INSTITUTO DE BIOCIÊNCIAS

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA ANIMAL

OMAR MACHADO ENTIAUSPE NETO

REVISÃO INTEGRATIVA DO COMPLEXO DE ESPÉCIES *PHALOTRIS* AFF. *LEMNISCATUS* (SERPENTES: DIPSADIDAE)

PORTO ALEGRE

2024

OMAR MACHADO ENTIAUSPE NETO

REVISÃO INTEGRATIVA DO COMPLEXO DE ESPÉCIES *PHALOTRIS* AFF. *LEMNISCATUS* (SERPENTES: DIPSADIDAE)

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Animal, Instituto de Biociências da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do título de Mestre em Biologia Animal.

Área de Concentração: Biologia Comparada
Linha de Pesquisa: Sistemática e Biogeografia
Orientador: Prof. Dr. Márcio Borges-Martins

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

PORTO ALEGRE

2024

OMAR MACHADO ENTIAUSPE NETO

REVISÃO INTEGRATIVA DO COMPLEXO DE ESPÉCIES *PHALOTRIS* AFF. *LEMNISCATUS* (SERPENTES: DIPSADIDAE)

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Animal, Instituto de Biociências da Universidade Federal do Rio Grande do Sul, como requisito parcial à obtenção do título de Mestre em Biologia Animal.

Aprovada em _____ de _____ de _____.

_____

Dr. Daniel Loebmann (FURG)

_____

Dr. Nelson Jurandi Rosa

Fagundes (UFRGS)

_____

Dr. Ismael Franz (UFRGS)

SUMÁRIO

“I have found snake systematics to be a hard test to intellectual honesty.”

Garth Underwood

AGRADECIMENTOS

Ao meu orientador, Márcio Borges-Martins, pela paciência, ensinamentos e tempo que foram dedicados à minha formação. Serei sempre grato por todas reuniões e discussões que tivemos, tentando resolver este e outros quebra-cabeças com peças incompletas na taxonomia e sistemática de Serpentes. Sobretudo, pelo acolhimento em um período tão difícil durante o início de minha dissertação.

Ao Felipe Gobbi Grazziotin, pela recepção em seu laboratório, orientação, e todo tempo despendido em reuniões e aulas semanais, indispensáveis para meu amadurecimento profissional, bem como ensinamentos pessoais que levarei pro resto da minha vida.

Ao Instituto de Ciências Biológicas da Universidade Federal do Rio Grande do Sul e ao Instituto Butantan, indispensáveis para o desenvolvimento do meu projeto.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, pela bolsa de pesquisa, sem a qual este estudo não seria exequível.

Aos meus amigos e colegas do Laboratório de Herpetologia da Universidade Federal do Rio Grande do Sul, pela convivência e parceria: Samuel “Samuka” Gohlke, Diogo Reis, Diego Janisch Alvares, Dener Heiermann, Ísis Homrich, Marcelo Freire, Gustavo Cubas, Gabriele Volkmer e Natália Dallagnol Vargas.

Aos meus amigos e colegas do Laboratório de Coleções Zoológicas do Instituto Butantan, pela recepção e acolhimento que recebi ao chegar em São Paulo: Arthur Abegg, Lucas Neves, Weverton Azevedo, Julia Mayumi, Frederico Alcântara, Bruno Rocha, Larissa Silva, Guilherme Oliveira, Gabriele Oliveira e Jayme Massim.

Ao Valdir José Germano, pelo tempo e paciência ao me ensinar protocolos de análise e preparo de material colecionado, assim como as discussões e ensinamentos sobre taxonomia de Serpentes.

À Juliana Wingert e Lívia Corrêa, pela paciência em me ensinar técnicas de laboratório em Biologia Molecular, na Universidade Federal do Rio Grande do Sul e Instituto Butantan, respectivamente.

Ao Jayme Massim, pela imensa ajuda com dados de tomografia computadorizada e discussão de caracteres de osteologia craniana.

À Claudia Koch, Thaís Barreto Guedes e Arthur Tiutenko, pela colaboração de longa data para as discussões de taxonomia e sistemática em Elapomorphini.

À minha companheira, Bruna Beraldo, por todo apoio, carinho e paciência comigo durante essa etapa.

Aos meus amigos, Grégori Rodrigues, Vinícius Ott, Brenda Henz, Gabriele Schiavon, Sylvio Iago, Jackson Rodrigues, Lucas Collares, Heliks Mesenburg, por todo apoio durante a minha jornada no mestrado.

À minha família: Maria José, Paulo Roberto Cardoso Gouvêa, Daniela Entiauspe, Carlos Francisco Mauch, Ludmila Entiauspe, Martina Ribeiro, Nathália Vellinho, Diego Ribeiro, Rubens Vellinho, por todo apoio desde a minha infância, culminando em minha carreira enquanto pesquisador.

Aos curadores de coleções científicas que permitiram meu acesso ao material sob seus cuidados, e outros que colaboraram com dados ou fotos de espécimes: Antônio Argôlo (UESC), Alejandro Giraudo (INALI), Ana Lúcia Prudente (MPEG), Arnold Kluge, Dan Rabosky (UMMZ), Carl Franklin, Eric Smith, Jonathan Campbell (UTA), Diva Maria Borges-Nojosa

(CHUFC), Daniel Klingberg Johansson, Paul Møller (ZMUC), Diego Meneghelli (UFRO-H), Diego Santana (ZUFMS), Eliza Maria Xavier Freire (MUFAL), Felipe Curcio, Christine Strüssmann (UFMT-R), Felipe Gobbi Grazziotin, Giuseppe Puorto (IBSP), Gabriel Skuk (in memorian, MUFAL), Gisele Cotta (FUNED), Günther Köhler (SMF), Hussam Zaher (MZUSP), Lucindo Gonzales, Steffen Reichle (MNKR), Laura Vohnnahme, Daniel Kizirian, Leroy Nuñez (AMNH), Fernanda Werneck (INPAH), Vicente Mata (FMNH), Jonathan Martínez (MCZ), Júlio César Moura-Leite (MHNCI), Martha Calderón, Guido Medina Fabián-Rangel (ICN), Mário Moura (MZUFV), Mark Oliver Röddel (ZMB), Ned Gilmore (ANSP), Nicolás Vidal (MNHN), Pier Cacciali (MHNP), Paulo Manzani (ZUEC), Paulo Passos (MNRJ), Patrick Campbell (BMNH), Rejane Lira-da-Silva (MZUFBA), Rafe Brown (KU), Víctor Zaracho (UNNEC), e Silke Schweiger (NMW).

Aos pesquisadores argentinos Alejandro Giraudo, Gustavo Scrocchi, Víctor Zaracho, pela ótima recepção, saídas de campo e disponibilidade em ajudar este projeto.

Às equipes esportivas “Hype” e “Beco”: Grégori Rodrigues, Lucas Collares, Sylvio Iago, Jackson Rodrigues, e Lucas Neves, Larissa Silva, Jayme Massim, Weverton Azevedo, Gabriele Rodrigues, Júlia Mayumi, respectivamente.

RESUMO

A tribo Elapomorphini representa um grupo de serpentes opistóglifas da família Dipsadidae, contendo os gêneros *Apostolepis* Cope, 1862, *Coronelaps* Lema and Hofstadler-Deiques, 2010, *Elapomorphus* Wiegmann (in Fitzinger), 1843, *Parapostolepis* Amaral, 1930, e *Phalotris* Cope, 1862. Estes gêneros são diagnosticados com base em caracteres de folidose, coloração, osteologia e morfologia hemipeniana. Destes, *Phalotris* é um gênero diverso de serpentes fossoriais de médio porte, com 19 espécies que ocorrem em áreas abertas do norte do Brasil até o sul da Argentina. O grupo *Phalotris bilineatus* contém ao menos 11 nomes disponíveis, com um extenso histórico taxonômico, e opiniões conflitantes entre autores acerca da validade de suas espécies. Neste trabalho, realizamos uma revisão do complexo *Phalotris* aff. *lemniscatus* e espécies próximas, revisando 243 espécimes e designando unidades taxonômicas operacionais (OTUs), que foram analisadas para morfologia interna e externa, e tiveram suas relações filogenéticas inferidas para dados moleculares. Para avaliar a composição diferencial em toxinas de veneno, geramos transcriptomas de glândula de veneno para duas espécies do complexo *P.* aff. *lemniscatus*. Nossas inferências filogenéticas e rede de haplótipos recuperam relações parafiléticas entre *P. bilineatus*, *P. lemniscatus*, *P. trilineatus* e *P. reticulatus*, com clados correspondendo a regiões geográficas, e não a espécies morfológicas. Nossas análises multivariadas baseadas em dados morfológicos também não apresentam distinções significativas entre espécies do grupo *P. bilineatus*. Encontramos perfis de expressão similares em famílias de toxinas entre *P. lemniscatus* e *P. reticulatus*. Propomos alocar *P. trilineatus* e *P. reticulatus* como sinônimos juniores de *P. lemniscatus*, e apresentamos comentários sobre outros táxons do grupo *P. bilineatus*.

