010. The role of pasture and soybean in deforestation of the Brazilian Amazon. Environmental Research Letters 5:024002.

DE BARROS FERRAZ, S. F., C. A. VETTORAZZI, D. M. THEOBALD, AND M. V. R. BALLESTER. 2005. Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondônia, Brazil: assessment and future scenarios. Forest Ecology and Management 204:69–85.

BLONDER, B. 2016. Do hypervolumes have holes? The American Naturalist 187:E93-105.

BLONDER, B. ET AL. 2018. New approaches for delineating *n* -dimensional hypervolumes. Methods in Ecology and Evolution 9:305–319.

BLONDER, B. 2018. Hypervolume concepts in niche- and trait-based ecology. Ecography 41:1441–1455.

BLONDER, B., AND D. J. HARRIS. 2018. Package ‘ hypervolume ’: high dimensional geometry and set operations using Kernel density estimation, support vector machines, and convex hulls.

BLONDER, B., C. LAMANNA, C. VIOLLE, AND B. J. ENQUIST. 2014. The *n* -dimensional hypervolume. Global Ecology and Biogeography 23:595–609.

BONVICINO, C. R., AND M. WEKSLER. 2012. Speciation in Amazonia: patterns and predictions of a network of hypotheses. Pp. 259–282 in Bones, clones, and biomes: the history and geography of recent Neotropical mammals (B. D. Patterson & L. P. Costa, eds.). The University of Chicago Press, Chicago.

BUSH, M. B., AND P. E. DE OLIVEIRA. 2006. The rise and fall of the Refugial Hypothesis of Amazonian speciation: a paleoecological perspective. Biota Neotropica 6:bn0010601.

CARNAVAL, A. C., M. J. HICKERSON, C. F. HADDAD, M. T. RODRIGUES, AND C. MORITZ. 2009. Stability predicts genetic diversity in the brazilian atlantic forest hotspot. Science 323:785–789.

CARNAVAL, A. C., AND C. MORITZ. 2008. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. Journal of Biogeography 35:1187–1201.

CASTELLANO, S., AND E. BALLETTO. 2002. Is the partial Mantel test inadequate? Evolution 56:1871–1873.

CASTELLO, L. ET AL. 2013. The vulnerability of Amazon freshwater ecosystems. Conservation Letters 6:217–229.

CATCHEN, J., P. A. HOHENLOHE, S. BASSHAM, A. AMORES, AND W. A. CRESKO. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3124–3140.

CHAN, L. M., J. L. BROWN, AND A. D. YODER. 2011. Integrating statistical genetic and geospatial methods brings new power to phylogeography. Molecular Phylogenetics and Evolution 59:523–537.

CHENG, H. ET AL. 2013. Climate change patterns in Amazonia and biodiversity. Nature Communications 4:1411.

COLINVAUX, P. A., G. IRION, M. E. RÄSÄNEN, M. B. BUSH, AND J. N. DE MELLO. 2001. A paradigm to be discarded: geological and paleoecological data falsify the Haffer & Prance refuge hypothesis of Amazonian speciation. Amazoniana 16:609–646.

COLINVAUX, P. A., P. E. DE OLIVEIRA, J. E. MORENO, M. C. MILLER, AND M. B. BUSH. 1996. A long pollen record from lowland Amazonia: forest and cooling in glacial times. Science 274:85–88.

COSTA, L. P., AND Y. L. R. LEITE. 2012. Historical fragmentation shaping vertebrate diversification in the Atlantic Forest biodiversity hotspot. Pp. 283–306 in Bones, clones, and biomes: the history and geography of recent Neotropical mammals (B. D. Patterson & L. P. Costa, eds.). The University of Chicago Press, Chicago, IL.

COSTA, M. H., AND J. A. FOLEY. 1998. A comparison of precipitation datasets for the Amazon Basin. Geophysical Research Letters 25:155–158.

CRONIN, M. A., A. CÁNOVAS, D. L. BANNASCH, A. M. OBERBAUER, AND J. F. MEDRANO. 2015. Single nucleotide polymorphism (SNP) variation of wolves (*Canis lupus*) in southeast Alaska and comparison with wolves, dogs, and coyotes in North America. Journal of Heredity 106:26–36.

D'ANDREA, R., AND A. OSTLING. 2016. Challenges in linking trait patterns to niche differentiation. Oikos 125:1369–1385.

DINERSTEIN, E. ET AL. 2017. An ecoregion-based approach to protecting half the terrestrial realm. BioScience 67:534–545.

DINIZ-FILHO, J. A. F. ET AL. 2013. Mantel test in population genetics. Genetics and Molecular Biology 36:475–485.

DORMANN, C. F. ET AL. 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36:27–46.

DRAY, S., AND A. B. DUFOUR. 2007. The ade4 package: implementing the duality diagram for ecologists. Journal of Statistical Software 22:1–20.

DUPUIS, D. J., AND M.-P. VICTORIA-FESER. 2013. Robust VIF regression with application to variable selection in large data sets. The Annals of Applied Statistics 7:319–341.

EMMONS, L. H. 1982. Ecology of *Proechimys* (Rodentia, Echimyidae) in southeastern Peru. Tropical Ecology 23:280–290.

ESPURT, N., P. BABY, S. BRUSSET, M. RODDAZ, W. HERMOZA, AND J. BARBARAND. 2010. The Nazca Ridge and uplift of the Fitzcarrald Arch: implications for regional geology in northern South America. Pp. 89–100 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

VAN ETTEN, J. 2018. gdistance: distances and routes on geographical grids.

EXCOFFIER, L., P. E. SMOUSE, AND J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491.

FABRE, P. H., J. L. PATTON, AND Y. L. R. LEITE. 2016. Family Echimyidae (hutias, South American spiny-rats and coypu). Pp. 552–641 in Handbook of the Mammals of the World. Vol 6. Lagomorphs and Rodents I (D. E. Wilson, T. E. J. Lacher & R. A. Mittermeier, eds.). Lynx Edicions, Barcelona.

FEARNSIDE, P. M. 2015. Amazon dams and waterways: Brazil's Tapajós Basin plans. Ambio 44:426–439.

FINER, M., C. N. JENKINS, S. L. PIMM, B. KEANE, AND C. ROSS. 2008. Oil and gas projects in the Western Amazon: threats to wilderness, biodiversity, and indigenous peoples. PLoS ONE 3:e2932.

FISCHER, M. C., M. FOLL, G. HECKEL, AND L. EXCOFFIER. 2014. Continental-scale footprint of balancing and positive selection in a small rodent (*Microtus arvalis*). PLoS ONE 9:e112332.

FRANKHAM, R. 2010. Challenges and opportunities of genetic approaches to biological conservation. Biological Conservation 143:1919–1927.

FRANKHAM, R., AND K. RALLS. 1998. Inbreeding leads to extinction. Nature 392:441–442.

GASCON, C. ET AL. 2000. Riverine barriers and the geographic distribution of Amazonian species. Proc Natl Acad Sci U S A 97:13672–13677.

HAFFER, J. 1969. Speciation in Amazonian forest birds. Science 165:131–137.

HAFFER, J. 1997. Alternative models of vertebrate speciation in amazonia an overview. Biodiversity and Conservation 6:451–476.

HÄGGI, C. ET AL. 2017. Response of the Amazon rainforest to late Pleistocene climate variability. Earth and Planetary Science Letters 479:50–59.

HARTL, D. L., AND A. G. CLARK. 1997. Principles of population genetics. 4th edition. Sinauer Associates, Inc. Publishers, Sunderland.

HEDRICK, P. W. 2005. A standardized genetic differentiation measure. Evolution 59:1633–1638.

HEDRICK, P. W., AND S. T. KALINOWSKI. 2000. Inbreeding depression in conservation biology. Annual Review of Ecology and Systematics 31:139–162.

HICKERSON, M. J. ET AL. 2010. Phylogeography's past, present, and future: 10 years after Avise, 2000. Mol Phylogenet Evol 54:291–301.

HIJMANS, R. J. 2017. geosphere: spherical trigonometry.

HIJMANS, R. J., S. E. CAMERON, J. L. PARRA, P. G. JONES, AND A. JARVIS. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965–1978.

HOORN, C. ET AL. 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. Science 330:927–931.

HOORN, C., J. GUERRERO, G. A. SARMIENTO, AND M. A. LORENTE. 1995. Andean tectonics as a cause for changing drainage patterns in Miocene northern South America. Geology 23:237.

HUBERT, N., AND J.-F. RENNO. 2006. Historical biogeography of South American freshwater fishes. Journal of Biogeography 33:1414–1436.

HUTCHINSON, G. E. 1957. Concludig remarks. Cold Spring Harbor Symposia on Quantitative Biology 22:415–427.

JOMBART, T., AND I. AHMED. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071.

JOST, L. 2008. G ST and its relatives do not measure differentiation. Molecular Ecology 17:4015–4026.

KAMVAR, Z. N., J. F. TABIMA, AND N. J. GRÜNWALD. 2014. *Poppr*: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.

KNOWLES, L. L. 2009. Statistical phylogeography. Annual Review of Ecology, Evolution, and Systematics 40:593–612.

KNOWLES, L. L., R. MASSATTI, Q. HE, L. E. OLSON, AND H. C. LANIER. 2016. Quantifying the similarity between genes and geography across Alaska's alpine small mammals. Journal of Biogeography 43:1464–1476.

LAH, L. ET AL. 2016. Spatially explicit analysis of genome-wide SNPs detects subtle population structure in a mobile marine mammal, the Harbor Porpoise. PLOS ONE 11:e0162792.

LANDE, R. 1988. Genetics and demography in biological conservation. Science 231:129–133.

LEE, C. R., AND T. MITCHELL-OLDS. 2011. Quantifying effects of environmental and geographical factors on patterns of genetic differentiation. Molecular Ecology 20:4631–4642.

LEGENDRE, P., AND M. J. FORTIN. 2010. Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources 10:831–844.

LEITE, R. N. 2013. Biogeography and evolution of Neotropical small mammals, with emphasis on hystricognath spiny rats of the genus *Proechimys* (Family Echimyidae). Brigham Young University.

LEITE, R. N., AND D. S. ROGERS. 2013. Revisiting Amazonian phylogeography: insights into diversification hypotheses and novel perspectives. Organisms Diversity & Evolution 13:639–664.

LEUTNER, B., N. HORNING, AND J. SCHWALB-WILLMANN. 2018. RStoolbox: tools for remote sensing data analysis.

LI, S., S. XU, H. WAN, H. JI, K. ZHOU, AND G. YANG. 2013. Genome-wide SNP and population divergence of finless porpoises. Genome Biology and Evolution 5:758–768.

LIN, D., D. P. FOSTER, AND L. H. UNGAR. 2011. VIF regression: a fast regression algorithm for large data. Journal of the American Statistical Association 106:232–247.

LOUGHEED, S. C., C. GASCON, D. A. JONES, J. P. BOGART, AND P. T. BOAG. 1999. Ridges and rivers: a test of competing hypotheses of Amazonian diversification using a dart-poison frog (*Epipedobates femoralis*). Proceedings of the Royal Society of London. Series B: Biological Sciences 266:1829–1835.

LUNDBERG, J. G., L. G. MARSHALL, J. GUERRERO, B. K. HORTON, M. C. S. L. MALABARBA, AND F. WESSELINGH. 1998. The stage for neotropical fish diversification: A history of tropical South American rivers. Pp. 13–48 in Phylogeny and Classification of Neotropical Fishes (L. R. Malabarba, R. E. Reis, R. P. Vari, Z. M. Lucena & C. A. S. Lucena, eds.). Edipucrs, Porto Alegre.

MALCOLM, J. R. 1992. Use of tooth impressions to identify and age live *Proechimys guyannensis* and *P. cuvieri* (Rodentia: Echimyidae). Journal of Zoology 227:537–546.

MALHI, Y., J. T. ROBERTS, R. A. BETTS, T. J. KILLEEN, W. LI, AND C. A. NOBRE. 2008. Climate Change, Deforestation, and the Fate of the Amazon. Science 319.

MANTEL, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer research 27:209–220.

MATOCQ, M. D., J. L. PATTON, AND M. N. F. DA SILVA. 2000. Population genetic structure of two ecologically distinct amazonian spiny rats: separating history and current ecology. Evolution 54:1423–1432.

MATTHYSEN, E. 2005. Density-dependent dispersal in birds and mammals. Ecography 28:403–416.

MEIRMANS, P. G. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. Evolution 60:2399–2402.

MEIRMANS, P. G., AND P. W. HEDRICK. 2011. Assessing population structure: F_{ST} and related measures. Molecular Ecology Resources 11:5–18.

MITTERMEIER, R. A., N. MYERS, J. B. THOMSEN, G. A. B. DA FONSECA, AND S. OLIVIERI. 1998. Biodiversity hotspots and major tropical wilderness areas: approaches to setting conservation priorities. Conservation Biology 12:516–520.

MORA, A. ET AL. 2010. Tectonic history of the Andes and sub-Andean zones: Implications for the development of the Amazon drainage basin. Pp. 38–60 in Amazonia, Landscape and Species Evolution: A Look into the Past (C. Hoorn & F. P. Wesselingh, eds.). Blackwell Publishing, Hoboken.

MORIN, P. A., K. K. MARTIEN, AND B. L. TAYLOR. 2009. Assessing statistical power of SNPs for population structure and conservation studies. Molecular Ecology Resources 9:66–73.

MORITZ, C. 2002. Strategies to protect biological diversity and the evolutionary processes that sustain it. Systematic biology 51:238–54.

MORITZ, C., AND D. P. FAITH. 1998. Comparative phylogeography and the identification of genetically divergent areas for conservation. Molecular Ecology 7:419–429.

MUSCARELLA, R. ET AL. 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in Ecology and Evolution 5:1198–1205.

MYERS, N., R. A. MITTERMEIER, C. G. MITTERMEIER, G. A. B. DA FONSECA, AND J. KENT. 2000. Biodiversity hotspots for conservation priorities. Nature 403:853–858.

NAZARENO, A. G., C. W. DICK, AND L. G. LOHMANN. 2017. Wide but not impermeable: Testing the riverine barrier hypothesis for an Amazonian plant species. Molecular Ecology 26:3636–3648.

NEI, M., AND R. K. CHESSER. 1983. Estimation of fixation indices and gene diversities. Annals of Human Genetics 47:253–259.

OKSANEN, J. ET AL. 2015. vegan: Community Ecology Package. *R package version 2.2-1*. .

OLIVEIRA, U., M. F. VASCONCELOS, AND A. J. SANTOS. 2017. Biogeography of Amazon birds: rivers limit species composition, but not areas of endemism. Scientific Reports 7:2992.

OMERNIK, J. M. 2004. Perspectives on the nature and definition of ecological regions. Environmental Management 34:S27–S38.

OTTO-BLIESNER, B. L., S. J. MARSHALL, J. T. OVERPECK, G. H. MILLER, AND A. HU. 2006. Simulating arctic climate warmth and icefield retreat in the last interglaciation. Science 311:1751–1753.

PARIS, J. R., J. R. STEVENS, AND J. M. CATCHEN. 2017. Lost in parameter space: a road map for stacks. Methods in Ecology and Evolution 8:1360–1373.

PATTON, J. L., AND M. A. ROGERS. 1983. Systematic implications of non-geographic variation in the spiny rat genus *Proechimys* (Echimyidae). Zeitschrift für Säugetierkunde 48:363–370.

PATTON, J. L., M. N. DA SILVA, AND J. R. MALCOLM. 2000. Mammals of the rio Juruá and the evolutionary and ecological diversification of Amazonia. Bulletin American Museum of Natural History 244:1–306.

PATTON, J. P., AND R. N. LEITE. 2015. Genus *Proechimys* J. A. Allen, 1889. Pp. 950–989 in Mammals of South America, Volume 2: Rodents (J. L. Patton, U. F. J. Pardiñas & G. D'Elía, eds.). University Of Chicago Press, Chicago.

PAYNTER, R. A. 1982. Ornithological gazetteer of Venezuela. Bird Department, Museum of Comparative Zoology, Harvard University, Cambridge.

PAYNTER, R. A. 1992. Ornithological gazetteer of Bolivia. 2nd ed. Bird Department, Museum of Comparative Zoology, Harvard University, Cambridge.

PAYNTER, R. A. 1993. Ornithological gazetteer of Ecuador. 2nd ed. Bird Department, Museum of Comparative Zoology, Harvard University, Cambridge.

PAYNTER, R. A. 1997. Ornithological gazetteer of Colombia. Bird Department, Museum of Comparative Zoology, Harvard University, Cambridge.

PAYNTER, R. A., AND M. A. TRAYLOR. 1991a. Ornithological gazetteer of Brazil A-M. Bird Department, Museum of Comparative Zoology, Harvard University, Cambridge.

PAYNTER, R. A., AND M. A. TRAYLOR. 1991b. Ornithological gazetteer of Brazil N-Z. Bird Department, Museum of Comparative Zoology, Harvard University, Cambridge.

PETERSON, A. T., AND Y. NAKAZAWA. 2007. Environmental data sets matter in ecological niche modelling: an example with *Solenopsis invicta* and *Solenopsis richteri*. Global Ecology and Biogeography 0:071113201427001-???

PETERSON, B. K., J. N. WEBER, E. H. KAY, H. S. FISHER, AND H. E. HOEKSTRA. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

PHILLIPS, S. J., R. P. ANDERSON, AND R. E. SCHAPIRE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231–259.

PIONTEKOWSKI, V. J., E. A. T. MATRICARDI, M. A. PEDLOWSKI, AND L. C. FERNANDES. 2014. Avaliação do desmatamento no estado de Rondônia entre 2001 e 2011. Floresta e Ambiente 21:297–306.

PORTER, J. H., AND J. L. DOOLEY. 1993. Animal dispersal patterns: a reassessment of simple mathematical models. Ecology 74:2436–2443.

PRADO, J. R., A. R. PERCEQUILLO, A. T. THOMAZ, AND L. L. KNOWLES. 2019. Similar but different: Revealing the relative roles of species-traits versus biome properties structuring genetic variation in South American marsh rats. Journal of Biogeography 46:770–783.

R-DEVELOPMENT CORETEAM. 2018. R Development Core Team. Viena. *R: A Language and Environment for Statistical Computing*. .

RAUFASTE, N., AND F. ROUSSET. 2001. Are partial Mantel tests adequate? Evolution 55:1703–1705.

RESENDE-MOREIRA, L. C. ET AL. 2019. Evolving in isolation: Genetic tests reject recent connections of Amazonian savannas with the central Cerrado. Journal of Biogeography 46:196–211.

RIBAS, C. C., A. ALEIXO, A. C. R. NOGUEIRA, C. Y. MIYAKI, AND J. CRACRAFT. 2012. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. Proceedings of the Royal Society of London B: Biological Sciences 279.

RIDLEY, M. 2004. The theory of Natural Selection. Pp. 95–136 in Evolution. 3rd edition. Malden.

ROSA, I. M. D., C. SOUZA, AND R. M. EWERS. 2012. Changes in size of deforested patches in the Brazilian Amazon. Conservation Biology 26:932–937.

SCHETINO, M. A. A. 2008. Código de barras de DNA em espécies de *Proechimys* (Rodentia: Echimyidae) da Amazônia. Instituto Nacional de Pesquisas da Amazônia.

SHIPLEY, B., F. DE BELLO, J. H. C. CORNELISSEN, E. LALIBERTÉ, D. C. LAUGHLIN, AND P. B. REICH. 2016. Reinforcing loose foundation stones in trait-based plant ecology. Oecologia 180:923–931.

SILVA, C. R. ET AL. 2018. The role of Pleistocene climate change in the genetic variability, distribution and demography of *Proechimys cuvieri and P. guyannensis* (Rodentia: Echimyidae) in northeastern Amazonia. PLoS One 13:e0206660.

DA SILVA, M. N., AND J. L. PATTON. 1998. Molecular phylogeography and the evolution and conservation of amazonian mammals. Molecular Ecology 7:475–486.

SMITH, B. T. ET AL. 2014. The drivers of tropical speciation. Nature 515:406–409.

SMOUSE, P. E., J. C. LONG, AND R. R. SOKAL. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. Systematic Zoology 35:627–632.

STEINER, C., P. SOURROUILLE, AND F. CATZEFLIS. 2000. Molecular characterization and mitochondrial sequence variation in two sympatric species of *Proechimys* (Rodentia: Echimyidae) in French Guiana. Biochemical Systematics and Ecology 28:963–973.

STEPHENS, L., AND M. A. TRAYLOR. 1983. Ornithological gazetteer of Peru. Bird Department, Museum of Comparative Zoology, Harvard University, Cambridge.

STEPHENS, L., AND M. A. TRAYLOR. 1985. Ornithological gazetteer of the Guianas. Bird Department, Museum of Comparative Zoology, Harvard University, Cambridge.

THOME, M. T., K. R. ZAMUDIO, J. G. GIOVANELLI, C. F. HADDAD, F. A. BALDISSERA JR., AND J. ALEXANDRINO. 2010. Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic Forest. Molecular Phylogenetics and Evolution 55:1018–1031.

VALE, C. G., P. TARROSO, AND J. C. BRITO. 2014. Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara-Sahel transition zone. Diversity and Distributions 20:20–33.

VEGA, R., E. VÁZQUEZ-DOMÍNGUEZ, T. A. WHITE, D. VALENZUELA-GALVÁN, AND J. B. SEARLE. 2017. Population genomics applications for conservation: the case of the tropical dry forest dweller *Peromyscus melanophrys*. Conservation Genetics 18:313–326.

VELASCO, J. A., AND C. GONZÁLEZ-SALAZAR. 2019. Akaike information criterion should not be a "test" of geographical prediction accuracy in ecological niche modelling. Ecological Informatics 51:25–32.

VONHOF, H. B., F. P. WESSELINGH, AND G. M. GANSSEN. 1998. Reconstruction of the Miocene western Amazonian aquatic system using molluscan isotopic signatures. Palaeogeography, Palaeoclimatology, Palaeoecology 141:85–93.

VOSS, R. S., D. P. LUNDE, AND N. B. SIMMONS. 2001. The mammals of Paracou, French Guiana: a neotropical lowland rainforest fauna part 2. Nonvolant species. Bulletin of the American Museum of Natural History 263:3–236.

WALLACE, A. R. 1854. On the monkeys of the Amazon. Annals and Magazine of Natural History 14:451–454.

WANG, C. ET AL. 2010. Comparing spatial maps of human population-genetic variation using Procrustes analysis. Statistical Applications in Genetics and Molecular Biology 9:1–22.

WANG, C., S. ZÖLLNER, AND N. A. ROSENBERG. 2012. A quantitative comparison of the similarity between genes and geography in worldwide human populations. PLoS Genetics 8:e1002886.

WANG, I. J., AND G. S. BRADBURD. 2014. Isolation by environment. Molecular Ecology 23:5649–5662.

WASER, P. M. 1985. Does competition drive dispersal? Ecology 66:1170–1175.

WINTER, D. J. 2012. mmod : an R library for the calculation of population differentiation statistics. Molecular Ecology Resources 12:1158–1160.

WOODS, AND KILPATRICK. 2005. Infraorder Hystricognathi Brandt, 1855. Pp. 1538–1600 in Mammal Species of the World. A Taxonomic and Geographic Reference (D. E. Wilson & D. M. Reeder, eds.). Volume 2. Johns Hopkins University Press, Baltimore.

WRIGHT, S. 1943. Isolation by Distance. Genetics 28:114–38.

ZUUR, A. F., AND E. N. IENO. 2016. A protocol for conducting and presenting results of regression-type analyses. Methods in Ecology and Evolution 7:636–645.

ZUUR, A. F., E. N. IENO, AND C. S. ELPHICK. 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1:3–14.

Tables

Table S1: Global fixation indexes and nucleotide diversity (π) for the three sympatric species of *Proechimys* from Western Amazon. H_S = Heterozygosity expected taking account populations division. H_T = Heterozygosity expected without population division. G_{ST} = standard Nei G_{ST}. G''_{ST} = Hedrick's G''_{ST}. D = Jost's D. Φ_{ST} = Meirmans Φ_{ST}.

	P. brevicauda	***P. simonsi***	***P. steerei***
H_S	0.083	0.137	0.139
H_T	0.208	0.155	0.193
G_{ST}	0.600	0.118	0.281
G''_{ST}	0.702	0.156	0.371
D	0.163	0.025	0.076
Φ_{ST}	0.891	0.178	0.592
π	0.209	0.154	0.191

Table 2: Mantel and partial Mantel tests with all populations for each one of the three sympatric species of *Proechimys* of the Western Amazon (ALL and in italics). Tests eliminating one population at a time are also indicated with the population name eliminated in the "Populations" column. Significant values are in bold, r (Pearson's correlation coefficient) and p-value are also reported.

Species	Populations	Mantel Test		Partial Mantel Test	
		r	p-value	r	p-value
P. brevicauda	GALVEZ	0.653	0.086	-0.113	0.600
	MADRE	0.714	0.183	0.754	0.142
	BOLIVIA	0.540	0.185	0.688	0.142
	AMAZONAS	0.701	0.067	0.263	0.250
	ACRE	0.628	0.096	0.314	0.225
	IQUITOS	0.665	0.083	-0.133	0.600
	ALL	*0.622*	*0.063*	*0.468*	*0.074*
P. simonsi	SOLIMOES	0.293	0.088	0.105	0.322
	CENTRAL ANDES	**0.686**	**0.005**	-0.697	0.989
	MADEIRA	**0.388**	**0.038**	0.192	0.242
	GALVEZ	0.315	0.143	-0.032	0.431
	YUNGAS	**0.426**	**0.032**	0.440	0.054
	MADRE	0.298	0.195	0.153	0.365
	JURUA	0.314	0.158	0.037	0.500
	ALL	*0.337*	*0.095*	*-0.010*	*0.515*
P. steerei	MADRE-BENI	0.372	0.224	0.479	0.100
	PANDO	0.382	0.211	0.621	0.075
	SOLIMOES	0.706	0.111	0.014	0.450
	ACRE	0.676	0.081	0.423	0.092
	JURUA	0.588	0.076	0.347	0.158
	JAINU	0.427	0.152	0.309	0.183
	ALL	***0.512***	***0.037***	***0.663***	***0.010***

Table 3: Analysis of molecular variance (AMOVA) for the three sympatric species of *Proechimys* from Western Amazon, considering the strata: strutural archs, areas of endemism (AOE), and ecoregions. Degrees of freedom (d.f.), percentage of genomic variation, Φ indices, and p-values for the genomic heterogeneity are presented.

Species	Strata	Source of variation	d.f.	Sum of squares	Variation (%)	Φ	p-value
P. brevicauda	**Structural Archs**	Among Archs	2	8,057.27	57.13	$\Phi_{ST} = 0.784$	0.036
		Among Samples Within Archs	3	2,937.21	21.25	$\Phi_{SC} = 0.496$	<0.001
		Within Samples	10	2,610.08	21.63	$\Phi_{CT} = 0.571$	<0.001
		Total	15	13,604.56			
	Areas of Endemism (AOE)	Among AOE	2	9545.30	69.93	$\Phi_{ST} = 0.773$	<0.001
		Among Samples Within AOE	3	1449.18	7.41	$\Phi_{SC} = 0.247$	0.001
		Within Samples	10	2610.08	22.66	$\Phi_{CT} = 0.699$	<0.001
		Total	15	13604.56			
	Ecoregions	Among Ecoregions	3	6346.86	-3.86	$\Phi_{ST} = 0.736$	0.561
		Among Samples Within Ecoregions	2	4647.62	77.41	$\Phi_{SC} = 0.745$	0.011
		Within Samples	10	2610.08	26.45	$\Phi_{CT} = -0.039$	<0.001
		Total	15	13604.56			

Table 3: Continuation.

Species	Strata	Source of variation	d.f.	Sum of squares	Variation (%)	Φ	p-value
P. simonsi	**Structural Archs**	Among Archs	4	2,277.64	8.94	$\Phi_{ST} = 0.133$	0.088
		Among Samples Within Archs	2	921.02	4.34	$\Phi_{SC} = 0.048$	0.416
		Within Samples	8	3,306.33	86.72	$\Phi_{CT} = 0.089$	0.009
		Total	14	6,505.00			
	Areas of Endemism (AOE)	Among AOE	1	898.45	12.92	$\Phi_{ST} = 0.174$	0.000
		Among Samples Within AOE	5	2300.22	4.49	$\Phi_{SC} = 0.052$	0.117
		Within Samples	8	3306.33	82.58	$\Phi_{CT} = 0.129$	0.008
		Total	14	6505.00			
	Ecoregions	Among Ecoregions	2	1080.39	1.22	$\Phi_{ST} = 0.125$	0.423
		Among Samples Within Ecoregions	4	2118.28	11.28	$\Phi_{SC} = 0.114$	0.042
		Within Samples	8	3306.33	87.50	$\Phi_{CT} = 0.012$	0.010
		Total	14	6505.00			

Table 3: Continuation.

Species	Strata	Source of variation	d.f.	Sum of squares	Variation (%)	Φ	p-value
P. steerei	**Structural Archs**	Among Archs	3	5282.61	46.04	Φ_{ST} = 0.470	<0.001
		Among Samples Within Archs	2	953.21	1.00	Φ_{SC} = 0.019	0.066
		Within Samples	8	3652.00	52.96	Φ_{CT} = 0.460	<0.001
		Total	13	9887.82			
	Areas of Endemism (AOE)	Among AOE	1	1495.70	12.75	Φ_{ST} = 0.482	<0.001
		Among Samples Within AOE	4	4740.13	35.44	Φ_{SC} = 0.406	<0.001
		Within Samples	8	3652.00	51.81	Φ_{CT} = 0.127	<0.001
		Total	13	9887.82			
	Ecoregions	Among Ecoregions	2	2431.57	-3.23	Φ_{ST} = 0.426	0.479
		Among Samples Within Ecoregions	3	3804.25	45.88	Φ_{SC} = 0.444	0.015
		Within Samples	8	3652.00	57.35	Φ_{CT} = -0.032	<0.001
		Total	13	9887.82			

Table 4: Comparison between hypervolumes of the sympatric *Proechimys* species from Western Amazon. Lower diagonal values are Sørensen similarities indexes for the environmental hypervolumes for each pair of species, and in the upper diagonal in bold, the Sørensen similarity for the morphological hypervolumes. Values closer to 1 have greater similarity. M volume is the size of the morphological hypervolume for each species while E volume for the environmental hypervolume.

	P. brevicauda	***P. simonsi***	***P. steerei***
P. brevicauda	-	**0.002**	**0.022**
P. simonsi	0.757	-	**0.001**
P. steerei	0.461	0.521	-
M Volume	585.12	1133.24	892.65
E Volume	1772.44	1929.36	751.46

Table 5: Results summary for genetic structure, models of isolation by distance (IBD), by barriers (IBB), and by environment (IBE) and the hypervolume similarity in each sympatric species of *Proechimys*. See Results and Discussion sections for more details.

Models	Analyses	*P. brevicauda*	*P. simonsi*	*P. steerei*
	Genetic Structure	Deep	Shallow	Intermediate
IBD	**Mantel Test**	No effect	Significant without some populations	Significant with all populations
	Procrustes	Significant	Significant	Significant
IBB	**Structural Archs**	Significant (57.13%)	No effect	Significant (46.04%)
	Areas of Endemism	Significant (69.93%)	Significant (12.92%)	Significant (12.75%)
	Ecoregions	No effect	No effect	No effect
	Historical Fragmentation	Yes	No	No
	Association between Genetic Clusters & Stable Area Over Time	No	No	No
	Suitability Areas Reduction Over Time	No	No	No
IBE	**Partial Mantel**	No effect	No effect	Significant with all populations
	dbRDA	No effect	No effect	No effect
	Morphological Hypervolume	little similarity	little similarity	little similarity
	Climatic Hypervolume	great similarity	great similarity	great similarity

Figures

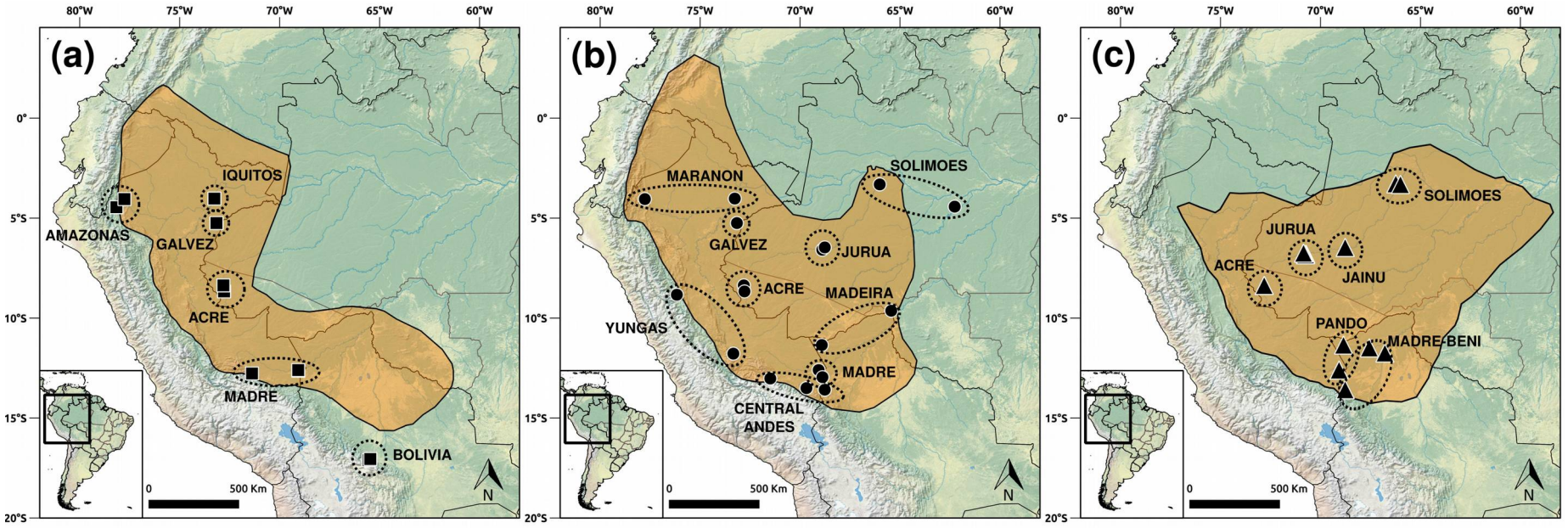


Figure 1: Geographical distribution of genetic samples used in the genomics analyses for *Proechimys brevicauda* (a), *Proechimys simonsi* (b), and *Proechimys steerei* (c) with the divisions into populations and their names (Table S1). Orange areas are the known distribution for these species, according to IUCN (www.iucnredlist.org). Shades of green represent vegetation, lighter tones is open areas and dark tones forests. Main rivers and the mountains are also represented.

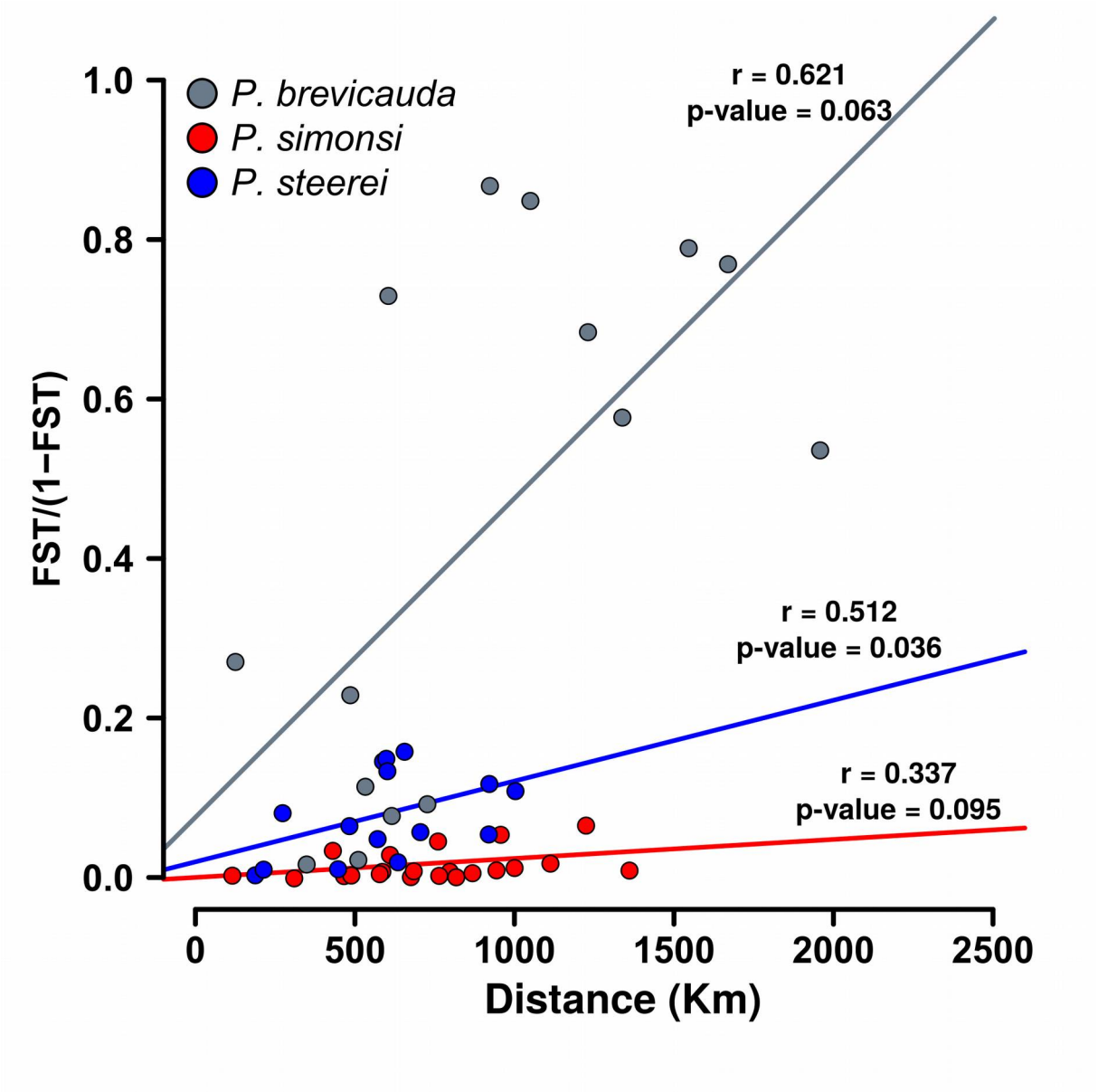


Figure 2: Isolation by distance for the three sympatric species of *Proechimys* in Western Amazon. Circles represent the pairwise comparisons between populations of the same species, lines are linear models between genetic and geographical distances, and the correlation coefficient (r) and p-value are also informed.