PALAVRAS-CHAVE: Dipsadidae; Elapomorphini; Sistemática; Taxonomia; Transcriptomas

INTRODUÇÃO GERAL

A família Dipsadidae Bonaparte, 1840 representa uma radiação de serpentes avançadas (Caenophidia Hoffstetter, 1939) particularmente diversa, contendo aproximadamente 800 espécies que possuem uma grande variedade de especializações morfológicas e hábitos ecológicos, tendo conquistado a maior parte dos habitats disponíveis nas Américas e Índias Ocidentais, com adaptações para hábitos terrestres, aquáticos, arbóreos e fossoriais, assim como atividades diurnas, noturnas ou catémeras (Hartmann & Marques, 2005; Hedges et al., 2009; Zaher et al., 2009; Grazziotin et al. 2012; Harrington et al. 2018; Zaher et al. 2019). Três subfamílias de serpentes dipsadídeas são conhecidas: Dipsadinae Bonaparte 1838, Carphophiinae Zaher, Grazziotin, Cadle, Moura-Leite, & Bonatto, 2009, e Xenodontinae Bonaparte 1845; destas, Xenodontinae são um grupo monofilético de serpentes que não apresentam sinapomorfia morfológica, e contém 14 tribos (Zaher et al. 2009, 2019).

A tribo Elapomorphini, que contém aproximadamente 45 espécies de hábitos fossoriais ou criptozóicos, fora sujeita a um longo histórico de instabilidade taxonômica e controvérsias entre propostas de arranjos taxonômicos e sistemáticos conflitantes (Ferrarezzi 1993; Harvey 1999; Entiauspe-Neto et al. 2019, 2020a, 2021). Cinco gêneros são conhecidos: *Apostolepis* Cope, 1861, *Coronelaps* Lema and Hofstadler-Deiques, 2010, *Elapomorphus* Wiegmann (in Fitzinger), 1843, *Parapostolepis* Amaral, 1930, e *Phalotris* Cope, 1862, que apresentam como sinapomorfias morfológicas putativas uma sutura frontoparietal em forma de "U", um processo dentígero do dentário curto, ossos frontais dorsalmente incluídos pelo processo anterolateral do parietal, poucas escamas supralabiais (5–7), segunda supralabial em contato com o olho, uma

placa nasal não dividida, entre outros caracteres morfológicos (Savitzky 1979; Ferrarezzi 1993, 1994; Zaher 1994; Zaher et al. 2009).

Os gêneros de Elapomorphini podem ser diferenciados com base em caracteres de folidose e coloração, os quais podem estar associados à adaptações e reduções provocadas por hábitos fossoriais e criptozóicos (Lema, 2001; Ferrarezzi, 1993). *Apostolepis* apresenta um pair de prefrontais possivelmente fusionadas com internasais, 15 dorsais, e uma ponta da cauda preta (este último caráter está ausente em *Apostolepis quinquelineata* Boulenger, 1896; *Parapostolepis* apresenta uma ponta da cauda preta e 17 dorsais; *Phalotris* possui uma placa prefrontal única, separada das internasais; os gêneros *Coronelaps* e *Elapomorphus* apresentam internasais e prefrontais pareadas, porém, *Coronelaps* apresenta um colar nucal parietal amarelo e uma sutura transversa entre o exoccipital e supraoccipital, ausentes em *Elapomorphus*.

O gênero *Phalotris* contém 19 espécies de pequeno e médio porte, que ocorrem desde o nordeste brasileiro e Bolívia, até o sul da Argentina (Ferrarezzi et al. 1993; Entiauspe-Neto et al. 2021; Uetz et al. 2023). As espécies de *Phalotris* são alocadas em três grupos morfológicos, propostos por Ferrarezzi (1993) em uma dissertação de mestrado não publicada: grupo *P. bilineatus*, que apresentam listras dorsais largas e ventrais pretas; grupo *P. nasutus*, que apresentam uma escama rostral projetada, e as temporais medial e posterior possivelmente fusionadas; grupo *P. tricolor*, que apresentam um padrão dorsal de coloração uniforme e colares nucais largos.

Até o presente momento, ao menos 12 nomes estão disponíveis para o grupo *P. bilineatus*: *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854, *Phalotris melanopleurus* Cope, 1885, *Elapomorphus trilineatus* Boulenger, 1889, *Elapomorphus lemniscatus divittatus* Lema, 1984, *Elapomorphus reticulatus* Peters, 1860, *Elapomorphus*

iheringi Strauch, 1884, *Elapomorphus bilineatus* Duméril, Bibrón, & Duméril 1854, *Elapomorphus suspectus* Amaral, 1924, *Phalotris normanscotti* Cabral & Cacciali, 2015, *Elapomorphus bollei* Mertens, 1954, *Elapomorphus spegazzinii* Boulenger, 1913, e *Phalotris illustrator* Scrocchi, Nenda & Giraudo 2022. Entretanto, os arranjos taxonômicos disponíveis apresentam hipóteses de validades conflitantes entre si, variando entre quatro (Ferrarezzi & Puorto, 1993; Giraudo & Scrocchi, 2002; Cabral & Cacciali, 2015; Nogueira et al., 2019; Guedes et al., 2023) e oito (Scrocchi et al. 2023) espécies válidas para o grupo *P. bilineatus*.

Com base neste cenário taxonômico incerto, apresentamos uma revisão integrativa para o complexo *Phalotris* aff. *lemniscatus*, revisando 243 espécimes e designando unidades taxonômicas operacionais (OTUs), que foram analisadas para morfologia interna e externa, e tiveram suas relações filogenéticas inferidas para dados moleculares, e caracterização de toxinas de veneno a partir de dados de transcriptoma. Com base nestes dados, apresentamos uma proposta que visa testar a validade de espécimes associados à *P. bilineatus*, *P. lemniscatus*, *P. reticulatus*, *P. illustrator*, *P. multipunctatus*, *P. suspectus*, *P. spegazzinii* e *P. normanscotti*, utilizados para a delimitação de unidades taxonômicas operacionais (OTUs) neste trabalho.

REFERÊNCIAS BIBLIOGRÁFICAS

Cabral, H., & Cacciali, P. (2015). A new species of Phalotris (Serpentes: Dipsadidae) from the Paraguayan Chaco. Herpetologica, 71(1), 72-77.

Entiauspe-Neto, O. M., Renner, M. F., Mario-da-Rosa, C., Abegg, A. D., Loebmann, D., & de Lema, T. (2017). Redescription, geographic distribution and ecological niche modeling of *Elapomorphus wuchereri* (Serpentes: Dipsadidae). Phyllomedusa: Journal of Herpetology, 16(2), 225-242.

Entiauspe-Neto, O. M., de Sena, A., Tiutenko, A., & Loebmann, D. (2019). Taxonomic status of *Apostolepis barrioi* Lema, 1978, with comments on the taxonomic instability of *Apostolepis* Cope, 1862 (Serpentes, Dipsadidae). Zookeys, 841, 71-78.

Entiauspe-Neto, O. M., Rangel, G. F. M., Guedes, T. B., & Tiutenko, A. (2020a). Raising of Lazarus: Rediscovery and redescription of *Apostolepis niceforoi* Amaral, 1935 (Serpentes: Dipsadidae: Elapomorphini). Holotipus, 1, 21-35.

Entiauspe-Neto, O. M., Guedes, T. B., Loebmann, D., & de Lema, T. (2020b). Taxonomic status of two simultaneously described *Apostolepis* Cope, 1862 species (Dipsadidae: Elapomorphini) from Caatinga enclaves moist forests, Brazil. Journal of Herpetology, 54(2), 225-234.

Entiauspe-Neto, O. M., Koch, C., Harvey, M. B., Colli, G. R., & Guedes, T. B. (2021a). Redescription of *Apostolepis ambiniger* (Peters, 1869) (Serpentes: Dipsadidae: Elapomorphini). Vertebrate Zoology, 71, 231-251.

Entiauspe-Neto, O. M., Tiutenko, A., Azevedo, W. D. S., & Abegg, A. D. (2021b). Extraordinary claims require extraordinary evidence: on the taxonomic identity of *Phalotris cerradensis* Silveira, 2020. Revue suisse de Zoologie, 128(1), 53-60.

Entiauspe-Neto, O. M., Koch, C., Gray, R. J., Tiutenko, A., Loebmann, D., & Guedes, T. B. (2021c). Taxonomic status of *Apostolepis tertulianobeui* Lema, 2004 based on an integrative revision of *Apostolepis assimilis* (Reinhardt, 1861)(Serpentes: Dipsadidae). Zoologischer Anzeiger, 291, 123-138.

Entiauspe-Neto, O. M., Koch, C., Guedes, T. B., Paredero, R. C., Tiutenko, A., & Loebmann, D. (2022). Unveiling an enigma from the Cerrado: taxonomic revision of two sympatric

species of *Apostolepis* Cope, 1862 (Dipsadidae: Xenodontinae: Elapomorphini) from central Brazil. European Journal of Taxonomy, 817, 143-182.

Ferrarezzi, H. (1993). Sistemática filogenética de *Elapomorphus*, *Phalotris* e *Apostolepis* (Serpentes, Colubridae, Xenodontinae). MsC dissertation, Universidade de São Paulo, Brazil.

Ferrarezzi, H. (1994). Uma sinopse dos gêneros e classificação das serpentes (Squamata): II. Família Colubridae. In: Nascimento LB, Bernardes AT, Cotta GA (Eds) Herpetologia no Brasil 1. PUCMG, Belo Horizonte, 81–91.