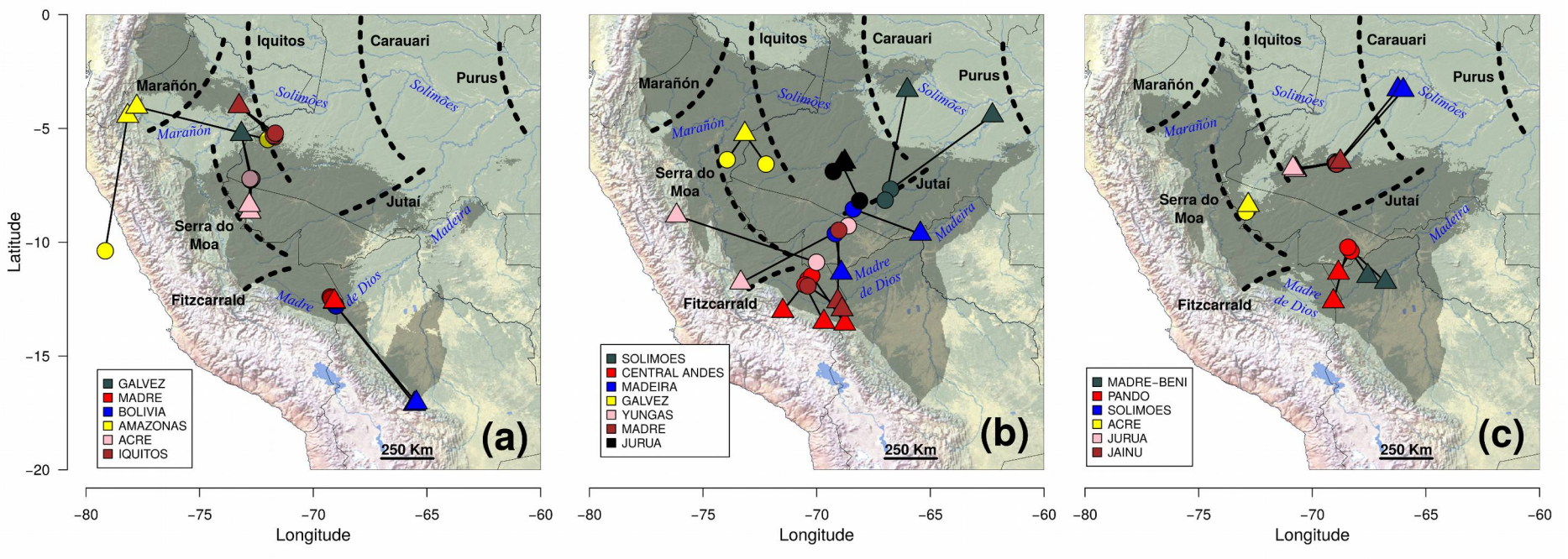


Figure 3: Procrustes analyses for the sympatric species: *P. brevicauda* (a), *P. simonsi* (b), and *P. steerei* (c). Colors represent different populations according to the legends boxes. Triangles represent the geographical localities, circles are the genetic distances among samples, and lines the deviation between geography and genetics distances. Greater isolation by distance, and consequently greater genetic structure, smaller are the lines size (the deviations between geography and genetics distances). Dotted lines indicate the structural archs locations with their respective names. Madre de Dios and Madeira rivers represent the division between Inambari and South Inambari areas of endemism, and Marañón and Solimões rivers between Inambari and Napo areas of endemism. Dark gray shades represent stable areas over time using the overlap between ENM for present-day, Middle Holocene, LGM, and LIG. Shadesof greens on the maps represent the differences in vegetation, with darker green for forest areas and lighter green/yellow, open areas.

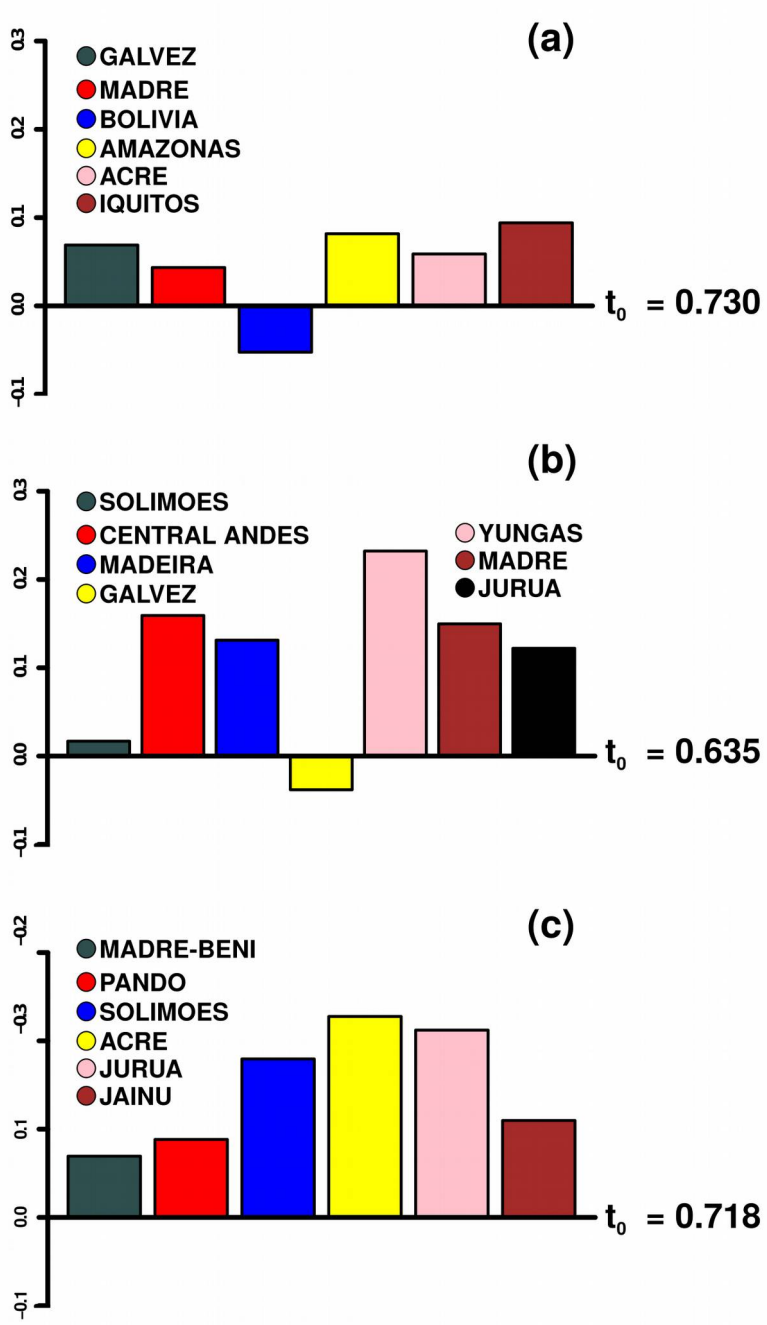


Figure 4: Variations in the correlation values between geography and genomic variation when one population are excluded at time in the Procrustes analyses for the three sympatric species: (a) *P. brevicauda*, (b) *P. simonsi*, and (c) *P. steerei*. Positive values indicate an increase in the association between geography and genomics when the population is excluded, and negative values indicate a decrease in the correlation. Correlation values with all populations is indicated at t_0. The colors represent the populations, the same ones were used in Figure 3.

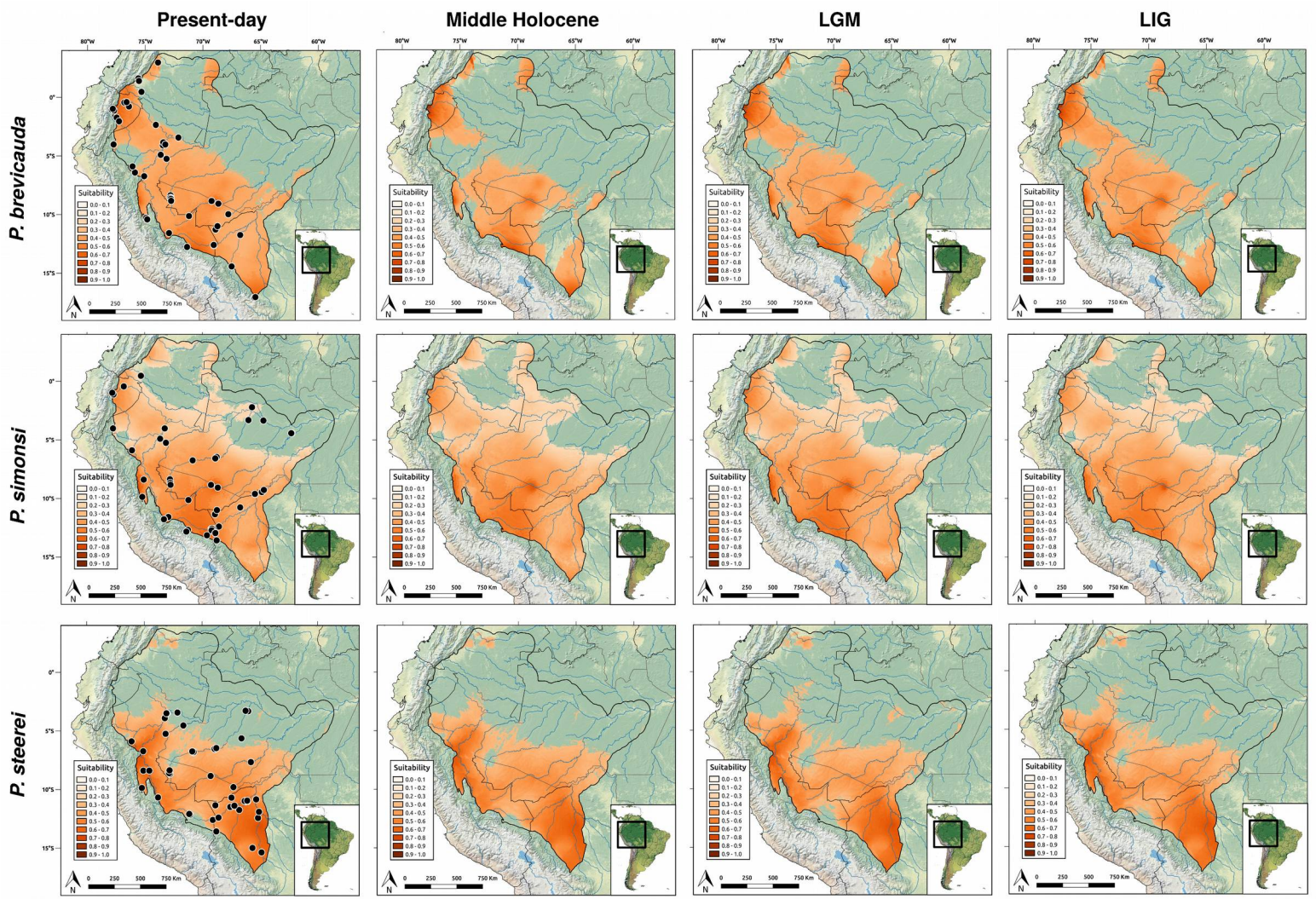

Present-day
Middle Holocene
LGM
LIG
P. brevicauda
P. simonsi
P. steerei
Suitability
0.0 - 0.1
0.1 - 0.2
0.2 - 0.3
0.3 - 0.4
0.4 - 0.5
0.5 - 0.6
0.6 - 0.7
0.7 - 0.8
0.8 - 0.9
0.9 - 1.0
80°W
75°W
70°W
65°W
60°W
0°
5°S
10°S
15°S
0 250 500 750 Km
N

Figure 5: Ecological niche models (ENM) for the sympatric species of *Proechimys* from Western Amazon: *P. brevicauda, P. simonsi* and *P. steerei*. Models are arranged from the oldest on the right to the present-day model on the left. The study area is within the thicker black line. Relief and major rivers are represented on the map to visualize the geographic landmarks. Dots on the present-day model represent the unbiased occurrence points used to build the models. Shades of greens on the maps represent the differences in vegetation, with darker green for forest area and lighter green/yellow, open areas. LIG = Last Interglacial, and LGM = Last Glacial Maximum.

Supplementary Material

Appendices

Appendix A: Details on preparation and sequencing of the genomic libraries based on the Peterson et al. (2012) protocol.

Genomic DNA from liver and muscle samples were extracted with DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA), following the manufacturer's recommendations, except for the DNA elution step where we used double distilled water (ddH_20) instead of elution buffer. Genomic DNA from skin and dry muscle samples were extracted following the same protocol as fresh tissues but with some modifications before the digestion step. In a sterile environment the hairs were removed from the skin samples. Afterwards, both the skin and dry muscle samples were hydrated for three days with ddH_20, replacing the water every 24 hours. After the hydration, the material was washed twice with 1X STE buffer (Bi *et al.* 2013), and then cut into small pieces to facilitate the digestion. During the digestion step I added 1 mM of dithiotreitol (DTT), a reducing agent, in 20 µL of volume per sample (Rohland & Hofreiter 2007). Extracted DNA was quantified by Qubit fluorometer (Life Technologies, Grand Island, NY, USA), and it was diluted with ddH_20 or concentrated in the SpeedVac Concentrator (ThermoFisher Scientific, Waltham, MA, USA) at 43º C (medium temperature) to reach the concentration of 17.6 ng/µL.

I followed the protocol from Peterson *et al.* (2012) for the preparation of genomic libraries (see Material and Methods for details about the number of samples and libraries) using the ddRAD-Seq technique. In this approach 300 ng of genomic DNA (*i.e.*, 17 µL of extracted DNA) were cut in variable-sized fragments, using two restriction enzymes: *Eco*-RI and *Mse*-I. The resulting solution was cleaned with commercial Ampure XP Beads (Beckman Coulter, Brea, CA, USA) and quantified in Qubit fluorometer (Life Technologies, Grand Island, NY, USA). Then, I used 50 ng of fragmented DNA in a volume of 33 µL per sample for the ligation step, in which the ends of the fragmented DNA were bonded to the Illumina adapters and a unique barcode per sample. After the reaction samples were pooled together and the solution was cleaned again with commercial Ampure XP Beads (Beckman Coulter, Brea, CA, USA). DNA fragments were automatically selected by size (between 350 and 450 bp) through Pippin Prep (Sage Science, Beverly, MA, USA) and selected fragments were

amplified by PCR. The libraries were cleaned with the beads again, quantified and sequenced. All libraries were sequenced in three lanes of HiSeq2500 (Illumina, San Diego, CA, USA) according to instructions of the manufacturer to generate 150 base pairs, single-end reads in the Hospital for Sick Children (Toronto, Ontario, Canada).

Bi, K., Linderoth, T., Vanderpool, D., Good, J.M., Nielsen, R. & Moritz, C. (2013) Unlocking the vault: next-generation museum population genomics. *Molecular Ecology* 22, 6018–6032.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species L. Orlando (Ed). *PLoS ONE* 7, e37135.

Rohland, N. & Hofreiter, M. (2007) Ancient DNA extraction from bones and teeth. *Nature Protocols* 2, 1756–1762.

Appendix B: Details on reads processing for phylogeographic analyses using the STACKS pipeline.

For studies of population genomics and phylogeography I used the STACKS 1.45 pipeline (Catchen *et al.* 2013). I processed the database separately for each species, and STACKS 1.45 processed the raw sequences in 5 steps. In the first step, the *process_radtags* demultiplexed the reads according to the barcodes list (parameter *-b*), with value 2 for the distance allowed between barcodes (*--barcode_dist*). Reads without barcodes, with uncalled nucleotides, with deficient restriction enzyme cut sites, or with low quality scores for Illumina, Phed ≤ 32 (*-E*) were excluded. At the end of this step the high-quality reads had 140 bp (without 10 pb of barcodes) and they were separated into individual files by individuals. The following step, *ustacks*, the reads of each individual were aligned by the *de novo* approach, and arranged in stacks with identical reads by individual. I excluded stacks with lower than 6 reads (*-m*) and I merged the stacks in loci, allowing up to 3 stacks by loci (*-max_locus_stacks*) and 3 nucleotides of distance between stacks (*-M*) (Paris *et al.* 2017). At the end of this stage, we had a set of putative loci with polymorphism and alleles inferred per those loci. *cstacks*, the third step, grouped the loci across all individuals and create a unique catalog with all loci for each one of the target species. When two loci were grouped, *cstacks* joined their SNP in the catalog, with 3 fixed differences expected between individuals (*-n*) (Paris *et al.* 2017). In the fourth step, *sstacks,* the loci from each individual were compared to all catalog loci, recording the matches into a new file. Individual loci similar to more than one loci of the catalog were excluded because their information is ambiguous. The *populations* program was the last step and it processed the reads individually using the same catalog-matched data created on the previous step. In this step I split individuals into different populations (see Material and Methods for details). Only the loci present in at least two populations in each lineage (*-p*), and with minimum depth of coverage for each loci equal to 6 (*-m*) were used to create the output in Variant Call Format (VCF). The VCF was edited, eliminating the very variable loci using a script in R plataform (R-Development CoreTeam 2018), and removing loci and individuals up to 20% of missing data in PLINK 1.9 (Purcell *et al.* 2007). *populations* program was performed again using the filtered data, and only one SNP per loci randomly chosen (*--write_random_snp*) was used to create the outputs for the subsequent analyzes.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22, 3124–3140.