Ferrarezzi, H., Barbo, F. E., & Albuquerque, C. E. (2005). Phylogenetic relationships of a new species of *Apostolepis* from Brazilian Cerrado with notes on the assimilis group (Serpentes: Colubridae: Xenodontinae: Elapomorphini). Papéis avulsos de Zoologia, 45(16), 215-229.

Guedes, T. B., Entiauspe-Neto, O. M., & Costa, H. C. (2023). Lista de répteis do Brasil: atualização de 2022. Herpetologia Brasileira 12(1): 56-161.

Giraudo, A.R. & Scrocchi, G.J. 2002. Argentinian Snakes: an annotated checklist. Smithsonian Herpetological Information Service, 132:1-53.

Grazziotin, F. G., Zaher, H., Murphy, R. W., Scrocchi, G., Benavides, M. A., Zhang, Y. P., & Bonatto, S. L. (2012). Molecular phylogeny of the new world Dipsadidae (Serpentes: Colubroidea): a reappraisal. Cladistics, 28(5), 437-459.

Harvey, M.B. (1999). Revision of Bolivian *Apostolepis* (Squamata: Colubridae). Copeia 1999: 388–409.

Savitzky, A. H. (1979). The origin of the New World proteroglyphous snakes and its bearing on the study of venom delivery systems in snakes. PhD dissertation, University of Kansas, USA.

Zaher, H. (1994). Phylogenie des pseudoboini et evolution des Xenodontinae sud-americains (Serpentes, Colubridae). PhD dissertation, Musée National D'Histoire Naturelle, France.

Zaher, H., Grazziotin, F. G., Cadle, J. E., Murphy, R. W., Moura-Leite, J. C. D., & Bonatto, S. L. (2009). Molecular phylogeny of advanced snakes (Serpentes, Caenophidia) with an emphasis on South American Xenodontines: a revised classification and descriptions of new taxa. Papéis Avulsos de Zoologia, 49(11), 115-153.

Zaher, H., Murphy, R. W., Arredondo, J. C., Graboski, R., Machado-Filho, P. R., Mahlow, K., ... & Grazziotin, F. G. (2019). Large-scale molecular phylogeny, morphology, divergence-time estimation, and the fossil record of advanced caenophidian snakes (Squamata: Serpentes). PloS one, 14(5), e0216148.

ARTIGO I

Interspecific conservatism in toxin expression revealed by venom-gland transcriptomics of the *Phalotris lemniscatus* species complex

Toxicon

Highly conserved and extremely variable: the paradoxical pattern of toxin expression revealed by comparative venom-gland transcriptomics of *Phalotris* (Serpentes: Dipsadidae)

Omar M. Entiauspe-Neto[1],[2],[*]; Pedro G. Nachtigall[3]; Márcio Borges-Martins[2]; Inácio L. M. Junqueira-de-Azevedo[3]; & Felipe G. Grazziotin[1]

[1] Laboratório de Coleções Zoológicas, Instituto Butantan, 05503-90, Av. Vital Brazil, 1500, Butantã, São Paulo, SP, Brazil.

[2] Programa de Pós-graduação em Biologia Animal, Departamento de Zoologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, CEP 91501-970, Porto Alegre, RS, Brazil.

[3] Laboratório Especial de Toxinologia Aplicada, Instituto Butantan, 05503-90, Av. Vital Brazil, 1500, Butantã, São Paulo, SP, Brazil.

*Corresponding author: Email: omarentiauspe@hotmail.com (OME-N)

ABSTRACT

Although "non-front fanged" snakes represent almost two-thirds of snake diversity, most studies on venom composition focus exclusively on "front-fanged" species, which account for most of the clinically relevant snake accidents. Reports on venom composition and accidents caused by "non-front fanged" snake species are scarce or rare. In this work, we evaluate the case of the poorly known and medically relevant Neotropical snake genus *Phalotris*, using venom-gland transcriptomes to evaluate venom gene composition and expression. We generated venom-gland transcriptomes for three *Phalotris* species, herein identified as *Phalotris reticulatus* associated to the Araucaria Pine forests populations, *Phalotris lemniscatus* associated to the Pampa Grasslands populations, and *Phalotris mertensi* associated to the Cerrado populations. Furthermore, we also evaluate the distinctiveness of toxin expression among these taxa, testing the influence of taxonomic arrangements upon toxin expression. We uncover unexpectedly similar profiles for major expressed toxin transcripts families among evaluated species, which share highly conserved Kunitz-type inhibitors. However, these species also share a different expression profile of snake venom metalloproteinases (SVMP), which presents a distinct set of SVMP isoforms and also truncated type III SVMPs resembling type II and/or type I SVMPs. Our results suggest a conserved evolutionary history of Kunitz-type inhibitor toxins, which may have independent origins among snake clades, and a highly diverse composition of snake venom metalloproteinases, similar to the diversity observed in some viper species.

Key Words: Evolution, Systematics, Toxins, Reptilia.

Introduction

Snakes have several phenotypic innovations, although venom and the associated venom delivery system are considered the key evolutionary novelties that shaped their recent diversification (Greene, 1983; Kochva, 1987; Cundall & Greene, 2000). The development of these traits during the beginning of the Paleogene (Zaher et al., 2019) allowed venomous snakes to chemically-induce prey immobilization and death, as well to evolve an effective defense against potential threats and predators (Savitzky, 1980; Kardong, 1986). Venom and their toxic compounds are considered direct products of selective pressures, consequently, snakes have been used as models to study phenotypic evolution and adaptation (Holding et al., 2021, 2022; Mason et al., 2022). However, most of these studies are focused on medically relevant species, restricting most of the knowledge on snake venom evolution to the families Viperidae and Elapidae (Dawson et al., 2021).

Considering the differences regarding venom delivery systems, snakes have been classified based on aspects of dentition in four main groups: (1) aglyphous, with fixed, ungrooved maxillary teeth, with or without enlarged maxillary rear-fangs; (2) opisthoglyphous, presenting fixed, grooved, and enlarged maxillary rear-fangs; (3) proteroglyphous, with anteriorly positioned, fixed or partially mobile canaliculated maxillary fangs; and (4) solenoglyphous, with anteriorly positioned, mobile canaliculated maxillary fangs. The first two groups are informally known as rear-fanged, non-front-fanged or non-venomous snakes, while the latter two are known as front-fanged or venomous snakes. Besides the fang position and potent toxins that can cause serious human envenomation, front-fanged snakes are also frequently characterized by the presence of venom glands with lumen and muscles that surround and actively compress the glands (Jackson, 2003; Jackson et al., 2017).

Although widely used, snake classification based on the venom delivery system does not represent our current understanding about phenotypic diversity and natural groups (Zaher et al., 2019). Frequently, in this fang-centered scheme, non-front-fanged snakes were classically considered as transitional groups to the highly specialized front-fanged families (Kardog, 1980). However, recent studies have shown that all fanged snakes belongs to the clade Endoglyptodonta (Zaher et al., 2019), and both fangs and venom glands being derived from the same embryological layer, having a common origin (Jackson et al., 2017; Oliveira et al., 2023). Consequently, based on current phylogenetic hypotheses, non-front-fanged groups of Endoglyptodonta are phylogenetically more diverse than the front-fanged groups (Zaher et al., 2019). Additionally, non-front-fanged snakes are phenotypically more variable than front-fanged groups, particularly when considering venom composition and the anatomy of the venom delivery system (McGivern et al., 2014; Bayona-Serrano et al., 2020; 2023; Oliveira et al., 2023).

Non-front-fanged snakes were historically regarded as harmless or not medically relevant (Puorto & França, 2009). However, since the deaths of eminent herpetologists in the twentieth century by "colubrid" envenomation (e.g., Karl Patterson Schmidt in 1957, and Robert Friedrich Wilhelm Mertens in 1975) the high venom toxicity of some opisthoglyphous species, like *Dispholidus typus* (Smith, 1828) and *Thelotornis capensis* (Smith, 1849), became undeniable. Other aglyphous colubrids with enlarged rear-fangs, such as the Asian species *Rhabdophis tigrinus* (Boie, 1826), can also induce life-threatening envenomation in humans (Smeets et al., 1991; Silva et al., 2014). Besides the high venom toxicity, these snakes also present modified venom delivery systems, departing from the typical non-front-fanged structure. For example, *D. typus*, like viperids and elapids, has cephalic muscles that can compress the venom glands (Lake

& Trevor-Jones, 1996). All together, those data indicate the relevance of studies focusing on characterizing and understanding venom composition in these neglected and understudied groups of snakes classified as “harmless”.

In South America, potentially dangerous or life-threatening accidents have been reported for the non-front-fanged genera *Boiruna* Zaher, 1996, *Clelia* Fitzinger, 1826*, Erythrolamprus* Wagler, 1830, *Heterodon* Latreille in Sonnini & Latreille, 1801, *Phalotris* Cope, 1862, and *Philodryas* Wagler, 1830 (Puorto & França, 2009; Weinstein et al., 2011). These snakes belong to the highly diverse family Dipsadidae, which has been the subject of recent phylogenetic studies that revealed novel evolutionary relationships (e.g., Grazziotin et al., 2012; Trevine et al., 2022; Abegg et al., 2022). Also recently, there has been an increased number of studies addressing the venom composition of dipsadids, showing their highly complex set of toxins (e.g., Bayona-Serrano et al., 2020, 2023; Heptinstall et al., 2023; Tioyama et al., 2023).