Paris, J.R., Stevens, J.R. & Catchen, J.M. (2017) Lost in parameter space: a road map for stacks S. Johnston (Ed). *Methods in Ecology and Evolution* 8, 1360–1373.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. & Sham, P.C. (2007) PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* 81, 559–575.

R-Development CoreTeam (2018) R Development Core Team. *R: A Language and Environment for Statistical Computing* 55, 275–286.

Tables

Table S1: 52 samples used in the genomics analyzes with information about the locality, species, institution of origin of the samples, populations, and strata used in the analyses of molecular variance (AMOVA): Endesmism, Ecoregions and Structural Archs. **AMNH-AMCC:** Ambrose Monell Cryo Collection, American Museum of Natural History, New York, USA; **FMNH:** Field Museum of Natural History, Chicago, Illinois, USA; **LMUSP:** Laboratório de Mamíferos da Escola Superior de Agricultura "Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil; **MSB:** Museum of Southwestern Biology, Alburqueque, New Mexico, USA; **MVZ:** Museum of Vertebrate Zoology, Berkeley, California, USA; **MZUSP:** Museu de Zoologia da Universidade de São Paulo, São Paulo, Brazil; **NMNH:** National Museum of Natural History, Washington, D.C., USA; **TTU:** Texas Tech University, Lubbock, Texas, USA. Table with more information about locality in ***.csv*** format is available on https://github.com/jdalapicolla.

Catalog Number	Source	Species	Populations	Endemism	Ecoregions	Structural Archs	Longitude	Latitude
AMNH272698	AMNH-AMCC	*P. brevicauda*	GALVEZ	Inambari	Southwest Amazon moist forests	Iquitos – Serra do Moa	-73.162	-5.250
AMNH272700	AMNH-AMCC	*P. brevicauda*	GALVEZ	Inambari	Southwest Amazon moist forests	Iquitos – Serra do Moa	-73.162	-5.250
MUSM13338	AMNH-AMCC	*P. brevicauda*	GALVEZ	Inambari	Southwest Amazon moist forests	Iquitos – Serra do Moa	-73.162	-5.250
FMNH175255	FMNH	*P. brevicauda*	MADRE	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-71.385	-12.772
MSB238391	MSB	*P. brevicauda*	BOLIVIA	South Inambari	Bolivian Yungas	South Fitzcarrald	-65.550	-17.107
MSB70574	MSB	*P. brevicauda*	BOLIVIA	South Inambari	Bolivian Yungas	South Fitzcarrald	-65.550	-17.107

Table S1: Continuation.

Catalog Number	Source	Species	Populations	Endemism	Ecoregions	Structural Archs	Longitude	Latitude
MSB70575	MSB	*P. brevicauda*	BOLIVIA	South Inambari	Bolivian Yungas	South Fitzcarrald	-65.467	-17.050
MVZ194439	MVZ	*P. brevicauda*	ACRE	Inambari	Iquitos várzea	Iquitos – Serra do Moa	-72.783	-8.667
MVZ194463	MVZ	*P. brevicauda*	ACRE	Inambari	Iquitos várzea	Iquitos – Serra do Moa	-72.817	-8.367
MVZ194485	MVZ	*P. brevicauda*	ACRE	Inambari	Iquitos várzea	Iquitos – Serra do Moa	-72.817	-8.367
MVZ155121	MVZ	*P. brevicauda*	AMAZONAS	Napo	Ucayali and Napo moist forests	Marañón – Andes	-78.168	-4.457
MVZ157905	MVZ	*P. brevicauda*	AMAZONAS	Napo	Ucayali and Napo moist forests	Marañón – Andes	-77.767	-4.050
MVZ157855	MVZ	*P. brevicauda*	MADRE	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-69.070	-12.633
MVZ168953	MVZ	*P. brevicauda*	MADRE	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-69.073	-12.600
TTU101179	TTU	*P. brevicauda*	IQUITOS	Napo	Iquitos várzea	Iquitos – Serra do Moa	-73.268	-4.024
TTU101195	TTU	*P. brevicauda*	IQUITOS	Napo	Iquitos várzea	Iquitos – Serra do Moa	-73.268	-4.024
TTU101213	TTU	*P. brevicauda*	IQUITOS	Napo	Iquitos várzea	Iquitos – Serra do Moa	-73.268	-4.024
MUSM13339	AMNH-AMCC	*P. simonsi*	GALVEZ	Inambari	Southwest Amazon moist forests	Iquitos – Serra do Moa	-73.162	-5.250

Table S1: Continuation.

Catalog Number	Source	Species	Populations	Endemism	Ecoregions	Structural Archs	Longitude	Latitude
MUSM13342	AMNH-AMCC	*P. simonsi*	GALVEZ	Inambari	Southwest Amazon moist forests	Iquitos – Serra do Moa	-73.162	-5.250
FMNH175275	FMNH	*P. simonsi*	CENTRAL_ANDES	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-71.492	-13.024
EFA37	LMUSP	*P. simonsi*	SOLIMOES	Inambari	Juruá-Purus moist forests	Purus – Carauari	-62.267	-4.428
MSB236594	MSB	*P. simonsi*	MADEIRA	Inambari	Southwest Amazon moist forests	South Fitzcarrald	-68.917	-11.350
MVZ194602	MVZ	*P. simonsi*	ACRE	Inambari	Iquitos várzea	Iquitos – Serra do Moa	-72.817	-8.367
MVZ194635	MVZ	*P. simonsi*	ACRE	Inambari	Iquitos várzea	Iquitos – Serra do Moa	-72.783	-8.667
MVZ194711	MVZ	*P. simonsi*	SOLIMOES	Inambari	Juruá-Purus moist forests	Purus – Carauari	-66.017	-3.317
MVZ166815	MVZ	*P. simonsi*	YUNGAS	Inambari	Peruvian Yungas	Serra do Moa – Andes	-76.167	-8.833
MVZ194703	MVZ	*P. simonsi*	JURUA	Inambari	Southwest Amazon moist forests	Carauari – Iquitos	-68.892	-6.583
MVZ194775	MVZ	*P. simonsi*	JURUA	Inambari	Southwest Amazon moist forests	Carauari – Iquitos	-68.767	-6.467
MVZ157968	MVZ	*P. simonsi*	MARANON	Napo	Napo moist forests	Marañón – Andes	-77.767	-4.050
MVZ168955	MVZ	*P. simonsi*	MADRE	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-69.073	-12.600

Table S1: Continuation.

Catalog Number	Source	Species	Populations	Endemism	Ecoregions	Structural Archs	Longitude	Latitude
MJ529	MZUSP	*P. simonsi*	MADEIRA	Inambari	Southwest Amazon moist forests	South Fitzcarrald	-65.440	-9.634
TTU101118	TTU	*P. simonsi*	MARANON	Napo	Iquitos várzea	Iquitos – Serra do Moa	-73.268	-4.024
USNM619007	NMNH	*P. simonsi*	YUNGAS	Inambari	Peruvian Yungas	Serra do Moa – Andes	-73.341	-11.780
USNM579697	NMNH	*P. simonsi*	CENTRAL_ANDES	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-69.684	-13.504
USNM619001	NMNH	*P. simonsi*	CENTRAL_ANDES	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-68.767	-13.583
USNM619003	NMNH	*P. simonsi*	CENTRAL_ANDES	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-69.683	-13.505
USNM619008	NMNH	*P. simonsi*	MADRE	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-68.882	-12.957
MSB236698	MSB	*P. steerei*	PANDO	Inambari	Southwest Amazon moist forests	South Fitzcarrald	-68.850	-11.351
MSB236689	MSB	*P. steerei*	MADRE_BENI	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-67.560	-11.490
MSB236807	MSB	*P. steerei*	MADRE_BENI	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-66.780	-11.749
MVZ194874	MVZ	*P. steerei*	ACRE	Inambari	Iquitos várzea	Iquitos – Serra do Moa	-72.817	-8.367

Table S1: Continuation.

Catalog Number	Source	Species	Populations	Endemism	Ecoregions	Structural Archs	Longitude	Latitude
MVZ194879	MVZ	*P. steerei*	ACRE	Inambari	Iquitos várzea	Iquitos – Serra do Moa	-72.817	-8.367
MVZ195034	MVZ	*P. steerei*	JAINU	Inambari	Southwest Amazon moist forests	Carauari – Iquitos	-68.767	-6.467
MVZ195036	MVZ	*P. steerei*	JAINU	Inambari	Southwest Amazon moist forests	Carauari – Iquitos	-68.767	-6.467
MVZ194909	MVZ	*P. steerei*	JURUA	Inambari	Iquitos várzea	Carauari – Iquitos	-70.734	-6.800
MVZ194914	MVZ	*P. steerei*	JURUA	Inambari	Iquitos várzea	Carauari – Iquitos	-70.734	-6.800
MVZ194987	MVZ	*P. steerei*	JURUA	Inambari	Iquitos várzea	Carauari – Iquitos	-70.750	-6.833
MVZ194997	MVZ	*P. steerei*	JURUA	Inambari	Iquitos várzea	Carauari – Iquitos	-70.850	-6.750
MVZ168942	MVZ	*P. steerei*	PANDO	Inambari	Southwest Amazon moist forests	South Fitzcarrald	-69.073	-12.600
MVZ190951	MVZ	*P. steerei*	SOLIMOES	Inambari	Juruá-Purus moist forests	Purus – Carauari	-66.233	-3.283
MVZ190954	MVZ	*P. steerei*	SOLIMOES	Inambari	Juruá-Purus moist forests	Purus – Carauari	-66.000	-3.317
USNM619002	NMNH	*P. steerei*	MADRE_BENI	South Inambari	Southwest Amazon moist forests	South Fitzcarrald	-68.767	-13.583

Table S2: 19 environmental variables used in the ENM analyses.

Variables	Description
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

Table S3: 138 unbiased occurrence points for the three sympatric species of *Proechimys* in Western Amazon used in ENM analyses, with information about the locality, voucher and on what dataset the identification at the specific level was based: morphology, genomic (Chapter 1), or mtDNA. **AMNH:** American Museum of Natural History, New York, USA; **FMNH:** Field Museum of Natural History, Chicago, Illinois, USA; **LMUSP:** Laboratório de Mamíferos da Escola Superior de Agricultura "Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil; **MCZ:** Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts, USA; **MN:** Museu Nacional da Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; **MPEG**: Museu Paraense Emílio Goeldi, Belém, Pará, Brazil; **MSB:** Museum of Southwestern Biology, Alburqueque, New Mexico, USA; **MVZ:** Museum of Vertebrate Zoology, Berkeley, California, USA; **MZUSP:** Museu de Zoologia da Universidade de São Paulo, São Paulo, Brazil; **NMNH:** National Museum of Natural History, Washington, D.C., USA; **TTU:** Texas Tech University, Lubbock, Texas, USA; **UFMG:** Coleção de Mamíferos, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; **UFPB:** Univerisade Federal da Paraíba, João Pessoa, Paraíba, Brazil. **UMMZ:** University of Michigan Museum of Zoology, Ann Arbor, Michigan, USA. Table with more information about locality in ***.csv*** format is available on https://github.com/jdalapicolla.

Species	Longitude	Latitude	Voucher	Based on	Source	Locality
P. brevicauda	-73.9006	2.9691	AMNH142145	morphology	AMNH	Colombia: Meta: Serranía de la Macarena
P. brevicauda	-74.9349	-10.2973	AMNH213482	morphology	AMNH	Peru: Pasco: Puerto Bermudez
P. brevicauda	-74.8333	-10.4168	AMNH230869	morphology	AMNH	Peru: Pasco: Nevati Mission
P. brevicauda	-67.5221	-14.4336	AMNH247905	morphology	AMNH	Bolivia: Beni: Rurrenabaque
P. brevicauda	-68.9167	-11.2833	AMNH249059	morphology	AMNH	Bolivia: Pando: Nareuda River
P. brevicauda	-73.1621	-5.2495	AMNH268275	morphology	AMNH	Peru: Loreto: Nuevo San Juan, Río Galvez
P. brevicauda	-75.6044	1.6151	AMNH33709	morphology	AMNH	Colombia: Caquetá
P. brevicauda	-77.4839	-1.7345	AMNH67322	morphology	AMNH	Ecuador: Pastaza: Sarayacu
P. brevicauda	-74.0879	-2.3516	AMNH71885	morphology	AMNH	Peru: Loreto: Boca do Rio Curaray
P. brevicauda	-72.1333	-3.4333	AMNH73797	morphology	AMNH	Peru: Loreto: Orosa
P. brevicauda	-75.0996	-6.7333	AMNH75275	morphology	AMNH	Peru: Loreto: Sarayacu

Table S3: Continuation.

Species	Longitude	Latitude	Voucher	Based on	Source	Locality
P. brevicauda	-75.9000	-6.4167	AMNH98249	morphology	AMNH	Peru: San Martín: Achinamisa, Rio Huallaga
P. brevicauda	-71.3854	-12.7717	FMNH175255	genomic (this study)	FMNH	Peru: Madre de Dios: Maskoitania
P. brevicauda	-75.3333	0.4667	FMNH71165	morphology	FMNH	Colombia: Putumayo: Rio Mecaya
P. brevicauda	-75.5500	1.4000	FMNH71173	morphology	FMNH	Colombia: Caquetá: Montanita
P. brevicauda	-68.7468	-10.9984	LMUSP210	morphology	LMUSP	Brazil: Acre: Reserva Extrativista Chico Mendes
P. brevicauda	-77.8975	-1.0741	MCZ37893	morphology	MCZ	Ecuador: Oriente, Rio JatunYacu [=Rio Jatunyacu]
P. brevicauda	-76.1094	-5.8948	MN35874	morphology	MN	Peru: Loreto: Yurimaguás
P. brevicauda	-73.4774	-4.0951	MN35875	morphology	MN	Peru: Loreto: Estrada Iquitos-Nauta
P. brevicauda	-73.6674	-4.9063	MN35880	morphology	MN	Peru: Loreto: Genaro Herrera, Rio Ucayali
P. brevicauda	-73.4427	-3.8435	MN35882	morphology	MN	Peru: Loreto: Mishana, Rio Nanay
P. brevicauda	-77.2746	-2.0444	MN46789	morphology	MN	Ecuador: Pastaza: Rio Bobonaza
P. brevicauda	-76.4167	-0.7833	MN75774	morphology	MN	Ecuador: Napo: San Francisco
P. brevicauda	-69.2607	-8.8474	MPEG10666	morphology	MPEG	Brazil: Acre: BR-364, Km 08
P. brevicauda	-68.6714	-9.0792	MPEG10820	morphology	MPEG	Brazil: Acre: Bairro do Triângulo
P. brevicauda	-72.8000	-8.5664	MPEG28347	morphology	MPEG	Brazil: Acre: Opposite Ocidente, left bank Rio Juruá
P. brevicauda	-72.8509	-8.6004	MPEG28349	morphology	MPEG	Brazil: Acre: Flora (=Fazenda Santa Fé), left bank Rio Juruá
P. brevicauda	-72.8167	-8.3667	MPEG28363	morphology	MPEG	Brazil: Acre: Sobral, left bank Rio Juruá
P. brevicauda	-72.7633	-8.8339	MPEG775	morphology	MPEG	Brazil: Acre: Seringal Oriente
P. brevicauda	-66.7797	-11.7495	MSB236806	genomic (this study)	MSB	Bolivia: Beni: Boca del Rio Biata

Table S3: Continuation.

Species	Longitude	Latitude	Voucher	Based on	Source	Locality
P. brevicauda	-65.5500	-17.1000	MSB70574	genomic (this study)	MSB	Bolivia: Cochabamba: El Palmar, Rio Cochi Mayu
P. brevicauda	-65.4667	-17.0500	MSB70575	genomic (this study)	MSB	Bolivia: Cochabamba: Villa Tunari
P. brevicauda	-71.2167	-10.1333	MVZ136648	morphology	MVZ	Peru: Ucayali: Balta, Rio Curanja
P. brevicauda	-69.0680	-12.6333	MVZ157855	genomic (this study)	MVZ	Peru: Madre de Dios: Lago Sandoval, Rio Madre de Dios
P. brevicauda	-77.7510	-4.0220	MVZ157905	genomic (this study)	MVZ	Peru: Amazonas: La Poza, Rio Santiago
P. brevicauda	-69.0729	-12.6000	MVZ168956	morphology	MVZ	Peru: Madre de Dios: Albergue Cusco Amazonica
P. brevicauda	-72.7830	-8.6666	MVZ194439	genomic (this study)	MVZ	Brazil: Acre: Igarapé Porongaba, right bank Rio Juruá
P. brevicauda	-73.2684	-4.0240	TTU101213	genomic (this study)	TTU	Peru: Loreto: Estacion Biologica Allpahuayo
P. brevicauda	-67.8099	-9.9747	UFMG1616	morphology	UFMG	Brazil: Acre: Fazenda Humaitá
P. brevicauda	-77.7105	-1.1304	UMMZ164870	morphology	UMMZ	Ecuador: Napo: Rio Shalcana
P. brevicauda	-77.8170	-0.9830	UMMZ80047	morphology	UMMZ	Ecuador: Napo: San Francisco, Rio Napo, Aguarico
P. brevicauda	-76.8172	-0.4400	UMMZ80065	morphology	UMMZ	Ecuador: Napo: Llunchi, Rio Napo, Napo-Pastaza
P. brevicauda	-76.6242	-0.4106	USNM513654	morphology	NMNH	Ecuador: Napo: Limoncocha
P. brevicauda	-72.9499	-11.5834	USNM582993	morphology	NMNH	Peru: Cusco: Camisea, San Martin 3
P. simonsi	-68.5754	-12.3986	AMNH263060	morphology	AMNH	Bolivia: Pando: Isla Gargantua
P. simonsi	-68.9168	-11.3501	AMNH263062	morphology	AMNH	Bolivia: Pando: Santa Rosa
P. simonsi	-73.1621	-5.2495	AMNH272677	mtDNA	Schetino 2008	Peru: Loreto: Nuevo San Juan, Río Galvez
P. simonsi	-62.2675	-4.4280	EFA037	genomic (this study)	LMUSP	Brazil: Amazonas: Rio Purus
P. simonsi	-62.3071	-4.4315	EFA039	morphology	LMUSP	Brazil: Amazonas: Rio Purus, Lago Ayapuá

Table S3: Continuation.

Species	Longitude	Latitude	Voucher	Based on	Source	Locality
P. simonsi	-75.3333	0.4667	FMNH71154	morphology	FMNH	Colombia: Putumayo: Rio Mecaya
P. simonsi	-68.7468	-10.9984	LMUSP208	genomic (this study)	LMUSP	Brazil: Acre: Reserva Extrativista Chico Mendes
P. simonsi	-65.4495	-9.6383	MJ251	morphology	MZUSP	Brazil: Rondônia: Abunã
P. simonsi	-64.8329	-9.4378	MJ397	morphology	MZUSP	Brazil: Rondônia: Caiçara
P. simonsi	-64.8600	-9.4567	MJ549	morphology	MZUSP	Brazil: Rondônia: Caiçara
P. simonsi	-65.4635	-9.6204	MJ572	morphology	MZUSP	Brazil: Rondônia: Abunã
P. simonsi	-65.4550	-9.6120	MJ833	morphology	MZUSP	Brazil: Rondônia: Abunã
P. simonsi	-73.6674	-4.9063	MN35881	morphology	MN	Peru: Loreto: Genaro Herrera, Rio Ucayali
P. simonsi	-69.2607	-8.8474	MPEG10633	morphology	MPEG	Brazil: Acre: BR-364, Km 08
P. simonsi	-68.6714	-9.0792	MPEG10803	morphology	MPEG	Brazil: Acre: Bairro do Triângulo
P. simonsi	-64.7199	-3.3539	MPEG22871	morphology	MPEG	Brazil: Amazonas: Tefé
P. simonsi	-65.7088	-2.2169	MPEG26357	morphology	MPEG	Brazil: Amazonas: Estação Ecológica Mamirauá
P. simonsi	-72.8166	-8.3667	MPEG28400	morphology	MPEG	Brazil: Acre: Sobral, left bank Rio Juruá
P. simonsi	-66.0167	-3.3167	MPEG28412	morphology	MPEG	Brazil: Amazonas: Lago Vai-quem-quer, right bank, Rio Juruá
P. simonsi	-70.8500	-6.7500	MPEG28424	morphology	MPEG	Brazil: Amazonas: Seringal Condor, left bank Rio Juruá
P. simonsi	-68.8922	-6.5828	MPEG28432	morphology	MPEG	Brazil: Amazonas: Altamira, right bank Rio Juruá
P. simonsi	-72.7830	-8.6666	MPEG28456	morphology	MPEG	Brazil: Acre: Igarapé Porongaba, right bank Rio Juruá
P. simonsi	-71.2167	-10.1333	MVZ136650	morphology	MVZ	Peru: Ucayali: Balta, Rio Curanja
P. simonsi	-77.7510	-4.0220	MVZ157968	genomic (this study)	MVZ	Peru: Amazonas: La Poza, Rio Santiago

Table S3: Continuation.

Species	Longitude	Latitude	Voucher	Based on	Source	Locality
P. simonsi	-69.0729	-12.6000	MVZ168955	morphology	MVZ	Peru: Madre de Dios: Albergue Cusco Amazonica
P. simonsi	-68.7667	-6.4667	MVZ194775	genomic (this study)	MVZ	Brazil: Amazonas: Barro Vermelho, left bank Rio Juruá
P. simonsi	-72.7633	-8.8339	MZUSP25837	morphology	MZUSP	Brazil: Acre: Seringal Oriente
P. simonsi	-64.6795	-9.2769	SA11659	morphology	MZUSP	Brazil: Rondônia: Resgate UHE Jirau
P. simonsi	-73.2684	-4.0240	TTU101118	genomic (this study)	TTU	Peru: Loreto: Iquitos, 25 Km S, Estacion Biologica Allpahuayo
P. simonsi	-76.1094	-5.8948	UFPB2990	morphology	UFPB	Peru: Loreto: Yurimaguás
P. simonsi	-71.3938	-12.8298	UMMZ160537	morphology	UMMZ	Peru: Madre de Dios: Hacienda Erika
P. simonsi	-71.2617	-12.6800	UMMZ160538	morphology	UMMZ	Peru: Madre de Dios: Aguas Calientes
P. simonsi	-77.7105	-1.1304	UMMZ164869	morphology	UMMZ	Ecuador: Napo: Rio Shalcana
P. simonsi	-76.8172	-0.4400	UMMZ80043	morphology	UMMZ	Ecuador: Napo: Llunchi, Rio Napo, Napo-Pastaza
P. simonsi	-77.8170	-0.9830	UMMZ80046	morphology	UMMZ	Ecuador: Napo: San Francisco, Rio Napo, Aguarico
P. simonsi	-69.2374	-12.5966	USNM364151	morphology	NMNH	Peru: Madre de Dios: Tambopata
P. simonsi	-75.2163	-9.8671	USNM364511	morphology	NMNH	Peru: Pasco: San Juan
P. simonsi	-69.1834	-12.5880	USNM390367	morphology	NMNH	Peru: Madre de Dios: Puerto Maldonado
P. simonsi	-75.0880	-8.3941	USNM461292	morphology	NMNH	Peru: Coronel Portillo: Pucallpa
P. simonsi	-69.2070	-12.7205	USNM530935	morphology	NMNH	Peru: Madre de Dios: Rio Tambopata
P. simonsi	-68.7700	-13.5800	USNM579259	morphology	NMNH	Bolivia: La Paz: Alto Rio Madidi
P. simonsi	-66.7333	-10.7667	USNM579616	morphology	NMNH	Bolivia: Pando: San Juan De Nuevo Mundo

Table S3: Continuation.

Species	Longitude	Latitude	Voucher	Based on	Source	Locality
P. simonsi	-69.6122	-13.1472	USNM579695	morphology	NMNH	Peru: Madre de Dios: Colpa De Guacamayos, Rio Tambopata
P. simonsi	-68.9022	-12.9589	USNM579698	morphology	NMNH	Peru: Puno: “Aguas Claras” Camp
P. simonsi	-72.9499	-11.5834	USNM582772	morphology	NMNH	Peru: Cusco: Camisea, Pagoreni
P. simonsi	-73.3258	-11.7667	USNM588063	morphology	NMNH	Peru: Cusco: Tangoshiari
P. simonsi	-73.3406	-11.7794	USNM619007	genomic (this study)	NMNH	Peru: Cusco: Cordillera de Vilcabamba
P. simonsi	-68.8817	-12.9566	USNM619008	genomic (this study)	NMNH	Peru: Madre de Dios: Santuario Nacional Pampas del Heath
P. steerei	-74.5688	-8.3931	AMNH147499	morphology	AMNH	Peru: Ucayali: Pucallpa
P. steerei	-65.3132	-10.8469	AMNH210348	morphology	AMNH	Bolivia: Beni: Guayaramerin
P. steerei	-64.8499	-15.3709	AMNH214648	morphology	AMNH	Bolivia: Beni: Buena Hora
P. steerei	-65.1551	-12.4315	AMNH214651	morphology	AMNH	Bolivia: Beni: Mamore River
P. steerei	-65.0679	-11.9060	AMNH214665	morphology	AMNH	Bolivia: Beni: Lagoinha
P. steerei	-75.2163	-9.8671	AMNH230895	morphology	AMNH	Peru: Pasco: San Juan
P. steerei	-67.2000	-11.3833	AMNH263056	morphology	AMNH	Bolivia: Pando: Bella Vista
P. steerei	-67.2167	-11.4000	AMNH263057	morphology	AMNH	Bolivia: Pando: La Cruz
P. steerei	-66.3167	-10.9833	AMNH263088	morphology	AMNH	Bolivia: Pando: Rio Madre de Dios
P. steerei	-68.5754	-12.3986	AMNH263112	morphology	AMNH	Bolivia: Pando: Isla Gargantua
P. steerei	-67.4500	-10.7000	AMNH263116	morphology	AMNH	Bolivia: Pando: Remanso
P. steerei	-73.1621	-5.2495	AMNH268279	morphology	AMNH	Peru: Loreto: Nuevo San Juan, Río Galvez
P. steerei	-73.0472	-3.5014	AMNH73288	morphology	AMNH	Peru: Loreto: Puerto Indiana

Table S3: Continuation.

Species	Longitude	Latitude	Voucher	Based on	Source	Locality
P. steerei	-72.1333	-3.4333	AMNH73803	morphology	AMNH	Peru: Loreto: Orosa
P. steerei	-73.8156	-10.6817	AMNH76072	morphology	AMNH	Peru: Ucayali: Santa Rosa
P. steerei	-75.0996	-6.7333	AMNH76261	morphology	AMNH	Peru: Loreto: Sarayacu
P. steerei	-73.0885	-3.4972	AMNH98797	morphology	AMNH	Peru: Loreto: Rio Panduro
P. steerei	-76.1094	-5.8948	MN35870	morphology	MN	Peru: Loreto: Yurimaguás
P. steerei	-73.2282	-3.9134	MN35877	morphology	MN	Peru: Loreto: Isla Muyuy
P. steerei	-72.7830	-8.6666	MNFS1136	mtDNA	Schetino 2008	Brazil: Acre: Igarapé Porongaba, right bank Rio Juruá
P. steerei	-69.2607	-8.8474	MPEG10642	morphology	MPEG	Brazil: Acre: BR-364, Km 08
P. steerei	-71.6237	-4.5357	MPEG1616	morphology	MPEG	Brazil: Amazonas: Estirão do Equador, Rio Javari
P. steerei	-70.7336	-6.8000	MPEG28537	morphology	MPEG	Brazil: Amazonas: Igarapé Nova Empresa, left bank, Rio Juruá
P. steerei	-70.8500	-6.7500	MPEG28558	morphology	MPEG	Brazil: Acre: Sacado (Condor), right bank Rio Juruá
P. steerei	-68.9165	-6.5336	MPEG28567	morphology	MPEG	Brazil: Amazonas: Boa Esperança, right bank, Rio Juruá
P. steerei	-72.8509	-8.6004	MPEG28570	morphology	MPEG	Brazil: Acre: Flora (=Fazenda Santa Fé), left bank Rio Juruá
P. steerei	-72.8166	-8.3667	MPEG28573	morphology	MPEG	Brazil: Acre: Nova Vida, right bank Rio Juruá
P. steerei	-66.2167	-11.0167	MSB211815	genomic (this study)	MSB	Bolivia: Pando: Agua Dulce
P. steerei	-67.5602	-11.4900	MSB236689	genomic (this study)	MSB	Bolivia: Pando: Opposite Independencia
P. steerei	-68.8498	-11.3506	MSB236698	genomic (this study)	MSB	Bolivia: Pando: La Cruz
P. steerei	-66.7797	-11.7495	MSB236807	genomic (this study)	MSB	Bolivia: Beni: Boca del Rio Biata
P. steerei	-69.0729	-12.6000	MVZ168943	morphology	MVZ	Peru: Madre de Dios: Albergue Cusco Amazonica

Table S3: Continuation.

Species	Longitude	Latitude	Voucher	Based on	Source	Locality
P. steerei	-66.2333	-3.2833	MVZ190951	genomic (this study)	MVZ	Brazil: Amazonas: Ilhazinha
P. steerei	-66.0000	-3.3167	MVZ190954	genomic (this study)	MVZ	Brazil: Amazonas: Ilha Paxiuba, right bank Rio Juruá
P. steerei	-70.7501	-6.8334	MVZ194987	genomic (this study)	MVZ	Brazil: Amazonas: Penedo, right bank Rio Juruá
P. steerei	-68.7667	-6.4667	MVZ195036	genomic (this study)	MVZ	Brazil: Amazonas: Jainu, right bank, Rio Juruá
P. steerei	-67.2833	-9.8000	MZUSP7347	morphology	MZUSP	Brazil: Acre: Iquiri [Colocação Iquiri, Rio Branco, AC]
P. steerei	-66.5853	-5.6364	UFMG6021	morphology	UFMG	Brazil: Amazonas: BAPE Suruwahá
P. steerei	-65.7716	-7.6638	USNM105537	morphology	NMNH	Brazil: Amazonas: Purus River, Hyantanhan [=Huitanaã]
P. steerei	-66.0913	-10.9525	USNM364009	morphology	NMNH	Bolivia: Pando: Manuripi
P. steerei	-66.0647	-10.9991	USNM391660	morphology	NMNH	Bolivia: Beni: Riberalta, Vaca Diez
P. steerei	-65.6349	-14.9876	USNM391665	morphology	NMNH	Bolivia: Beni: San Ignacio
P. steerei	-75.0880	-8.3941	USNM461342	morphology	NMNH	Peru: Coronel Portillo: Pucallpa
P. steerei	-71.0451	-12.1561	USNM530931	morphology	NMNH	Peru: Madre de Dios: Pakitza
P. steerei	-71.1120	-12.0905	USNM559415	morphology	NMNH	Peru: Madre de Dios: Rio Manu
P. steerei	-68.7667	-13.5833	USNM619002	genomic (this study)	NMNH	Bolivia: La Paz: Río Madidi, Moire Camp

Table S4: Samples used in morphological hypervolume analyses for the three sympatric species of *Proechimys* in Western Amazon, with information about the locality, voucher, and the institution of origin of the samples. **AMNH:** American Museum of Natural History, New York, USA; **LMUSP:** Laboratório de Mamíferos da Escola Superior de Agricultura "Luiz de Queiroz, Universidade de São Paulo, Piracicaba, São Paulo, Brazil; **MN:** Museu Nacional da Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil; **MPEG**: Museu Paraense Emílio Goeldi, Belém, Pará, Brazil; **MVZ:** Museum of Vertebrate Zoology, Berkeley, California, USA; **NMNH:** National Museum of Natural History, Washington, D.C., USA. Table with more information about locality and raw variables in ***.csv*** format is available on https://github.com/jdalapicolla.

Catalog Number	Alternative Number	Source	Species	Longitude	Latitude
MN35875		MN	*P. brevicauda*	-73.48	-4.10
MVZ153616		MVZ	*P. brevicauda*	-78.16	-4.46
MVZ155034		MVZ	*P. brevicauda*	-78.17	-4.45
MVZ155036		MVZ	*P. brevicauda*	-78.16	-4.46
MVZ157934		MVZ	*P. brevicauda*	-77.75	-4.02
MVZ157948		MVZ	*P. brevicauda*	-77.75	-4.02
MVZ157966		MVZ	*P. brevicauda*	-77.75	-4.02
AMNH263054		AMNH	*P. brevicauda*	-66.78	-11.75
AMNH263122		AMNH	*P. brevicauda*	-66.78	-11.75
MVZ157855		MVZ	*P. brevicauda*	-69.07	-12.63
MVZ168956		MVZ	*P. brevicauda*	-69.07	-12.60
MVZ168958		MVZ	*P. brevicauda*	-69.07	-12.60
AMNH268281		AMNH	*P. brevicauda*	-73.16	-5.25
AMNH272700		AMNH	*P. brevicauda*	-73.16	-5.25
MPEG28342		MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28343		MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28344		MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28346		MPEG	*P. brevicauda*	-72.85	-8.60
MPEG28347		MPEG	*P. brevicauda*	-72.80	-8.57
MPEG28349		MPEG	*P. brevicauda*	-72.85	-8.60
MPEG28350		MPEG	*P. brevicauda*	-72.78	-8.67
MPEG28351		MPEG	*P. brevicauda*	-72.78	-8.67
MPEG28353		MPEG	*P. brevicauda*	-72.78	-8.67
MPEG28354		MPEG	*P. brevicauda*	-72.78	-8.67

Table S4: Continuation.

Catalog Number	Alternative Number	Source	Species	Longitude	Latitude
MPEG28357		MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28358		MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28360	MVZ194463	MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28361		MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28362		MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28363		MPEG	*P. brevicauda*	-72.82	-8.37
MPEG28460		MPEG	*P. brevicauda*	-72.82	-8.37
MUSM11258		AMNH	*P. brevicauda*	-73.16	-5.25
MUSM11262		AMNH	*P. brevicauda*	-73.16	-5.25
MUSM11297		AMNH	*P. brevicauda*	-73.16	-5.25
MVZ190668		MVZ	*P. brevicauda*	-72.78	-8.67
MVZ190678		MVZ	*P. brevicauda*	-72.82	-8.37
AMNH263060		AMNH	*P. simonsi*	-68.58	-12.40
AMNH263062		AMNH	*P. simonsi*	-68.92	-11.35
AMNH268278		AMNH	*P. simonsi*	-73.16	-5.25
AMNH272699		AMNH	*P. simonsi*	-73.16	-5.25
AMNH272716		AMNH	*P. simonsi*	-73.16	-5.25
AMNH272717		AMNH	*P. simonsi*	-73.16	-5.25
EFA038		LMUSP	*P. simonsi*	-62.27	-4.43
MJ158		LMUSP	*P. simonsi*	-64.83	-9.45
MJ397		LMUSP	*P. simonsi*	-64.83	-9.44
MJ549		LMUSP	*P. simonsi*	-64.86	-9.46
MJ636		LMUSP	*P. simonsi*	-64.85	-9.44
MJ774		LMUSP	*P. simonsi*	-64.83	-9.44
MPEG28397		MPEG	*P. simonsi*	-68.77	-6.47
MPEG28398	MVZ194703	MPEG	*P. simonsi*	-68.89	-6.58
MPEG28400		MPEG	*P. simonsi*	-72.82	-8.37
MPEG28408		MPEG	*P. simonsi*	-72.82	-8.37
MPEG28409	MVZ194602	MPEG	*P. simonsi*	-72.82	-8.37
MPEG28414		MPEG	*P. simonsi*	-66.02	-3.32
MPEG28416		MPEG	*P. simonsi*	-66.02	-3.32
MPEG28417	MVZ194711	MPEG	*P. simonsi*	-66.02	-3.32
MPEG28419		MPEG	*P. simonsi*	-66.02	-3.32
MPEG28424		MPEG	*P. simonsi*	-70.85	-6.75
MPEG28427	MVZ194775	MPEG	*P. simonsi*	-68.77	-6.47
MPEG28429		MPEG	*P. simonsi*	-68.89	-6.58

Table S4: Continuation.

Catalog Number	Alternative Number	Source	Species	Longitude	Latitude
MPEG28430		MPEG	*P. simonsi*	-68.89	-6.58
MPEG28432		MPEG	*P. simonsi*	-68.89	-6.58
MPEG28433		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28437		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28438		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28443		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28444		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28447	MVZ194635	MPEG	*P. simonsi*	-72.78	-8.67
MPEG28448		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28449		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28451		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28452		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28453		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28455		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28456		MPEG	*P. simonsi*	-72.78	-8.67
MPEG28459		MPEG	*P. simonsi*	-72.82	-8.37
MUSM11283		AMNH	*P. simonsi*	-73.16	-5.25
MUSM11299		AMNH	*P. simonsi*	-73.16	-5.25
MUSM11314		AMNH	*P. simonsi*	-73.16	-5.25
MUSM13342		AMNH	*P. simonsi*	-73.16	-5.25
MUSM13343		AMNH	*P. simonsi*	-73.16	-5.25
MUSM13344		AMNH	*P. simonsi*	-73.16	-5.25
MVZ166814		MVZ	*P. simonsi*	-71.26	-12.68
MVZ168955		MVZ	*P. simonsi*	-69.07	-12.60
USNM364151		NMNH	*P. simonsi*	-69.24	-12.60
USNM390367		NMNH	*P. simonsi*	-69.18	-12.59
USNM530935		NMNH	*P. simonsi*	-69.21	-12.72
USNM578000		NMNH	*P. simonsi*	-72.95	-11.58
USNM579259		NMNH	*P. simonsi*	-68.77	-13.58
USNM579616		NMNH	*P. simonsi*	-66.73	-10.77
USNM579694		NMNH	*P. simonsi*	-69.61	-13.15
USNM579695		NMNH	*P. simonsi*	-69.61	-13.15
USNM579697		NMNH	*P. simonsi*	-69.68	-13.50
USNM582772		NMNH	*P. simonsi*	-72.95	-11.58
USNM582897		NMNH	*P. simonsi*	-72.95	-11.58
MPEG28369		MPEG	*P. steerei*	-70.75	-6.83

Table S4: Continuation.