Among these medically relevant dipsadid genera, *Phalotris* is likely the least studied due to its natural rarity and secretive behavior. The genus encompasses 19 small to medium-sized species, distributed from northern and northeastern Brazil southwards into Argentina and Uruguay (Ferrarezzi, 1993; Guedes et al., 2017; Entiauspe-Neto et al., 2021). These rear-fanged snakes share an extensive muscularization of the venom gland (Savitzky, 1979), which is relatively larger than in other Dipsadidae (Ferrarezzi, 1993). *Phalotris* also presents maxillary fangs that are positioned more anteriorly than most opisthoglyphous dipsadids (Savitzky, 1979). Campos et al. (2016) provided the first characterization of *Phalotris* venom, through the analysis of *P. mertensi* (Hoge, 1955) venom gland transcriptome, and high-resolution venom proteomics. These authors uncovered a high content of Kunitz-type proteins, C-type lectins, and snake venom metalloproteinases (SVMPs). They also evaluated venom activity through enzymatic

essays, which indicated that *P. mertensi* venom presents myotoxic activity approximately three times higher and venom coagulant activity approximately 20 times higher than the venom of *Bothrops jararaca* (Wied-Neuwied, 1824) (Campos et al., 2016).

Few human accidents have been recorded for *Phalotris* species. Those reports revealed that the envenomation symptoms ranged from mild local effects to serious systemic complications (Lema, 2007; Alba et al., 2017). Only three accidents with humans have been recorded, one in Brazil and two in Uruguay. In Brazil, the envenomation was caused by a juvenile of *P. lemniscatus* (Duméril, Bibron, & Duméril, 1854), producing local pain, renal insufficiency, hepatic and cerebral injuries. It was treated with blood transfusion and anti-bothropic anti-venom serum, with no assessment of efficacy for the treatment (Lema 2007). In Uruguay, two envenomations were caused by adult *P. lemniscatus* specimens (Alba et al. 2017). Both accidents caused blood coagulation alterations, subsequently treated with anti-bothropic anti-venom serum. Despite its evident medical importance, very little is known about the venom composition of the *Phalotris* species diversity.

Here, we expand the knowledge about the venom characterization of *Phalotris* through the analysis of the venom gland transcriptome of three species. We present the first assessment of the venom composition of *P. lemniscatus* and *P. reticulatus* (Peters, 1860). We also provide a new evaluation of the venom gland transcriptome of *P. mertensi*, comparing the composition and expression levels of toxins among these three species. Additionally, we also evaluate possible signals of positive selection shaping the diversity of toxin isoforms in the venom of *Phalotris*. Finally, we analyzed toxic isoforms of SVMPs and Kunitz to better understand their evolutionary history within the group.

Materials and Methods

Sample collection and identification

A total of four *Phalotris* specimens were collected (Table 1, Fig. 1) from three different localities in southern and southeastern Brazil between November 2017 and January 2018. These specimens are attributable to three species, *Phalotris reticulatus* (field number SB0254), that occurs in Araucaria Pine forests; *Phalotris lemniscatus* (SB0445, SB0446), that occurs in Pampa Grasslands; and *Phalotris mertensi* (SB0216) that occurs in the Cerrado of southeastern and central-western Brazil. Snout-to-vent length (SVL) was measured with a flexible ruler, from the tip of the rostral scale to the cloacal scale, and ranged from 257 to 610 mm.

We collected venom from the specimens using a modified protocol based on Rosenberg (1992), Hill and Mackessy (1997) and Hofman et al. (2018). Snakes were subjected to anesthesia, using an airway administration of isoflurane, followed by a subcutaneous ventral injection of pilocarpine (approximately 6 μg/g relative to the snake mass), a muscarinic cholinergic agonist to stimulate venom secretion. Afterwards, we collected venom using capillary tubes placed on each maxillary rear fang. The venom was vacuum dried and stored at −80 ºC. Four days after the venom collection, snakes were euthanized with an intracardiac injection of sodium pentobarbital (100mg/kg), according to standard approved AVMA guidelines. Immediately after euthanasia, both venom glands were excised and stored in RNAlater (Thermo Fisher Scientific, US) at −80°C until use.

RNA extraction and sequencing

Total RNA from the venom gland was extracted using Trizol Reagent (Invitrogen, US), following some modifications in the manufacturer's protocol as described in Tioyama et al.

(2023). To evaluate RNA integrity, quantity, and quality, we quantified the extracted RNA using a Quant-iTTMRiboGreen RNA reagent and kit (Thermo Fisher Scientific, US), measure the RNA integrity using the Agilent RNA 6000 Nano Kit (Agilent Technologies, US) and measure the contamination through UV absorbance using NanoDrop 1000 (Thermo Scientific, US). After ensuring sufficient RNA weight and quality for library preparation, we generated cDNA libraries with Illumina TruSeq Stranded RNA HT Kit (Illumina, US). Quantification of cDNA libraries was performed with a real-time PCR employing KAPA SYBR FAST Universal qPCR Kit, according to the manufacturer's protocol, using the StepOnePlusTM Real-Time PCR System (ThermoFisher, US). Pooled libraries were sequenced with a 2 x 150 base pair (bp) paired-end reads on an Illumina HiSeq 1500 platform at Instituto Butantan, São Paulo, Brazil.

Transcriptome assembly and annotation

We trimmed Illumina adapters and removed low-quality reads (i.e., reads with Phred score < 25 and length < 75 bp) using Trim Galore! (v0.4.4; https://github.com/FelixKrueger/TrimGalore). Then, the trimmed reads were merged using PEAR (Paired-End reAdmergeR; Zhang et al. 2014). The merged reads were used to perform a *De Novo* assembly methodology following the recommendations of Holding et al. (2018). We used five distinct assemblers: Trinity (kmer parameter set to 31; Haas et al. 2013), Extender (overlap parameters set to 120 and 150; Rokyta et al. 2012), NGen (on default parameters; Lasergene DNAStar software package, US, available at https://www.dnastar.com/t-nextgen-seqman-ngen.aspx), Bridger (kmer parameter set to 30; Chang et al., 2015), and rnaSPAdes (kmer parameter set to 31, 75, and 127; Bushmanova et al., 2019). To remove redundancy, the assemblies were combined and clustered with 100% identity

using the “RemoveRedundancy.py” script available in the repository of ToxCodAn (https://github.com/pedronachtigall/ToxCodAn).

We annotated toxins using ToxCodAn (Nachtigall et al. 2021b), which is an open-source python script designed to perform precise annotation of snake venom gland transcriptomes. The output of identified toxins (cds; redundancy filtered) and putative toxins (sp; filtered) were combined into a single file and chimeric transcripts were removed using the ChimeraKiller script (v0.7.4; https://github.com/masonaj157/ChimeraKiller). Afterwards, the final toxin coding sequences set, accounting for <99% similarity with *cd-hit*, to reduce redundancy of repeated transcripts and group allelic variation within unique loci, was assembled. The nontoxin venom components were annotated using the nontoxin contigs output by ToxCodAn. We performed coding sequence prediction using CodAn () with the vertebrate model, followed by a BLAST search of predicted sequences using a custom proteinDB incorporated in ToxCodAn repository (https://github.com/pedronachtigall/ToxCodAn) and the Swissprot database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/swissprot.tar.gz). For the coding sequences which were not annotated in the previous step, we employed a HMM search using hmmsearch (HMMER 3.2.1; http://hmmer.org/), using a Hidden Markov Model models from BUSCO (Simão et al. 2018), and hmmscan (HMMER 3.2.1; http://hmmer.org/) with the Hidden Markov Models from Pfam (https://pfam.xfam.org/). Then, we combined the toxin and nontoxin final sets to generate the final venom-gland transcriptome annotated set for each species and clustered it with 99% similarity using cd-hit (Fu et al., 2012) to reduce redundancy of repeated transcripts and group allelic variation at a single locus. For *P. lemniscatus* that had two specimens, we combined their final and annotated venom-gland transcriptome and clustered with 98% identity using CD-HIT

(Fu et al., 2012) reduce redundancy of repeated transcripts and group allelic variation at a single locus. The final species-specific consensus transcriptome was used in downstream analysis.

Expression analyses and ortholog identification

We performed transcript quantification by mapping the merged reads against the species consensus transcriptome using Bowtie2 (v2.3.5; Langmead and Salzberg, 2012) and setting the mismatch rate parameter to 0.02, while relative expression was calculated using RSEM (v1.3.1; Li and Dewey, 2011). We assessed the intraspecific variation for *P. lemniscatus* by generating a pairwise null distribution of expression divergence based on nontoxin expression to identify putative outliers that may represent divergence in expression level (Rokyta and Ward, 2017). The data was centered log-ratio (clr) transformed to normalize the expression distributions while accounting for the compositional nature of relative expression values (i.e., TPM). The toxin transcripts with pairwise divergence in expression outside the 99th percentile of the centered log-ratio transformed distribution of nontoxins were considered outliers that may represent differentially expressed toxins.