Catalog Number	Alternative Number	Source	Species	Longitude	Latitude
MPEG28486	MVZ194909	MPEG	*P. steerei*	-70.73	-6.80
MPEG28491	MVZ194914	MPEG	*P. steerei*	-70.73	-6.80
MPEG28492		MPEG	*P. steerei*	-70.73	-6.80
MPEG28493		MPEG	*P. steerei*	-70.73	-6.80
MPEG28529		MPEG	*P. steerei*	-70.75	-6.83
MPEG28531		MPEG	*P. steerei*	-70.73	-6.80
MPEG28532		MPEG	*P. steerei*	-70.73	-6.80
MPEG28533		MPEG	*P. steerei*	-70.73	-6.80
MPEG28534		MPEG	*P. steerei*	-70.73	-6.80
MPEG28535		MPEG	*P. steerei*	-70.73	-6.80
MPEG28537		MPEG	*P. steerei*	-70.73	-6.80
MPEG28540	MVZ194987	MPEG	*P. steerei*	-70.75	-6.83
MPEG28543	MVZ194997	MPEG	*P. steerei*	-70.85	-6.75
MPEG28546		MPEG	*P. steerei*	-70.85	-6.75
MPEG28558		MPEG	*P. steerei*	-70.85	-6.75
MPEG28559		MPEG	*P. steerei*	-68.77	-6.47
MPEG28562	MVZ195034	MPEG	*P. steerei*	-68.77	-6.47
MPEG28564	MVZ195036	MPEG	*P. steerei*	-68.77	-6.47
MPEG28567		MPEG	*P. steerei*	-68.92	-6.53
AMNH263056		AMNH	*P. steerei*	-67.20	-11.38
AMNH263057		AMNH	*P. steerei*	-67.22	-11.40
AMNH263108		AMNH	*P. steerei*	-66.22	-11.02
AMNH263109		AMNH	*P. steerei*	-67.20	-11.38
AMNH263112		AMNH	*P. steerei*	-68.58	-12.40
AMNH263113		AMNH	*P. steerei*	-68.58	-12.40
MVZ168942		MVZ	*P. steerei*	-69.07	-12.60
MVZ168943		MVZ	*P. steerei*	-69.07	-12.60
USNM530931		NMNH	*P. steerei*	-71.05	-12.16
USNM530932		NMNH	*P. steerei*	-71.05	-12.16
USNM559415		NMNH	*P. steerei*	-71.11	-12.09
USNM559426		NMNH	*P. steerei*	-71.11	-12.09
MPEG28569		MPEG	*P. steerei*	-72.85	-8.60
MPEG28570		MPEG	*P. steerei*	-72.85	-8.60
MPEG28571		MPEG	*P. steerei*	-72.82	-8.37
MPEG28572	MVZ194874	MPEG	*P. steerei*	-72.82	-8.37
MPEG28573		MPEG	*P. steerei*	-72.82	-8.37

Table S4: Continuation.

Catalog Number	Alternative Number	Source	Species	Longitude	Latitude
MPEG28574		MPEG	*P. steerei*	-72.82	-8.37
MPEG28575	MVZ194879	MPEG	*P. steerei*	-72.82	-8.37
MPEG28576		MPEG	*P. steerei*	-72.82	-8.37

Table S5: Descriptive statistics for the genomic data of the three sympatric species of *Proechimys* from Western Amazonia pre and post the reads processing step, which includes cleaning, alignment, and missing data removal. Comparison of sample size, number of populations, loci, and SNPs between pre and post reads processing for each species. The number of mutations per loci, and length of reads (Pairbase per loci). The maximum value of θ (genetic diversity) allowed by locus, which corresponds to 95% function of probability density (Fig. S3) is also presented.

		P. brevicauda	*P. simonsi*	*P. steerei*
Pre Reads Processing	**Sample size**	17	20	15
	Populations	6	9	7
	Mutations per locus	18	17	19
	Pairbase per locus	140	140	140
	Loci	84,622	74,036	72,036
	SNP	303,573	246,768	220,281
Post Reads Processing	**Sample size**	16	15	14
	Populations	6	7	6
	Mutations per locus	9	8	8
	Pairbase per locus	127	125	129
	Loci	5,050	4,629	5,819
	SNP	21,442	16,998	17,519
θ_{MAX} (Threshold 95%)		0.022	0.020	0.019

Table S6: Initial and filtered number of reads, depth of coverage per sample and per species for all samples of the three sympatric species of *Proechimys* from Western Amazon, followed by their standard deviation (SD) and the maximum number of reads per locus in the sequenced individual.

Species	Catalog Number	Alternative Number	Initial Reads	Filtered Reads	Coverage Depth	SD	Max
P. brevicauda (n = 17)	AMNH272698		1,370,001	1,353,880	17.86	9.63	862
	AMNH272700		1,764,494	1,737,006	20.88	11.21	335
	FMNH175255		442,305	434,908	20.19	13.13	187
	MSB238391		3,222,836	3,154,997	35.33	23.83	779
	MSB70574		3,285,537	3,231,860	36.52	24.75	388
	MSB70575		1,982,773	1,950,711	25.38	15.07	1370
	MUSM13338		708,124	700,457	12.88	7.28	481
	MVZ155121		1,811,744	1,791,793	22.48	13.37	1620
	MVZ157855		3,856,833	3,775,561	44.24	28.60	1175
	MVZ157905		3,388,028	3,351,326	35.67	22.10	1268
	MVZ168953		1,383,971	1,358,279	19.63	12.06	1175
	MVZ194439	INPA3442	2,240,727	2,212,090	24.61	15.74	1468
	MVZ194463	MPEG28360	1,722,621	1,704,610	20.48	11.14	378
	MVZ194485	MPEG28371	1,274,241	1,260,279	16.75	9.13	607
	TTU101179	TK73909	1,063,870	1,031,330	16.09	9.12	642
	TTU101195	TK73940	1,694,713	1,667,744	20.71	11.13	425
	TTU101213	TK73977	2,096,729	2,008,985	23.65	13.78	662
	TOTAL		**1,959,385**	**1,925,048**	**24.31**	**14.77**	**813.06**
P. simonsi (n = 20)	EFA37		1603218	1584577	21.71	10.96	280
	FMNH175275		3611229	3571386	36.73	21.42	491
	MJ529		1618498	1602490	20.31	10.07	399
	MSB236594		1306947	1291507	18.62	8.87	319
	MUSM13339		1165692	1144811	15.84	7.96	647
	MUSM13342		1578524	1553567	18.95	9.58	655
	MVZ157968		220401	203224	8.72	15.13	629
	MVZ166815		2699704	2661738	31.30	17.99	269
	MVZ168955		3705742	3652171	38.34	24.16	560
	MVZ194602	MPEG28409	238885	230238	8.56	8.67	317
	MVZ194635	MPEG28447	412782	406705	14.89	6.94	238
	MVZ194703	MPEG28398	1243929	1225588	15.83	7.25	267
	MVZ194711	MPEG28417	2066850	2041052	23.35	12.07	437

Table S6: Continuation.

Species	Catalog Number	Alternative Number	Initial Reads	Filtered Reads	Coverage Depth	SD	Max
P. simonsi (n = 20)	MVZ194775	MPEG28427	1266460	1246684	16.51	7.65	363
	TTU101118	TK73760	2134726	2103255	25.68	14.28	244
	USNM579697		802295	782670	15.22	8.11	319
	USNM619001	LHE0742	713045	702312	11.84	6.28	353
	USNM619003	LHE0820	2196620	2167069	25.07	14.80	814
	USNM619007	LLW424	1559317	1536613	20.99	10.62	237
	USNM619008	ACF076	2564448	2527600	29.76	16.63	502
	TOTAL		**1,635,466**	**1,611,763**	**20.91**	**11.97**	**417.00**
P. steerei (n = 15)	MSB236689		1063283	1035894	16.71	9.29	475
	MSB236698		3069695	3024333	33.15	19.76	477
	MSB236807		2840277	1548164	20.58	11.53	687
	MVZ168942		1900170	1875324	25.52	14.88	484
	MVZ190951		2456635	2427383	24.39	14.23	472
	MVZ190954		1045238	1034209	16.20	8.05	290
	MVZ194874	MPEG28572	2908472	2871602	26.76	15.84	850
	MVZ194879	MPEG28575	2183799	2156581	24.35	13.97	452
	MVZ194909	MPEG28486	2467462	2429956	26.44	14.45	355
	MVZ194914	MPEG28491	3478150	3432116	33.12	20.10	342
	MVZ194987	MPEG28540	2878306	2835589	27.84	15.68	563
	MVZ194997	MPEG28543	2145539	2121592	25.43	15.13	648
	MVZ195034	MPEG28562	875925	863864	13.96	8.20	529
	MVZ195036	MPEG28564	2890509	2850750	27.59	16.02	423
	USNM619002	LHE0747	646363	632498	11.16	7.43	615
	TOTAL		**2,189,988**	**2,075,990**	**23.55**	**13.64**	**510.80**

Table S7: Pairwise F_{ST} (upper diagonal and bold) and Jost's *D* (lower diagonal) values between populations of the three sympatric species of *Proechimys.*

P. brevicauda	GALVEZ	MADRE	BOLIVIA	AMAZONAS	ACRE	IQUITOS
GALVEZ	-	**0.465**	**0.442**	**0.105**	**0.019**	**0.215**
MADRE	0.250	-	**0.074**	**0.367**	**0.423**	**0.460**
BOLIVIA	0.253	0.031	-	**0.350**	**0.407**	**0.436**
AMAZONAS	0.078	0.242	0.245	-	**0.087**	**0.025**
ACRE	0.009	0.247	0.251	0.073	-	**0.188**
IQUITOS	0.106	0.258	0.261	0.032	0.103	-

P. simonsi	SOLIMOES	CENTRAL ANDES	MADEIRA	GALVEZ	YUNGAS	MADRE	JURUA
SOLIMOES	-	**0.064**	**0.010**	**0.015**	**0.012**	**0.020**	**0.010**
CENTRAL ANDES	0.049	-	**0.035**	**0.054**	**0.030**	**0.005**	**0.046**
MADEIRA	0.019	0.033	-	**0.008**	**0.003**	**0.002**	**0.005**
GALVEZ	0.030	0.045	0.018	-	**0.007**	**0.012**	**0.006**
YUNGAS	0.028	0.029	0.012	0.023	-	**0.004**	**0.005**
MADRE	0.033	0.008	0.013	0.028	0.014	-	**0.011**
JURUA	0.020	0.041	0.011	0.015	0.018	0.025	-

P. steerei	MADRE-BENI	PANDO	SOLIMOES	ACRE	JURUA	JAINU
MADRE-BENI	-	**0.006**	**0.107**	**0.057**	**0.139**	**0.129**
PANDO	0.006	-	**0.100**	**0.049**	**0.132**	**0.120**
SOLIMOES	0.112	0.109	-	**0.054**	**0.022**	**0.013**
ACRE	0.079	0.074	0.075	-	**0.077**	**0.063**
JURUA	0.115	0.113	0.025	0.077	-	**0.013**
JAINU	0.117	0.114	0.024	0.079	0.012	-

Table S8: Pairwise standard Nei G_{ST} (upper diagonal and bold) and Hedrick's G''_{ST} (lower diagonal) values between populations of the three sympatric species of *Proechimys*.

P. brevicauda	GALVEZ	MADRE	BOLIVIA	AMAZONAS	ACRE	IQUITOS
GALVEZ	-	**0.647**	**0.625**	**0.251**	**0.043**	**0.379**
MADRE	0.839	-	**0.204**	**0.546**	**0.612**	**0.645**
BOLIVIA	0.828	0.359	-	**0.534**	**0.594**	**0.624**
AMAZONAS	0.448	0.777	0.771	-	**0.224**	**0.120**
ACRE	0.091	0.819	0.809	0.412	-	**0.348**
IQUITOS	0.598	0.840	0.829	0.239	0.566	-

P. simonsi	SOLIMOES	CENTRAL ANDES	MADEIRA	GALVEZ	YUNGAS	MADRE	JURUA
SOLIMOES	-	**0.156**	**0.056**	**0.083**	**0.078**	**0.098**	**0.058**
CENTRAL ANDES	0.306	-	**0.107**	**0.135**	**0.090**	**0.029**	**0.126**
MADEIRA	0.122	0.220	-	**0.050**	**0.035**	**0.038**	**0.033**
GALVEZ	0.179	0.273	0.111	-	**0.061**	**0.078**	**0.040**
YUNGAS	0.168	0.190	0.079	0.136	-	**0.041**	**0.049**
MADRE	0.206	0.063	0.084	0.169	0.092	-	**0.070**
JURUA	0.128	0.255	0.075	0.090	0.109	0.153	-

P. steerei	MADRE-BENI	PANDO	SOLIMOES	ACRE	JURUA	JAINU
MADRE-BENI	-	**0.021**	**0.267**	**0.177**	**0.283**	**0.298**
PANDO	0.046	-	**0.255**	**0.165**	**0.273**	**0.286**
SOLIMOES	0.486	0.471	-	**0.158**	**0.074**	**0.074**
ACRE	0.356	0.337	0.328	-	**0.172**	**0.179**
JURUA	0.506	0.494	0.160	0.349	-	**0.039**
JAINU	0.523	0.508	0.160	0.358	0.0866	-

Table S9: Population genetics indexes for the populations of the three sympatric species of *Proechimys* from the Western Amazon: sample size (n), nucleotide diversity (π), Wright's inbreeing coefficient (F_{IS}), observed heterozygosity (H_{obs}), percentage of polymorphic sites (%pol).

Species	Populations	n	%poly	H_{obs}	π	F_{IS}
P. brevicauda	ACRE	3	0.155	0.088	0.094	0.011
	AMAZONAS	2	0.169	0.076	0.134	0.086
	BOLIVIA	3	0.105	0.060	0.064	0.007
	GALVEZ	3	0.126	0.069	0.078	0.015
	IQUITOS	3	0.136	0.073	0.082	0.017
	MADRE	2	0.064	0.040	0.049	0.014
P. simonsi	CENTRAL ANDES	3	0.174	0.073	0.103	0.056
	GALVEZ	2	0.203	0.142	0.152	0.016
	JURUA	2	0.194	0.130	0.147	0.026
	MADEIRA	2	0.187	0.122	0.141	0.030
	MADRE	2	0.172	0.113	0.131	0.027
	SOLIMOES	2	0.173	0.091	0.134	0.065
	YUNGAS	2	0.198	0.122	0.151	0.044
P. steerei	ACRE	2	0.248	0.173	0.187	0.021
	JAINU	2	0.157	0.109	0.120	0.016
	JURUA	4	0.248	0.124	0.131	0.014
	MADRE-BENI	2	0.171	0.114	0.130	0.023
	PANDO	2	0.160	0.106	0.122	0.024
	SOLIMOES	2	0.191	0.135	0.145	0.014

Table S10: Procrustes results for each species with the index of association between geography and genetics (t_0) and its p-value. The robustness of the Procrustes eliminating a population at time, that generated Figure 4 are also indicated. Each line corresponds to a population eliminated by species. **t"** = value of the association between geographic and genetic maps without the population; **t_0 - t''** = increase or decrease in association value between geographic and genetic maps when population was elimitated; **$\theta_{t''}$** = angle rotation between the genetic and geographic map in the new analysis with the population and its p-value. **t'** = association value between the original PCA and the new PCA without the population, values close to 1 indicate that withdraw population had little effect on the PCA structure, in other words, few differences were obtained when the population was eliminated; **$\theta_{t'}$** = angle rotation between original PCA and the new one, without the population.

	Populations	**t''**	**t_0 - t''**	**$\theta_{t''}$**	**p-value**	**t'**	**$\theta_{t'}$**
P. brevicauda $t_0 = 0.730$ p-value < 0.001	GALVEZ	0.799	0.069	43.239	<0.001	0.882	6.463
	MADRE	0.773	0.043	43.464	0.002	0.875	-0.329
	BOLIVIA	0.677	-0.052	34.850	0.007	0.842	-0.220
	AMAZONAS	0.811	0.082	49.005	<0.001	0.999	-3.272
	ACRE	0.788	0.059	-56.826	0.001	0.881	7.768
	IQUITOS	0.824	0.094	-63.405	<0.001	0.908	-6.397
P. simonsi $t_0 = 0.635$ p-value < 0.001	SOLIMOES	0.652	0.017	-85.196	0.006	0.826	-1.861
	CENTRAL ANDES	0.795	0.159	-58.970	<0.001	0.777	-27.626
	MADEIRA	0.767	0.131	-81.659	0.001	0.833	2.574
	GALVEZ	0.597	-0.038	-70.216	0.015	0.990	3.711
	YUNGAS	0.868	0.232	-77.835	<0.001	0.833	2.245
	MADRE	0.785	0.150	-76.684	<0.001	0.820	-9.007
	JURUA	0.758	0.122	-77.150	0.001	0.988	-3.730
P. steerei $t_0 = 0.718$ p-value < 0.001	MADRE-BENI	0.787	0.069	-68.952	0.001	0.837	16.073
	PANDO	0.806	0.088	-76.359	<0.001	0.837	18.462
	SOLIMOES	0.897	0.179	-82.648	<0.001	0.862	-3.601
	ACRE	0.945	0.228	-70.767	<0.001	0.840	-8.579
	JURUA	0.930	0.212	-89.884	<0.001	0.847	-9.609
	JAINU	0.828	0.110	-86.654	<0.001	0.863	-3.879

Table S11: Parameters used in the ecological niche models (ENM), their validation indexes, and threshold value used for the final construction of the models per each sympatric species. Initial and filtered occurrence points number used to create the models, the uncorrelated variables (Bio1-19) selected for each species is presented. Feature class and regularization multiplier values utilized in the models complexity, and number of samples for the training and test step, AUC values (AUC_{TRAIN} and AUC_{TEST}) with its standard deviation (AUC_{SD}), True Skill Statistic (TSS) and omission error for each model is also showed.