To perform interspecific variation analysis between *P. lemniscatus* and *P. reticulatus*, we first inferred orthology among toxin and nontoxins transcripts of both species using OrthoFinder (v2.3.14; Emms and Kelly, 2019). OrthoFinder allows the identification of groups of sequences derived from a single gene that share the same common ancestor of compared species (i.e., orthogroups), and also identifies conserved orthologs within orthogroups. Therefore, OrthoFinder output allows the identification of orthology among the analyzed sequences for a comparative analysis between species.

For nontoxins, we filtered low expressed transcripts by keeping only genes with TPM higher than 1 in both individuals of each species (i.e., genes with TPM lower or equal to 1 in one individual were removed from downstream analysis). Afterwards, we applied OrthoFinder to detect orthogroups among nontoxins to compare the expression level of these orthogroups between both species as mentioned above in the intraspecific variation. The toxin transcripts with expression placed out of the 99th percentile of the centered log-ratio transformed distribution of nontoxins were considered outliers that may represent differentially expressed toxins between the species.

We performed a divergence analysis of toxin and nontoxin orthologous of *P. lemniscatus* and *P. reticulatus* by calculating pairwise synonymous substitution rates (dS), nonsynonymous substitution rates (dN), and their dN/dS ratios (ω). Each orthologous pair was aligned by codon using PRANK (v.170427). The alignments were used as input to estimate dS, dN, and ω using codeml from paml package (v4.9). We used the dS, dN and ω to compare toxins against a background set of nontoxins (Rokyta et al., 2013) to identify if toxin genes present higher dS and/or dN and also to test for positive selection in toxins (i.e., higher values of ω). We removed orthologous pairs with dS <0.001, to eliminate the possibility of excessively inflated ω values, and pairs with dS <0.10 to eliminate putative misidentified orthologs. Statistical differences between toxins and nontoxins were tested using the Wilcoxon sign rank test.

We also explored phylogenetic relationships among Kunitz-type inhibitor toxins, pooling available sequences of venom-derived proteins putatively homologous with Kunitz-type inhibitor domain or WAP domain for fused toxins. We delimited intron-exon boundaries by comparing sequences of cDNAs and genes. We performed preliminary sequence alignment using MAFFT 1.3.6 (Katoh et al., 2013) plugin in Geneious (v. 7.1.8.; Kearse et al., 2012), and then manually

aligned for a correct reading frame. We inferred phylogenetic relationships under a Maximum Likelihood (ML) framework, computed using RAxML (Stamatakis, 2014) in CIPRES Science Gateway (available at https://www.phylo.org/), searching the most likely tree 100 times and conducting 1,000 nonparametric bootstrap replicates using the GTR + Γ. We also compared putative toxins recovered as Snake Venom Metalloproteinases with the ones described for *P. mertensi* by Campos et al. (2016), with the same parameters as for Kunitz-type inhibitors.

Results

The venom-gland transcriptome of *P. mertensi* (Fig. 2; Supplementary file 1, Table S1) consisted of 6485 transcripts, of which 62 were toxin encoding transcripts. Its composition was very similar to that described by Campos et al. (2016), and therefore, we are commenting and comparing its venom composition in the discussion. On the other hand, we are describing in detail the results for the venom gland transcriptome of *P. reticulatus* and *P. lemniscatus* since they were sequenced here for the first time.

The venom-gland transcriptome of *P. reticulatus* (Fig. 2; Table S2) consisted of 3778 transcripts, of which 51 represented toxin transcripts. Those toxin transcripts were classified into 24 distinct families (as ordered by expression proportion: Kunitz-type proteins, snake venom metalloproteinase, C-type lectin, C-type natriuretic peptide, l-amino acid oxidase, Vespryn, three finger toxin, snake venom matrix metalloproteinase, snake venom serine protease, phospholipase B, translationally controlled tumor protein, cysteine-rich secretory protein, nucleotidase, hyaluronidase, Cystatin, vascular endothelial growth factor, snake venom acid lipase, Papilin-like, Ficolin, wap domain containing protein, double wap domain containing protein, acetylcholinesterase). The venom-gland consensus transcriptome of *P. lemniscatus* (combination

of SB0445 and SB0446 transcriptomes; Fig. 2; Table S3) consisted of a total 6984 transcripts, of which 63 were toxin transcripts and classified into 24 distinct families (as ordered by mean expression proportion: Kunitz-type proteins, C-type lectin, snake venom metalloproteinase, three finger toxin, C-type natriuretic peptide, snake venom serine protease, l-amino acid oxidase, Vespryn, translationally controlled tumor protein, phospholipase B, cysteine-rich secretory protein, nucleotidase, snake venom matrix metalloproteinase, Ficolin, Cystatin, Papilin-like, acetylcholinesterase, wap domain containing protein, vascular endothelial growth factor, hyaluronidase, Ku-wap-fusin-like protein, snake venom acid lipase, phospholipase A2), with differential intraspecific expression observed for two families (KUWAPF and PLA2 were absent in SB0446).

The major expressed toxin families of *Phalotris reticulatus* were a single Kunitz-type inhibitor (KUNZ, 53.22 %; TPM = 410710.97, $n = 1$), 15 unique snake venom metalloproteinases (SVMP, 25.4 %; TPM range = 3030.48–72565.83, 6642.23 ± 13409.99, $n = 15$), 12 unique C-type lectins (CTL, 8.7 %; TPM range = 10.27–13848.8, 2845.75 ± 4617.82, $n = 12$), and two unique C-type natriuretic peptides (CNP, 4.2 %; TPM range = 12814.69–19601.3, 16207.9 ± 3393.3, $n = 2$). Similarly, *Phalotris lemniscatus* expressed two unique Kunitz-type inhibitors (KUNZ, 69 %; TPM range = 83229.14–467506.56, 248952.77 ± 166703.24, $n = 4$), fifteen unique C-type lectins (CTL, 11.7 %; TPM range = 0.87–467506.56; 5399.20 ± 9158.44, $n = 30$), twenty unique snake venom metalloproteinases (SVMP, 10.5 %; TPM range = 0.4–10118.97; 3786.51 ± 5624.67, $n = 40$), and three unique C-type natriuretic peptides (CNP, 2.52 %; TPM range = 35.77–7271.27; 6079.96 ± 7458.47, $n = 6$). Several other minor components (< 2% of both total and toxin expression) were recovered for all three species, such as a three-finger toxin (3FTx), snake venom matrix metalloproteinase (svMMP), Vespryn

(VESP), snake venom serine protease (SVSP), and translationally controlled tumor protein (TCTP). A complete overview of expressed toxins and non-toxins can be seen at Tables S1-S3 (Supplementary file 1).

As for our evaluated pairwise synonymous substitution rates (dS), nonsynonymous substitution rates (dN), and dN/dS ratios (ω) between the toxin and nontoxin orthologs (Fig. 5), we uncovered a higher ω ratio for toxins than nontoxins, with approximate mean values of 0.92, and 19 toxin isoforms had ω values greater than 1, indicating positive selection. We also detected that synonymous and nonsynonymous substitution rates were significantly higher in toxins when compared to nontoxins. Of these, a CTL isoform (CTL-1) had the highest ω value ($\omega = 3$), as well as the highest synonymous and nonsynonymous values among toxins. We also uncovered a significant amount of SVMP isoforms (SVMP 1–9) with positive selection ($\omega > 1$).

Regarding the major expressed toxin families, we uncovered a remarkably conserved pattern of toxin expression among Kunitz-type inhibitor toxins, with two isoforms bearing similar expression levels among all sampled *Phalotris* species (Fig. 3). On the other hand, we identified twenty distinct isoforms of SVMP, 12 for *P. mertensi* and 13 for *P. lemniscatus* and *P. reticulatus* (similar isoforms in both species). Eleven of the 12 different SVMP transcripts of *P. mertensi* were identified as a P-III class SVMP, except for one isoform that has truncated disintegrin-like domain (‘*P.mertensi|SVMP-5’*; Fig. 4). Similarly, 11 of the 13 different SVMP transcripts of *P. lemniscatus* and *P. reticulatus* are identified herein as P-III class SVMPs, except for two isoforms that have a truncated cystein-like domain (‘*P.reticulatus|SVMP-10’*; Fig. 4). This truncated isoform presents a cystein-like domain with half (six of 12) of the cysteines present in the non-truncated ones (Fig. 4). Considering the expression and definition of orthogroups of SVMPs, we uncovered a cluster of four sequential (SVMP 3–6) and three isolated

(SVMP 8, 11–12) isoforms bearing differential expression that are absent in *P. mertensi* and present among *P. lemniscatus* and *P. reticulatus*. Another cluster (SVMP 19–20) is expressed in *P. mertensi* and absent among *P. lemniscatus* and *P. reticulatus*. As for C-type lectins, most of the evaluated isoforms have a conserved pattern of toxin expression among species, except for a cluster of two isoforms (CTL 11–12) that are expressed by *P. mertensi* and absent in *P. lemniscatus* and *P. reticulatus* (Fig. 3).

Our Kunitz-type inhibitor toxin alignment consists of 204 sequences, with 309 base-pairs. We uncovered three major toxin groups (Fig. 6): "Classic" Kunitz proteins, with a signal peptide and Kunitz domain; Ku-wap-fusin-like proteins, with a signal peptide, Kunitz domain, whey-acidic protein domain (WAP), and a C-tail; "Terminal" Kunitz proteins, with a signal peptide, Kunitz domain, and a Terminal Kunitz domain. We recover two major clades of Kunitz-type inhibitor toxins: Clade A (bootstrap support = 99), composed of "Classic" Kunitz and Ku-wap-fusin-like proteins, and terminals of Colubridae, Dipsadidae, and Elapidae snake families; Clade B (bootstrap support = 99), composed of "Terminal" Kunitz proteins and Ku-wap-fusin-like proteins, and terminals of Colubridae, Australian Elapidae, and Viperidae snake families.

Discussion

Our results show striking differences when comparing toxin expression profiles among some species of *Phalotris* (Fig. 3), and seem to be associated with intrageneric phylogenetic relationships (Zaher et al., 2019). *Phalotris lemniscatus* and *P. reticulatus* are classified in the *Phalotris bilineatus* species group, whereas *P. mertensi* belongs to the *Phalotris tricolor* species group (sensu Ferrarezzi, 1993). Considering the three species analyzed here, the inferences of

Zaher et al., (2019) suggest that tricolor and bilineatus species groups do not share a sister relationship. Therefore, our analyses of levels of toxin expression are, in part, probably mirroring the evolutionary history within *Phalotris* and also suggesting that the toxin profile may be resulted from a common ancestor of both species groups.

In general, both species groups exhibit Kunitz-type inhibitors as the main toxin class in their venom-gland transcriptomes, suggesting they represent a key component in the *Phalotris* venom (Fig. 3). This expression pattern was already shown by Campos et al. (2016) for *P. mertensi*, and our results corroborate the dominance of Kunitz-type inhibitors in the venom of other two species of *Phalotris*. The expression levels of Kunitz-type isoforms seem to be conserved in all species, showing the highest levels among all sequenced transcripts (Tables S1-S3). Kunitz-type inhibitors act primarily by blocking protease activity, such as protein breakdown or catalysis of proteolysis; these Kunitz-type inhibitors usually have substrate-like inhibition of serine proteases, although in some Elapidae venoms they may present neurotoxic activity, through the release of acetylcholine that may act upon neuromuscular junctions (Dawson et al., 2021).

Species of *Phalotris* share nocturnal, fossorial, or cryptozoic habits, and feed upon elongated vertebrates, such as legless lizards, amphisbaenids, other snakes and caecilians (Carreira, 2002; Duarte, 2006; Marques et al., 2009, 2015). The physical constraints associated with fossoriality and ophiophagy likely caused substantial impacts on the phenotypic evolution of a fossorial predator like *Phalotris* (Savitzky, 1983; Gans, 1968). Killing and handling elongated, aggressive, and muscularly strong vertebrates (e.g., amphisbaenids, Gans, 1974) inside underground galleries or chambers can be seen as a challenge for a nonvenomous snake. The presence of a neurotoxic action in the venom of *Phalotris*, mediated by Kunitz-type

inhibitors could be an adaptive novelty that allows them to prey elongated vertebrates in underground environments. However, Campos et al. (2016) suggested that the Kunitz-type inhibitors of *P. mertensi* are more likely to function as true protease inhibitors than as neurotoxins.

Kunitz-type inhibitors have also been reported as a major component for other snake genera that present predominantly nocturnal, fossorial habits and ophiophagous diet, such as *Atractaspis* (Terrat et al. 2013; Al-Sadoon et al. 2020), *Bungarus* (Liu et al. 1983), *Micrurus* (Vivas et al. 2016; Oliveira et al., 2023), *Ophiophagus* (He et al. 2008), and some species of *Naja* (Suvilesh et al. 2017; Maritz et al. 2021). Considering that these taxa belong to different families (Elapidae, Atractaspididae, Dipsadidae), it is possible that Kunitz-type inhibitors might constitute a recurrent toxin phenotype in species that have convergent fossorial habits and ophiophagous diet. Moreover, it suggests that Kunitz-type inhibitors may have convergently evolved to perform in similar toxic roles for those distinct lineages.

Corroborating these similarities between the venom composition of *Phalotris* and elapids, our phylogenetic analysis of Kunitz-type inhibitor proteins also suggests different origins among snake families. “Classic” Kunitz-like proteins appear to be restricted to *Phalotris* and elapids, while “Terminal” Kunitz-like proteins occur among colubrids, Australian elapids, and viperids. These two forms of Kunitz-type inhibitor seem to have originated independently from different ancestral Ku-wap-fusin-like proteins (Fig. 6). However, our phylogenetic tree was only based on the analysis of the aligned sequences of the signal-peptide and the Kunitz domain—avoiding the complex WAP domain region. Consequently, our phylogenetic inference could not account for other models of sequence evolution, like exon shuffling (Patthy, 1999; Wang et al., 2017) or transcript fusion through trans-splicing events (Barresi et al., 2019). It has been hypothesized that

the evolution of Kunitz-type protease inhibitors in snakes is not restricted to events of sequence substitutions or indels, but also involves exon/domain rearrangements (Doley et al., 2010). Therefore, we suggest that only with whole genome data integrated to full-length transcriptome sequencing approaches (e.g., using long-reads technologies) may allow to properly analyze and compare the evolution of all Kunitz-type inhibitor domains present in snake venoms.

Conversely from the conserved expression of Kunitz-type inhibitors, SVMPs show contrasting differences of expression levels among species (Fig. 3). Although SVMP represents the second most expressed class of toxins in all analyzed species, the expression of isoforms between the species groups was contrastingly different. Snake venom metalloproteinases are a diverse family of zinc-dependent proteolytic enzymes, that act primarily upon the disruption of cell hemostasis, and therefore may cause consumption coagulopathy, cellular hemorrhage, tissue destruction, and hypovolemic shock, which can cause life-threatening envenomation in humans (Dawson et al., 2021).

Currently, three classes of SVMPs are known (P-I, P-II, and P-III), classified based on the presence or absence of domains downstream of the metalloproteinase domain (Fig. 4; Dawson et al., 2021). The P-III class of SVMPs consists of metalloproteinase, disintegrin-like and cysteine-rich domains. This class can be found in almost all families of endoglyptodonts (sensu Zaher et al., 2009), but it is particularly prevalent in the venom of viperids and some colubroideans. The P-II class is characterized by the loss of the cysteine-rich domain and a modification of the disintegrin-like domain to a disintegrin domain; while the P-I class lacks the cysteine-rich and the disintegrin domain, being formed only by metalloproteinase domain. Classes P-II and P-I are known only from viperid venoms, and they seem to have evolved from a P-III class isoform, through sequential losses of non-metalloproteinase domains (Giorgianni et

al., 2020; Dawson et al., 2021). The classes P-II and P-III likely have their toxic activities mediated or facilitated by their non-enzymatic accessory domains, which might account for systemic dispersal, and therefore, increased hemorrhagic activity. The toxic actions of the P-I class are not completely understood, although the lack of non-metalloproteinase domains seems to reduce the systemic activity, restricting the effects to local regions surrounding the inoculation area (Herrera et al. 2015).

Although simplified forms of SVMPs P-III have been described only for viperids, Campos et al. (2016) discovered that *P. mertensi* has a SVMPs P-III with truncated disintegrin-like domain (P-III-tD). We confirmed their results by recovering the P-III-tD in the transcriptome of the venom glands of a different individual of *P. mertensi* (Fig. 4). Our results further corroborate the tendency of simplification of SVMPs P-III in *Phalotris* through the discovery of a SVMPs P-III with truncated cystein-like domain (P-III-tC) in the venom gland transcriptome of *P. reticulatus*. Although we still need to confirm through proteomics the presence of P-III-TB in the venom of *P. reticulatus*, we can interpret this finding as a corroboration that species of *Phalotris*, like viperids, present simplified forms of SVMPs P-III and may present a similar genomic context for the SVMP locus (Giorgianni et al., 2020; Almeida et al., 2021; Nachtigall et al., 2024).

We still lack compelling evidence that these two simplified forms of SVMPs P-III (P-III-tD and P-III-tC) from *Phalotris* can be considered analogous to the classes P-I and P-II of viperids. However, the expression level of the SVMPs P-III-tD in *P. mertensi* is relatively high, representing the fifth most expressed SVMP in our study (first most expressed SVMP transcript in Campos et al., 2016) and the ninth most expressed toxin in the venom gland (Table S1). Such

high expression level can be interpreted as a direct result of selective pressures, maintaining the relative abundance of this isoform in the *P. mertensi* venom.

On the other hand, P-III-tC from *P. reticulatus* is a relatively lowly expressed transcript, representing the 10th most expressed SVMP and the 22nd most expressed toxin in the venom gland (Table S2). Although lowly expressed, the presence of P-III-tC supports previous results showing that the SVMP diversity in *Phalotris* venom is higher than expected for most species of Colubroidea (Junqueira-de-Azevedo et al., 2016). The venom gland transcriptomes of different dipsadid tribes have been studied in the last years (e.g., Alsophiini, Echinantherini, Hydropsini, Imantodini, Philodryadini, Pseudoboini, Tachymenini, Xenodontini; (Junqueira-de-Azevedo et al., 2016; Byona-Serrano, 2020, 2023; Cerda et al., 2022; Tioyama et al., 2023; Heptinstall et al., 2023) and several genera—like *Alsophis*, *Philodryas*, *Oxyrhopus*, *Siphlophis*—are currently known for high expression levels of SVMP in their venom glands. Despite the diversity of isoforms present in the *Phalotris* venom is rare (only some species of *Philodryas* have shown more than 10 isoforms), and simplified forms of SVMPs P-III were never found in other dipsadids (Junqueira-de-Azevedo et al., 2016), the increase of studies analyzing those neglected snake clades may reveal that such simplification of toxins may be more common than expected.