	P. brevicauda	***P. simonsi***	***P. steerei***
Initial Occurrence Points	182	166	120
Filtered Occurrence Points	44	48	46
Variables from WorldClim (Bio)	2, 4, 8, 13, 18	2, 4, 8, 16, 18	2, 4, 8, 13, 18
Feature Class	H	LQH	LQP
Regularization Multiplier	3.0	2.0	3.0
Training Samples	31	34	33
Test Samples	13	14	13
AUC_{TRAIN}	0.799	0.749	0.755
AUC_{TEST}	0.762	0.711	0.731
AUC_{SD}	0.029	0.05	0.029
True Skill Statistic (TSS)	0.342	0.195	0.338
Omission Error	0.198	0.133	0.154
Maximum Value of Suitability	0.652	0.657	0.719
Threshold 10%	0.420	0.266	0.391

Table S12: Contribution and importance of the environmental variables for the construction of the ecological niche models (ENM) for each sympatric species of *Proechimys* from Western Amazon.

Species	Variables	Contribution (%)	Permutation importance (%)
P. brevicauda	Bio2	44.4	55.8
	Bio8	27.1	10.3
	Bio18	13.1	19.0
	Bio13	9.1	3.0
	Bio4	6.3	11.8
P. simonsi	Bio2	64.8	51.2
	Bio18	17.3	22.5
	Bio8	8.8	13.1
	Bio4	5.5	7.6
	Bio16	3.6	5.6
P. steerei	Bio8	59.0	49.7
	Bio2	38.6	45.4
	Bio13	1.4	0.9
	Bio18	0.8	3.4
	Bio4	0.2	0.6

Table S13: Suitable area (in Km²) for the occurrence of the three sympatric species of *Proechimys* in the ecological niche models (ENM) over time, and the value of the suitable area that remains stable in all time periods.

Species	Present-day	Middle Holocene	LGM	LIG	Stable Area
P. brevicauda	1,413,575.52	972,096.11	1,122,454.24	1,182,064.78	937,955.37
P. simonsi	1,832,931.43	1,940,964.14	1,901,722.75	1,863,146.61	1,792,602.62
P. steerei	1,238,353.55	1,225,968.34	1,205,705.06	1,262,620.87	1,161,620.63

Table S14: Distance-based RDA results for the three sympatric species of *Proechimys* from Western Amazon. Two first lines for each species indicate the marginal tests, showing the results for association between genetic (G) and geographical distance(D) and between genetic distance and PC1 values (E) with values of F-statistic and p-values. Third line is the result of conditional tests, which shows the association of the genetic distance (G) with the PC1 values (E), being conditioned to the effect of the geographic distance (D), in order words, removing the D effect. NA values in conditional tests represents that "No residual component" returned in the analyses, and it was not possible calculate F-statistic and p-value. It may be interpreted as the geographic distance (D) as a "condition" practically does not affect in the model, and the geographic distance (D) is independent from genetic (G) and from PC1 values (E).

Species	dbRDA	F	P-value
P. brevicauda	**G~D**	21.690	0.072
	G~E	5.780	0.071
	G~E|D	NA	NA
P. simonsi	**G~D**	1.517	0.520
	G~E	3.406	0.108
	G~E|D	NA	NA
P. steerei	**G~D**	12.990	0.060
	G~E	8.142	0.056
	G~E|D	0.385	0.754

Table S15: Variables contribution to the five principal components (PC) used for the estimation of the environmental hypervolume, with indication of the variation amount explained by each PC. The three variables that contribute most to each PC, negatively or positively, are highlighted in bold.

Variables	PC1 (44.52%)	PC2 (28.69%)	PC3 (12.65%)	PC4 (6.29%)	PC5 (3.79%)
Bio1	0.039	**0.413**	-0.045	-0.164	-0.125
Bio2	-0.287	-0.136	0.111	-0.034	-0.170
Bio3	0.292	-0.028	0.280	0.067	-0.175
Bio4	-0.214	-0.094	**-0.329**	**-0.367**	**0.412**
Bio5	-0.231	0.268	-0.162	-0.110	0.007
Bio6	0.268	0.264	0.000	0.054	0.008
Bio7	**-0.327**	-0.082	-0.079	-0.097	-0.003
Bio8	-0.120	0.284	0.010	**-0.501**	-0.336
Bio9	0.184	0.339	-0.069	0.127	0.098
Bio10	-0.043	**0.381**	-0.208	-0.249	0.072
Bio11	0.139	**0.362**	0.122	0.092	-0.301
Bio12	**0.298**	-0.125	-0.237	-0.049	-0.090
Bio13	0.136	-0.119	**-0.531**	0.077	-0.291
Bio14	0.294	-0.123	0.078	-0.322	0.098
Bio15	-0.296	0.050	-0.186	0.301	-0.180
Bio16	0.152	-0.123	**-0.514**	0.131	-0.283
Bio17	**0.306**	-0.108	0.037	-0.298	0.127
Bio18	0.091	-0.319	0.039	**-0.399**	**-0.400**
Bio19	0.277	0.074	-0.240	0.055	**0.380**

Table S16: Variables contribution to the nine principal components (PC) used for the estimation of the morphological hypervolume, with indication of the variation amount explained by each PC. The three variables that contribute most to each PC, negatively or positively, are highlighted in bold.

Variables	PC1 (73.69%)	PC2 (5.32%)	PC3 (4.02%)	PC4 (2.95%)	PC5 (2.41%)	PC6 (2.14%)	PC7 (1.90%)	PC8 (1.51%)	PC9 (1.12%)
BaL	**-0.978**	-0.019	-0.001	0.087	0.010	-0.045	-0.047	-0.044	-0.044
BuL	-0.714	0.213	0.156	**-0.314**	**0.342**	**0.296**	**-0.313**	-0.014	-0.031
CDM1	-0.967	-0.064	0.001	-0.088	0.010	-0.025	0.005	0.076	0.091
CIL	**-0.984**	-0.048	-0.024	0.058	0.043	-0.026	-0.027	-0.015	-0.029
D	-0.940	-0.121	0.034	0.185	0.042	-0.042	-0.019	-0.109	-0.081
GSL	**-0.970**	-0.167	-0.038	0.080	0.004	-0.013	-0.015	0.041	0.010
IFL	-0.648	**0.521**	0.168	0.208	0.282	-0.149	0.284	-0.063	**0.166**
IFW	-0.517	0.227	**0.712**	0.207	**-0.291**	0.166	-0.043	0.092	-0.054
IOC	-0.809	-0.133	-0.038	-0.220	0.109	0.178	**0.331**	**0.266**	-0.023
LMD	-0.925	-0.056	-0.027	0.145	0.126	0.010	-0.025	-0.067	-0.164
MaxB	-0.823	0.091	0.054	**-0.264**	-0.182	**-0.310**	-0.157	-0.132	0.090
MB	-0.891	0.091	-0.015	-0.051	0.008	0.164	0.023	-0.062	**0.220**
MD	-0.953	0.061	-0.114	-0.003	0.083	-0.052	-0.038	-0.004	-0.061
MTRL	-0.528	**0.678**	**-0.323**	-0.105	-0.226	-0.078	-0.037	**0.242**	-0.121
NL	-0.898	**-0.273**	-0.043	0.169	-0.037	-0.037	-0.053	0.161	0.052
OccW	-0.725	0.097	**-0.371**	0.146	**-0.282**	**0.382**	0.070	**-0.219**	0.055
OL	-0.924	-0.130	-0.060	0.006	0.012	0.013	-0.036	0.017	-0.038

Table S16: Continuation.

Variables	PC1 (73.69%)	PC2 (5.32%)	PC3 (4.02%)	PC4 (2.95%)	PC5 (2.41%)	PC6 (2.14%)	PC7 (1.90%)	PC8 (1.51%)	PC9 (1.12%)
PLa	-0.904	0.203	-0.110	0.155	0.108	-0.080	-0.013	-0.073	-0.135
RB	-0.769	-0.111	0.201	**-0.381**	-0.128	-0.045	**0.287**	-0.208	**-0.181**
RD	-0.922	-0.169	0.063	-0.066	-0.119	-0.047	-0.031	0.091	0.111
RL	-0.924	-0.263	-0.008	0.114	-0.008	-0.060	-0.025	0.111	-0.024
ZB	-0.930	-0.021	0.000	-0.123	-0.063	-0.061	-0.048	-0.028	0.157

Figures

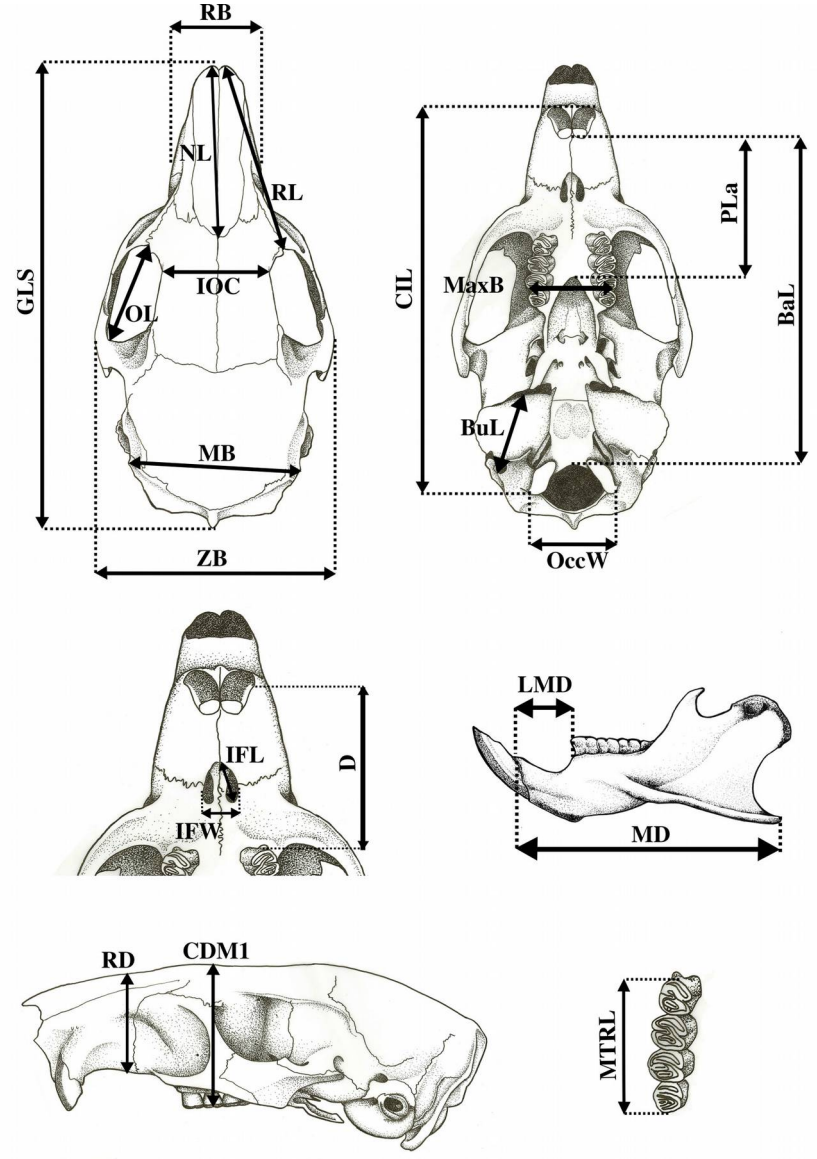


Figure S1: 22 cranial measurements with their respective acronyms (illustration of a *Trinomys dimidiatus* made by Gustavo S. Libardi). **BaL:** basilar length of Hensel; **BuL:** bullar length; **CDM1:** cranial depth at M1; **CIL:** condyloincisive length; **D:** diastema length; **GSL:** greatest length of skull; **IFL:** length of incisive foramen; **IFW**: maximum width of incisive foramen; **IOC**: interorbital constriction; **LMD:** length of mandibular diastema; **MaxB:** maxillary breadth at M2-M3; **MB:** greatest breadth across mastoid; **MD:** mandibular length; **MTRL**: alveolar length of upper molariforms; **NL:** nasal length; **OccW**: occipital condyle width; **OL:** orbit length; **PLa:** palatal length A; **RB:** rostral breadth; **RD:** rostral depth; **RL:** rostral length; **ZB:** zygomatic arch breadth.

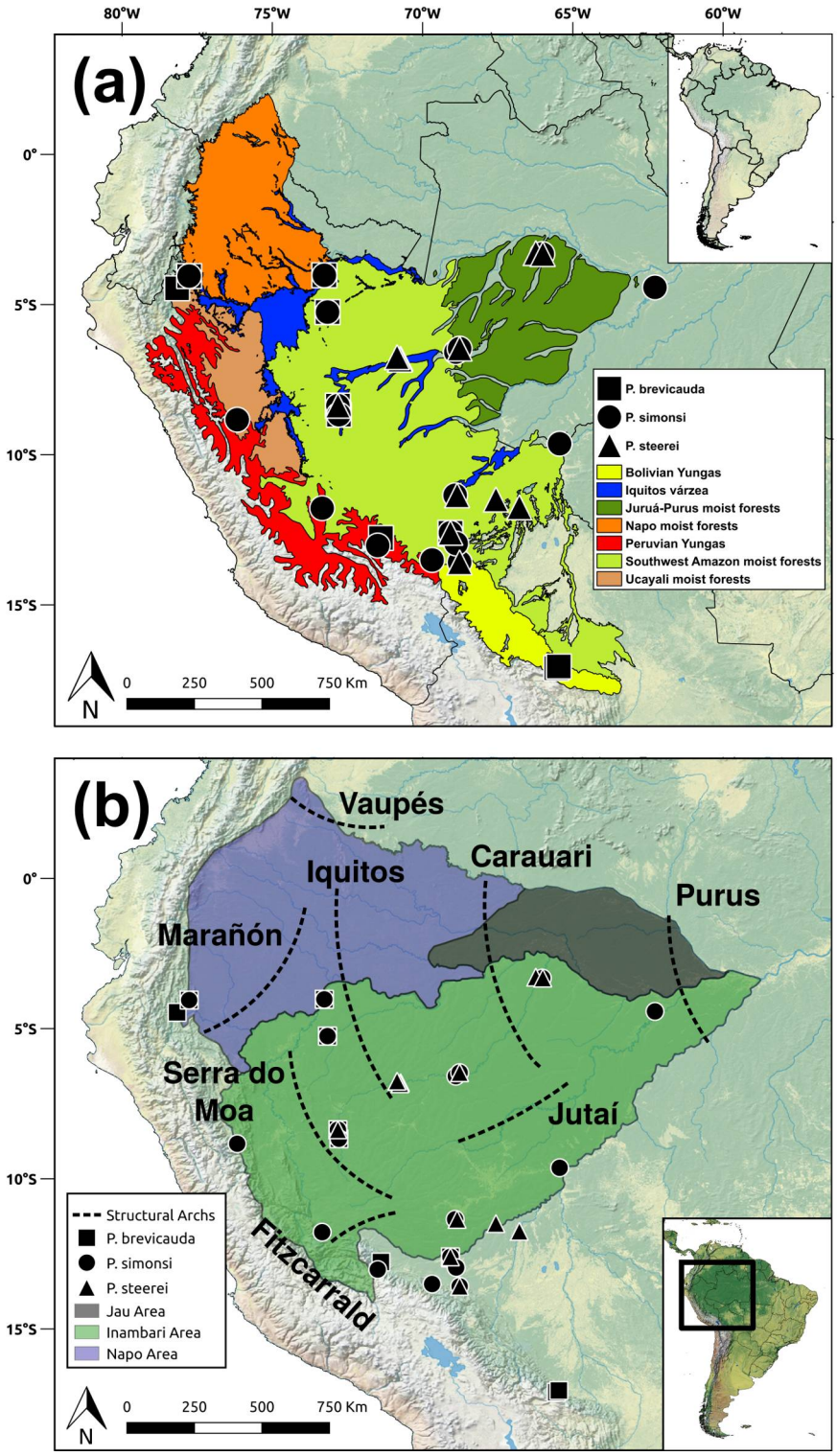


Figure S2: Ecoregions (a), areas of endemism and structural archs (b) used as strata in the analyses of molecular variance (AMOVA). Squares, circles, and triangles represent the genetic samples to *P. brevicauda*, *P. simonsi*, and *P. steerei*. Samples to the south of Inambari area were grouped into "South Inambari" level.