Concurring with the high number of isoforms, presence of transcripts with truncated domains, and differences in expression level among species, we found signs of positive selection in most of the SVMP transcripts when comparing closely related species (Fig. 5). These results suggest that selective pressures are likely driving species of *Phalotris* to increase the SVMP diversity in their venoms as previously described in viperids (Mason et al., 2020; Nachtigall et al., 2022). Although the expression levels of SVMPs are mostly similar between *P. lemniscatus* and *P. reticulatus* (Figs. 3, 5B), the level of nonsynonymous substitutions among the sequences

of each orthogroup are significantly higher when compared with other toxins (Fig. 5A). The presence of signs of selection shaping the divergence of SVMPs among closely related species of *Phalotris* brings another level of support on the prominence of SVMPs in the venom of this genus of fossorial, ophiophagous snake. The role of such diversity of SVMPs in *Phalotris*, however, needs to be further tested through functional and enzymatic essays.

Conclusion

Our transcriptome characterization of the venom glands of three *Phalotris* species shows that this dipsadid genus has similar patterns of venom composition to those present in some front-fanged snake families. Transcripts of Kunitz-type inhibitors are the most expressed toxin in all analyzed species, with expression levels similar to those found in other nocturnal, fossorial and ophiophagous snakes, like some species of elapids. Additionally, like in most viperids, species of *Phalotris* express a highly variable set of SVMP isoforms (more than 20 in some species), presenting simplified forms of SVMPs P-III with truncated non-metalloproteinase domains (P-III-tD and P-III-tC) resembling the truncated SVMPs from viperids (Giorgianni et al., 2020; Almeida et al., 2021).

Considering the overall conserved pattern in venom gland toxin expression seen between the *Phalotris tricolor* and *Phalotris bilineatus* species groups (high expression of Kunitz-type inhibitors and SVMPs), it is likely that other congeners might have similar venom profiles, and despite rare, human encounters with these other species might result in medically relevant accidents. Additionally, no venom characterization is available to other genera, like *Apostolepis* Cope, 1862, *Coronelaps* Lema & Hofstadler-Deiques, 2010, and *Elapomorphus* Wiegmann in Fitzinger, 1843, which are closely related to *Phalotris*, being classified in the tribe

Elapomorphini (Ferrarezzi, 1993). Further studies are necessary to bridge this gap and provide a comprehensive venom characterization to these rare, morphologically distinct, and medically relevant species of dipsadids. Moreover, expanding the set of species analyzed and better understanding their venom composition may bring novel opportunities for finding potential medicines and development of efficient antivenoms (Casewell et al., 2020; Oliveira et al., 2022).

Funding

OME-N and PGN thank to Fundação de Amparo à Pesquisa do Estado de São Paulo for their research grants (FAPESP 2021/13671-7 and 2018/26520-4, respectively). ILMJA and FGG are supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (304532/20134, 312016/2021-2 and 405518/2021-8). This research was supported by grants from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) to ILMJA (BIOTA/FAPESP 2016/50127-5; CEPID/FAPESP 2013/07467-1), and to FGG (FAPESP 2022/12660-4).

Ethical statement

All experimental procedures were reviewed and approved by the Ethics Committee on Animal Use of the Instituto Butantan (CEUAIB; Protocol Number 4479020217). The specimens used in the present study were registered in the Brazilian National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SisGen; Registration Number A6E9C41). The species were collected under the permits from Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio; permit numbers 56576, 57585 and 66597). The authors compiled with the ARRIVE guidelines (https://arriveguidelines.org/arri ve-guidelines).

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank Arthur D. Abegg (Projeto Dacnis), Diego J. Alvares (UFRGS), Marcelo R. Duarte (IBSP), and Michel de Aguiar Passos for kindly providing photographs of *Phalotris* specimens in life.

Literature Cited

Abegg, A. D., Santos Jr, A. P., Costa, H. C., Battilana, J., Graboski, R., Vianna, F. S., ... & Grazziotin, F. G. (2022). Increasing taxon sampling suggests a complete taxonomic rearrangement in Echinantherini (Serpentes: Dipsadidae). Frontiers in Ecology and Evolution, 10, 969263.

Almeida, D. D., Viala, V. L., Nachtigall, P. G., Broe, M., Gibbs, H. L., Serrano, S. M. D. T., ... & Junqueira-de-Azevedo, I. L. (2021). Tracking the recruitment and evolution of snake toxins using the evolutionary context provided by the *Bothrops jararaca* genome. Proceedings of the National Academy of Sciences, 118(20), e2015159118.

Al-Sadoon, M. K., Paray, B. A., Rudayni, H. A., Al-Mfarij, A. R., & Albeshr, M. F. (2020). Seasonal food composition of a burrowing asp, Atractaspis engaddensis Haas, 1950 from natural habitats of an arid Arabian desert. Journal of King Saud University-Science, 32(4), 2393-2396.

Bayona-Serrano, J. D., Viala, V. L., Rautsaw, R. M., Schramer, T. D., Barros-Carvalho, G. A., Nishiyama Jr, M. Y., ... & Junqueira-de-Azevedo, I. L. (2020). Replacement and parallel simplification of nonhomologous proteinases maintain venom phenotypes in rear-fanged snakes. Molecular Biology and Evolution, 37(12), 3563-3575.

Bayona-Serrano, J. D., Grazziotin, F. G., Salazar-Valenzuela, D., Valente, R. H., Nachtigall, P. G., Colombini, M., ... & Junqueira-de-Azevedo, I. L. M. (2023). Independent recruitment of different types of Phospholipases A2 to the venoms of Caenophidian snakes: the rise of PLA2-IIE within Pseudoboini (Dipsadidae). Molecular Biology and Evolution, 40(7), msad147.

Barresi, V., Cosentini, I., Scuderi, C., Napoli, S., Di Bella, V., Spampinato, G., & Condorelli, D. F. (2019). Fusion transcripts of adjacent genes: New insights into the world of human complex transcripts in cancer. International Journal of Molecular Sciences, 20(21), 5252.

Cabral, H., & Cacciali, P. (2015). A new species of *Phalotris* (Serpentes: Dipsadidae) from the Paraguayan Chaco. Herpetologica, 71(1), 72-77.

Campos, P. F., Andrade-Silva, D., Zelanis, A., Paes Leme, A. F., Rocha, M. M. T., Menezes, M. C., ... & Junqueira-de-Azevedo, I. D. L. M. (2016). Trends in the evolution of snake toxins underscored by an integrative omics approach to profile the venom of the colubrid Phalotris mertensi. Genome biology and evolution, 8(8), 2266-2287.

Carreira, S. (2002). Alimentacíon de ofídios de Uruguay. Asociacíon Herpetológica Española, Monografias de Herpetología 6, 126 pp.

Casewell, N. R., Wüster, W., Vonk, F. J., Harrison, R. A., & Fry, B. G. (2013). Complex cocktails: the evolutionary novelty of venoms. Trends in ecology & evolution, 28(4), 219-229.

Cerda, P. A., Crowe-Riddell, J. M., Gonçalves, D. J., Larson, D. A., Duda Jr, T. F., & Davis Rabosky, A. R. (2022). Divergent Specialization of Simple Venom Gene Profiles among Rear-Fanged Snake Genera (*Helicops* and *Leptodeira*, Dipsadinae, Colubridae). Toxins, 14(7), 489.

Costa H.C. & Bérnils R.S. (2018). Répteis do Brasil e suas unidades federativas: lista de espécies. Herpetologia Brasileira 1: 11–48.

Cundall, D., & Greene, H. W. (2000). Feeding in snakes. In K. Schwenk (Ed.). Feeding: Form, function and evolution in tetrapod vertebrates (pp. 293–333). Academic Press.

Dawson, C. A., Ainsworth S., Albulescu, L. O., & Casewell, N. (2021). Snake Venom Metalloproteinases. In: Mackessy, S. P. Reptile Venoms and Toxins: Unlimited Opportunities for Basic and Applied Research. In Handbook of Venoms and Toxins of Reptiles (pp. 3-18). CRC Press.

Debono, J., Dobson, J., Casewell, N. R., Romilio, A., Li, B., Kurniawan, N., ... & Fry, B. G. (2017). Coagulating colubrids: Evolutionary, pathophysiological and biodiscovery implications of venom variations between boomslang (Dispholidus typus) and twig snake (Thelotornis mossambicanus). Toxins, 9(5), 171.

Doley, R., Pahari, S., Reza, M. A., Mackessy, S. P., & Kini, R. M. (2010). The gene structure and evolution of ku-wap-fusin (Kunitz Waprin fusion protein), a novel evolutionary intermediate

of the Kunitz serine protease inhibitors and waprins from Sistrurus catenatus (Massasauga Rattlesnake) venom glands. The Open Evolution Journal, 4(1).

Duarte, M.R. (2006). Natural history notes. *Phalotris mertensi* (false coral snake) and Amphisbaena mertensi (NCN). Predation. Herpetological Review 37(2): 234.