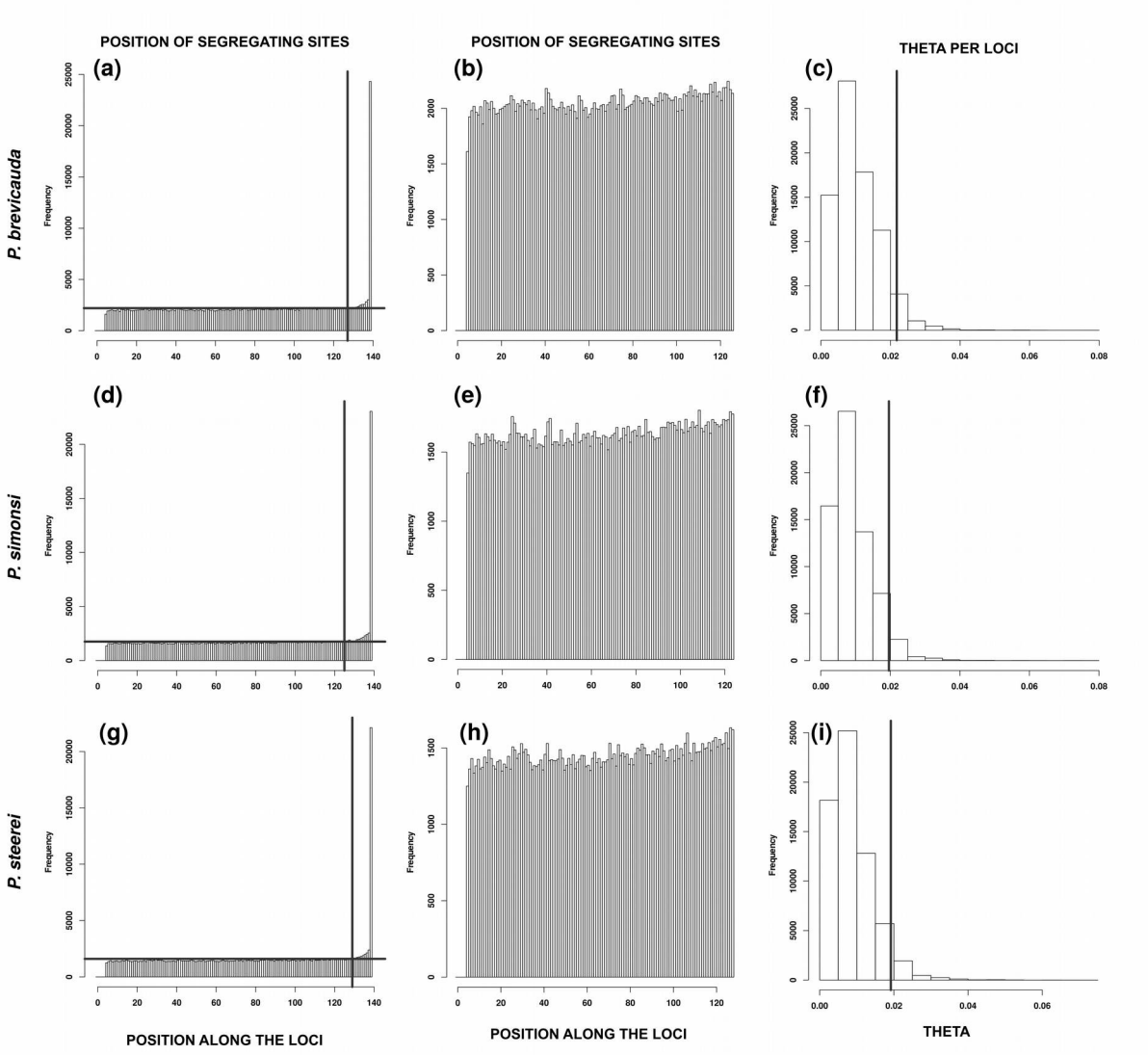
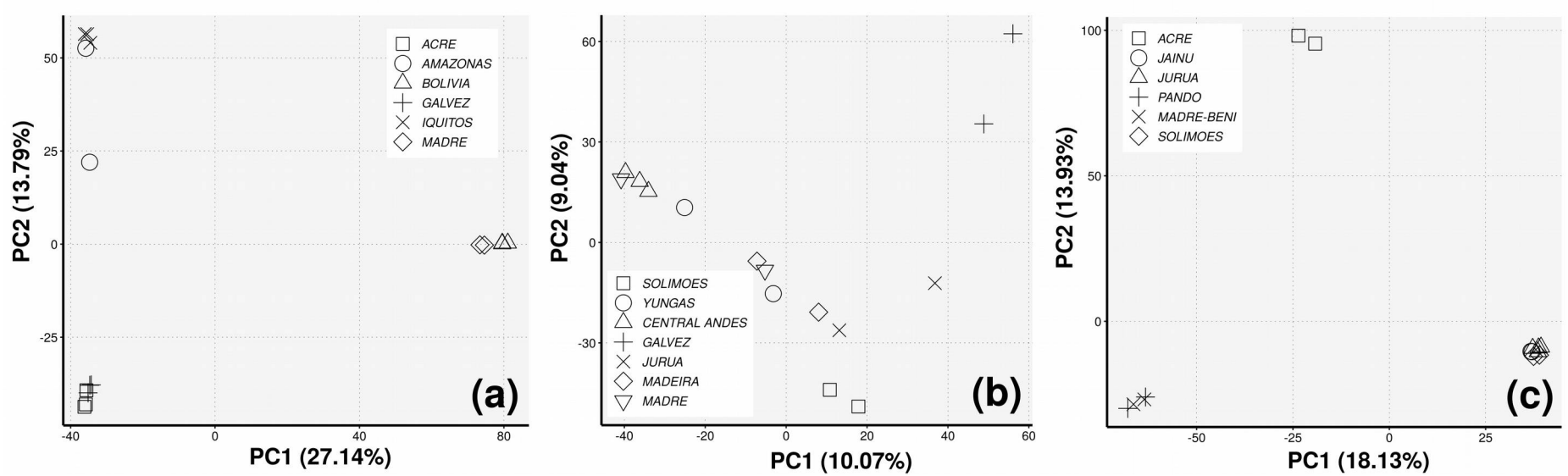


Figure S3: Segregating sites frequency by position within the loci for the three sympatric species of *Proechimys: P. brevicauda* (a-c)*, P. simonsi* (d-f)*,* and *P. steerei* (g-i). (a), (d), and (g) show the increase of polymorphism frequency at the final positions of the loci, after vertical and longitudinal gray lines. (b), (e), and (h) show the polymorphism frequency at positions at the along of the loci after removal of the more polymorphic positions. (c), (f), and (i) show the loci frequency with theta values greater than the 95% threshold that were removed in the reads processing step.

Figure S4: Principal Components Analysis (PCA) graphs using PC1 and PC2 for the three sympatric species of *Proechimys* from Western Amazon: *P. brevicauda* (a), *P. simonsi* (b) and *P. steerei* (c). Symbols represent the different populations from which species.



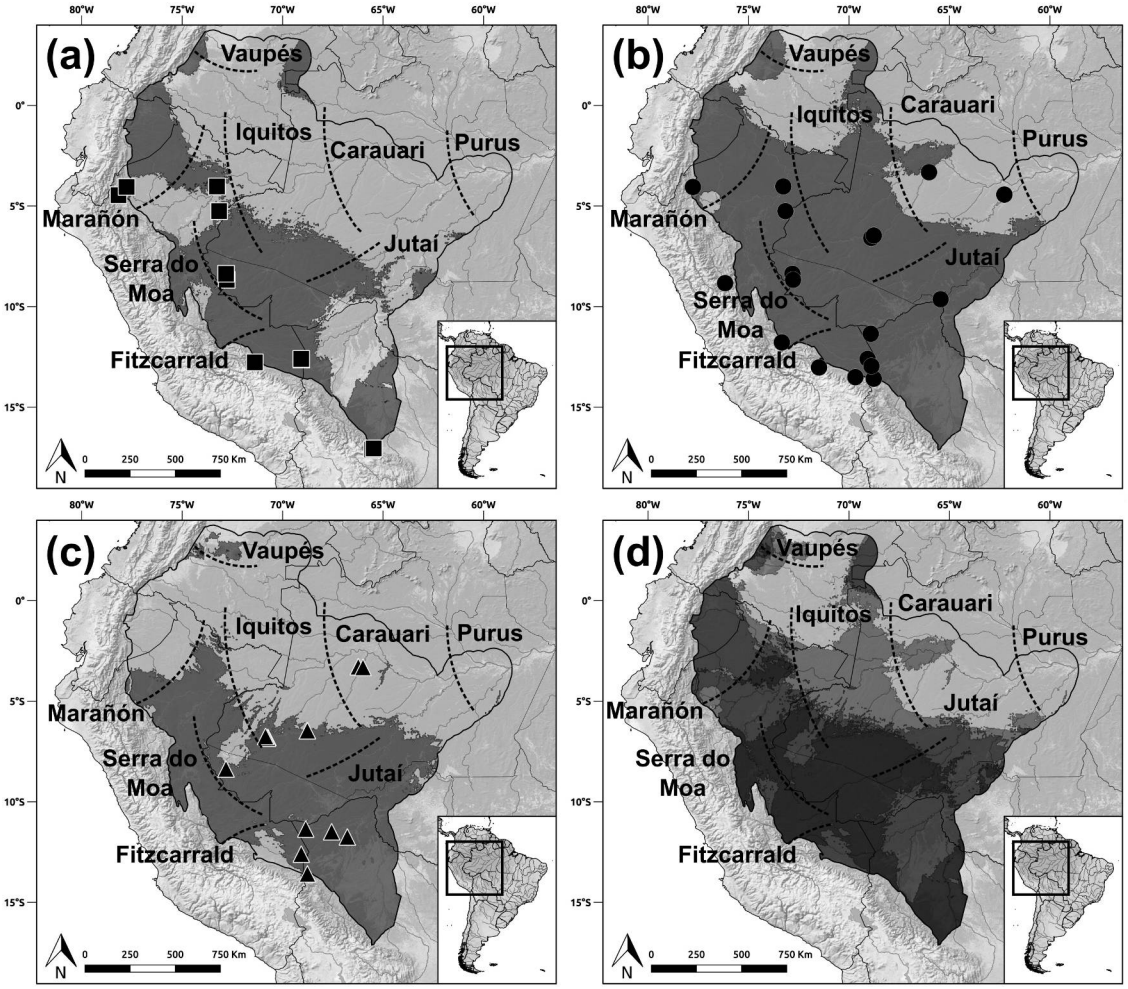


Figure S5: Areas of stability for the geographic distribution of the three sympatric species of *Proechimys* from the Western Amazon over time: (a) *P. brevicauda*, (b) *P. simonsi*, (c) *P. steerei,* (d) for the three species together. Stable areas were constructed by the intersections of the Ecological niche models (ENM) of each species from the Last Interglacial (LIG) to the present-day models. Squares, circles, and triangles represent the genetic samples to *P. brevicauda*, *P. simonsi*, and P. steerei, respectively. Dotted lines indicate the structural archs locations with their respective names. The study area is within the thicker black line. Relief and major rivers are represented on the map to view the geographic landmarks.

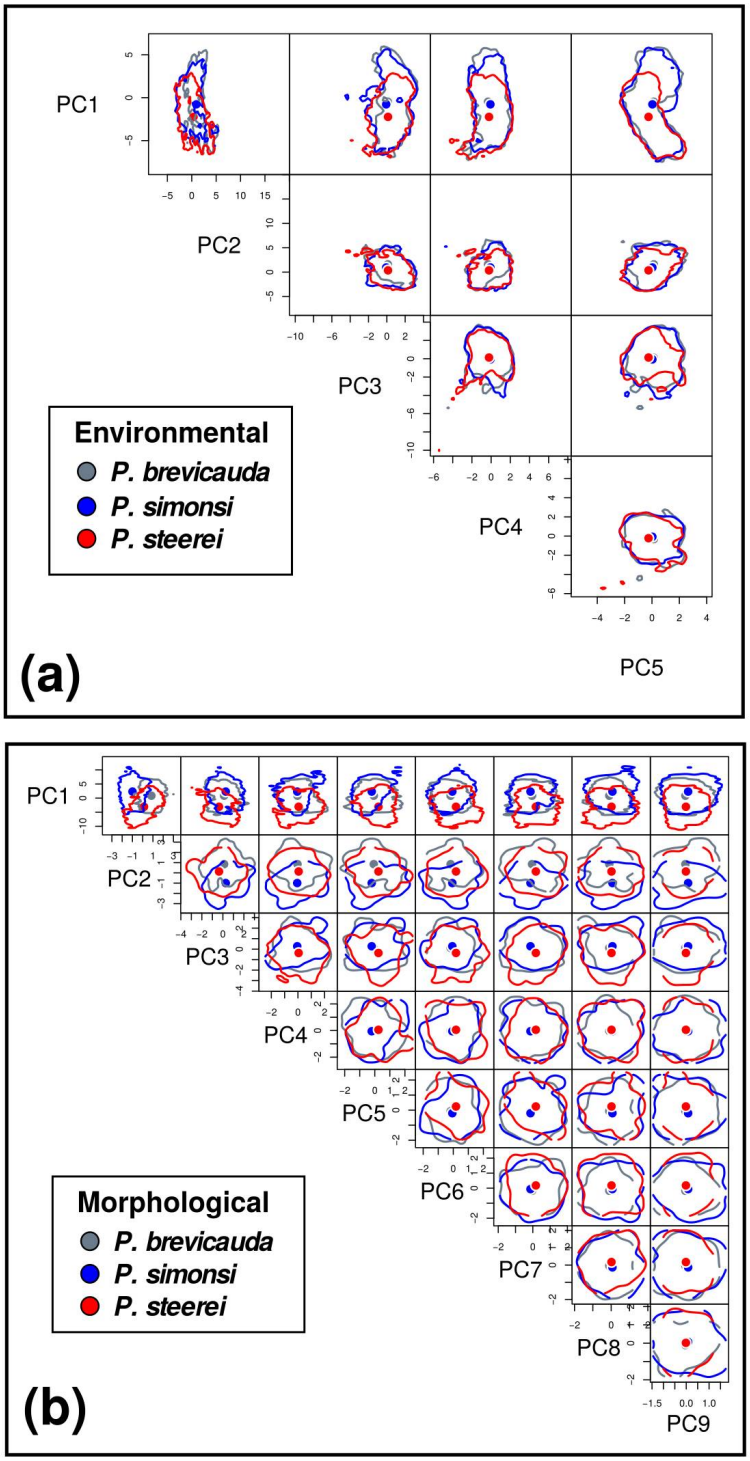


Figure S6: Pairwise comparisons by principal components (PC) of the overlap of environmental (a) and morphological (b) hypervolumes among the three sympatric species of *Proechimys* from Western Amazon. The colors represent each species and the number of PC used correspond to 95% of the variation in Principal Components Analysis (PCA) for each hypervolume (Table S15 and S16).

5. SYNTHESIS, RECOMMENDATIONS AND FUTURE PERSPECTIVES

During my Ph.D. thesis I applied interdisciplinary techniques and concepts, the Integrative Taxonomy approach, and I was able to unify results obtained from distributional, cytogenetic, mtDNA, nuDNA, genomic, morphologic/morphometric, and climatic/environmental data. I also used modern modelling and simulation methods, and multivariate analyses to integrate these data to test hypotheses on the ecology and evolution of an Amazonian group of rodents still poorly known to the science.

This thesis was the most comprehensive study for an echimyid rodent genus, either in the number of samples, the geographical scale, or different types of data used. I was able to quantify inter- and intra-genomic variation in the genus *Proechimys*, obtain a phylogenetic resolution for the basal portion of the tree recovering five main clades (A-E), delimit the putative species, and present a biogeographic history that contributed on understanding the evolution of the Amazonian landscape and the effect of past climate changes.

5.1. Taxonomic Implications

I integrated genomic, morphometric, cytogenetic, mtDNA and nuDNA data to recover the phylogenetic relationships and to delimit the species of the genus *Proechimys*. This broad approach has brought new information on the relationships of the genus *Hoplomys*, on the position of *Proechimys hoplomyoides*, and on the species diversity of *Proechimys* with the identification of 28 lineages with potential to be considered species, a higher number than the 22 valid species in the genus currently.

In this perspective, the main taxonomic implications of this study were to provide data for the future taxonomic rearrangements, and the indication of an increase in species/taxa diversity for the genus *Proechimys*, as well as a change in the arrangement of genera within the family Echimyidae. I could not associate an available nominal taxon to 12 of the 28 lineages, some of these may represent putative new species, while others may result in revalidations of taxa currently included in the synonymy of valid species.

5.2. Biogeographic and Ecological Implications

Employing genetic data coupled to fossil record calibration points, I estimated the divergence times for the genus *Proechimys* and its species and tested biogeographic models explain its current distribution. The most recent common ancestor for *Proechimys* was estimated to have originated about 11 Ma in the Miocene, being its ancestral area the Western Amazon. Biogeographic history of the genus in the Western Amazon was greatly influenced by the geological change promoted, among other events, by the Andes uplift in the Miocene, while for the taxa from Eastern Amazon the climatic oscillations of the Pleistocene were more important than geological changes. In addition, the biogeographic history of *Proechimys* is an example of how geology and climatic variation can affect populations promoting diversification.

Geographic distance and climatic differences among populations were not important to explain the genetic variation in three sympatric species of *Proechimys* from Western Amazon. However, geological barriers of the region showed great correlation with the genetic variation. Most of these barriers, such as the main Amazonian rivers, probably delimit areas of secondary contact between lineages rather than have played the role drivers for diversification in the genus. In addition, the origin of *Proechimys* and the ancestors of the five main clades (A-E) also presented similar ages for their origins, evidencing another example of a rapid diversification event in the family Echimyidae during the Miocene.

Hypervolumes results indicated little similarity between the n-dimensional spaces occupied by the sympatric species using morphometric data, while in the climatic hypervolume the similarity was greater. These results imply that sympatric species may be segregated at the microhabitat level but their segregation is more related to how these species use the habitat (morphological traits) than to climatic / environmental differences between localities or habitat, as non-flooded or seasonal flooded forests. These results may imply that the sympatric species of *Proechimys* from the Western Amazon are potential models for ecological studies on competition, partition resources in the environment, among others.

5.3. Conservation Implications

The sympatric species used as a model for the phylogeographic study are classified as least concern (LC) in the IUCN (https://www.iucnredlist.org/species/) but there is no information on the threats to the populations, the population size, distribution and ecological traits for these species. My data on the genetic diversity and inbreeding suggested that the populations do not show signs of diversity loss, confirming their classification as LC. The most important variables for ecological niche models (ENM) were those related to the temperature. This implies that in the context of climatic changes, increases or decreases in mean temperature would represent a threat to the species of *Proechimys*, apart from the habitat loss that is currently increasing in Amazon, especially on its southern border on the arc of deforestation on Mato Grosso and Rondônia.

The results presented here and the future taxonomic rearrangements based on them also would have implications on the study of the fauna in the Amazon, either in ecology and ethology, or on environmental impact studies and infrastructure projects in the region, since a better understanding of the genetic and species diversity may produce more efficient decision-making about environmental management.

5.4. Methodological Implications

This thesis also presented some methodological innovations, as an application of RADSeq data to study taxa with older divergence times from Neotropics, such as *Proechimys*. This cheaper and faster technique can be useful to taxonomic and evolutionary studies, especially if it was used with other tools, such as, simulations, or other datasets for integrative analyses.

Also merit emphasis the application of traditional morphometric data in species delimitation, considering that most methods use genetic data. These data have already been considered to be less informative for species delimitation in *Proechimys*. The present study showed that morphometric data can still be useful for the species delimitation, if properly treated, such as removing the effect of allometry and used in integration with other datasets.

Morphometric and distribution data as well all R scripts and tutorials for statistical analyses are available on the internet through my personal GitHub page (https://github.com/jdalapicolla). Only genetic data due to their size have not yet been available, and it should be after the chapters publication. It presents important implications for future studies on biodiversity, from other researchers, who wish perform similar analyses, and now they hold a start-point with background material. These initiatives make science more inclusive by allowing free access to data, protocols, and extra information that, in this case, was generated from public resources of Brazilian government.

5.5. Futures Perspectives

During the Ph.D. project, I identified morphotypes and geo-referenced localities of 3,104 specimens of *Proechimys* from 18 museums and collections in Brazil, United States and England, and evaluated the morphological variation of 22 quantitative characters in 1,503 specimens, and 58 qualitative characters of skull and skin in 315 individuals. In this period, I photographed 477 specimens, and I have images, and morphological and morphometric data of 62 of the 65 type-series of the genus *Proechimys*.

My perspective is to continue working on these data and to propose a taxonomic revision at the specific level and gather data about the three missing type-specimens to correlate morphologic and genetic variation to the name bearing taxa, and apply the appropriate names to the species that I recovered on my integrative approach. Therefore, I intend to establish formal species description, evaluation of morphological variation, and provide an identification key for the species of *Proechimys*. In addition, the revision will promote corrections of taxonomic errors and a better understanding of the taxonomic history of the group. Moreover, the emergency actions are to implement a broader phylogenetic assessment for *Hoplomys* and *Proechimys*, with the inclusion of additional samples of *Hoplomys* and *P. hoplomyoides*, to verify which taxonomic rearrangements would be most appropriate at generic level.