Entiauspe-Neto, O. M., Tiutenko, A., Azevedo, W. D. S., & Abegg, A. D. (2021). Extraordinary claims require extraordinary evidence: on the taxonomic identity of *Phalotris cerradensis* Silveira, 2020. Revue suisse de Zoologie, 128(1), 53-60.

Ferrarezzi, H. (1993). Sistemática filogenética de *Elapomorphus, Phalotris* e *Apostolepis* (Serpentes: Colubridae: Xenodontinae), Master Dissertation. Universidade de São Paulo.

Ferrarezzi, H. (1994). Uma sinopse dos gêneros e classificação das Serpentes (Squamata): II. Família Colubridae; pp. 81–91. In : Nascimento, L.B., Bernardes, T.A., and Cotta, G.A. (Eds.), Herpetologia do Brasil 1. Editora PUC-MG Fundação Ezequiel Dias. Fundação Biodiversitas. Belo Horizonte.

Fry, B. G. (2005). From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. Genome research, 15(3), 403-420.

Fry, B. G., & Wüster, W. (2004). Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. Molecular biology and evolution, 21(5), 870-883.

Fry, B. G., Scheib, H., van der Weerd, L., Young, B., McNaughtan, J., Ramjan, S. R., ... & Norman, J. A. (2008). Evolution of an arsenal: structural and functional diversification of the venom system in the advanced snakes (Caenophidia). Molecular & Cellular Proteomics, 7(2), 215-246.

Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28, 3150–3152.

Gans, C. (1968). Relative success of divergent pathways in amphisbaenian specialization. The American Naturalist, 102(926), 345-362.

Gans, C. (1974). Biomechanics. Approach to vertebrate biology. University of Michigan Press, Ann Arbor.

Giorgianni, M. W., Dowell, N. L., Griffin, S., Kassner, V. A., Selegue, J. E., & Carroll, S. B. (2020). The origin and diversification of a novel protein family in venomous snakes. Proceedings of the National Academy of Sciences, 117(20), 10911-10920.

Grazziotin, F. G., Zaher, H., Murphy, R. W., Scrocchi, G., Benavides, M. A., Zhang, Y. P., & Bonatto, S. L. (2012). Molecular phylogeny of the new world Dipsadidae (Serpentes: Colubroidea): a reappraisal. *Cladistics*, *28*(5), 437-459.

Greene, H. W. (1983). Dietary correlates of the origin and radiation of snakes. American Zoologist, 23(2), 431-441.

Guedes, T. B., Sawaya, R. J., Zizka, A., Laffan, S., Faurby, S., Pyron, R. A., ... & Antonelli, A. (2018). Patterns, biases and prospects in the distribution and diversity of Neotropical snakes. Global Ecology and Biogeography, 27(1), 14-21.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature Protocols 8, 1494–1512.

He, Y. Y., S. B. Liu, W. H. Lee, J. Q. Qian, Y. Zhang. 2008. Isolation, expression and characterization of a novel dual serine protease inhibitor, OH-TCI, from king cobra venom. Peptides. 29(10):1692–99.

Helfrich, P., Rieb, E., Abrami, G., Lücking, A., & Mehler, A. (2018). TreeAnnotator: versatile visual annotation of hierarchical text relations. Proceedings of the eleventh international conference on language resources and evaluation.

Heptinstall, T. C., Strickland, J. L., Rosales-Garcia, R. A., Rautsaw, R. M., Simpson, C. L., Nystrom, G. S., ... & Parkinson, C. L. (2023). Venom phenotype conservation suggests integrated specialization in a lizard-eating snake. Toxicon, 229, 107135.

Herrera, C., T. Escalante, M.-B. Voisin, A. Rucavado, D. Morazán, J.K.A. Macêdo, J.J. Calvete, L. Sanz, S. Nourshargh, J.M. Gutiérrez, J.W. Fox. (2015). Tissue localization and extracellular matrix degradation by PI, PII and PIII snake venom metalloproteinases: clues on the mechanisms of venom-induced hemorrhage. PLoS Neglected Tropical Diseases, 9:e0003731.

Hickman, C. P. (2020). Integrated principles of zoology. McGraw-Hill Education.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, G., Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 1978(15):1965–1978.

Hill, R.E., Mackessy, S.P., 1997. Venom yields from several species of colubrid snakes and differential effects of ketamine. Toxicon 35, 671–678.

Hofmann, E. P., Rautsaw, R. M., Strickland, J. L., Holding, M. L., Hogan, M. P., Mason, A. J., ... & Parkinson, C. L. (2018). Comparative venom-gland transcriptomics and venom proteomics

of four Sidewinder Rattlesnake (*Crotalus cerastes*) lineages reveal little differential expression despite individual variation. Scientific reports, 8(1), 1-15.

Holding, M.L., Margres, M.J., Mason, A.J., Parkinson, C.L., Rokyta, D.R., 2018. Evaluating the performance of de novo assembly methods for venom-gland transcriptomics. Toxins (Basel). 10, 249.

Holding, M. L., Strickland, J. L., Rautsaw, R. M., Hofmann, E. P., Mason, A. J., Hogan, M. P., ... & Parkinson, C. L. (2021). Phylogenetically diverse diets favor more complex venoms in North American pitvipers. Proceedings of the National Academy of Sciences, 118(17), e2015579118.

Holding, M. L., Trevine, V. C., Zinenko, O., Strickland, J. L., Rautsaw, R. M., Mason, A. J., ... & Rokyta, D. R. (2022). Evolutionary allometry and ecological correlates of fang length evolution in vipers. Proceedings of the Royal Society B, 289(1982), 20221132.

Jackson, K. (2003). The evolution of venom-delivery systems in snakes. Zoological Journal of the Linnean Society, 137(3), 337-354.

Jackson, T. N., Young, B., Underwood, G., McCarthy, C. J., Kochva, E., Vidal, N., ... & Fry, B. G. (2017). Endless forms most beautiful: the evolution of ophidian oral glands, including the venom system, and the use of appropriate terminology for homologous structures. Zoomorphology, 136, 107-130.

Junqueira-de-Azevedo, I. L., Campos, P. F., Ching, A. T., & Mackessy, S. P. (2016). Colubrid venom composition: an-omics perspective. Toxins, 8(8), 230.

Kardong, K. V. (1980). Evolutionary patterns in advanced snakes. American Zoologist, 20(1), 269-282.

Kardong, K. V. (1986). Predatory strike behavior of the rattlesnake, *Crotalus viridis oreganus*. Journal of Comparative Psychology, 100(3), 304.

Kasturiratne, A., Wickremasinghe, A. R., de Silva, N., Gunawardena, N. K., Pathmeswaran, A., Premaratna, R., ... & de Silva, H. J. (2008). The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. PLoS medicine, 5(11), e218.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

Kochva, E. (1987). The origin of snakes and evolution of the venom apparatus. Toxicon, 25(1), 65-106.

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution, 35(6), 1547.

Lema, T. (1984) Sobre o gênero *Elapomorphus* Wiegmann, 1843 (Serpentes, Colubridae, Elapomorphinae). Iheringia Série Zoologia, 64, 53−86.

Lema, T. (2007). Report of a Human Ophidic Accident by Phalotris trilineatus (Snakes, Colubridae) on Southern Cost of Brazil. Caderno de Pesquisa Série Biologia, 19(2):6-16.

Liu, C. S., T. C. Wu, T. B. Lo. 1983. Complete amino acid sequences of two protease inhibitors in the venom of Bungarus fasciatus. Int. J. Pept. Protein Res. 21:209–15.

Maritz, B., Barends, J. M., Mohamed, R., Maritz, R. A., & Alexander, G. J. (2021). Repeated dietary shifts in elapid snakes (Squamata: Elapidae) revealed by ancestral state reconstruction. Biological Journal of the Linnean Society, 134(4), 975-986.

Marques, O.A.V., Eterovic, A., Nogueira, C.C., Sazima, I. (2015). Serpentes do Cerrado: Guia Ilustrado, First Edition. Ribeirão Preto, São Paulo, Brazil, Holos Press.

Marques, O.A.V., Pereira, D.N., Barbo, F.E., Germano, V.J., Sawaya, R.J. (2009). Reptiles in São Paulo municipality: diversity and ecology of the past and present fauna. Biota Neotropica 9(2): 139–150.

Mason, A. J., Margres, M. J., Strickland, J. L., Rokyta, D. R., Sasa, M., & Parkinson, C. L. (2020). Trait differentiation and modular toxin expression in palm-pitvipers. BMC genomics, 21, 1-20.

McGivern, J. J., Wray, K. P., Margres, M. J., Couch, M. E., Mackessy, S. P., & Rokyta, D. R. (2014). RNA-seq and high-definition mass spectrometry reveal the complex and divergent venoms of two rear-fanged colubrid snakes. BMC genomics, 15(1), 1-18.

Modahl, C. M., Brahma, R. K., Koh, C. Y., Shioi, N., & Kini, R. M. (2020). Omics technologies for profiling toxin diversity and evolution in snake venom: Impacts on the discovery of therapeutic and diagnostic agents. Annual review of animal biosciences, 8, 91-116.

Nachtigall, P. G., Grazziotin, F. G., & Junqueira-de-Azevedo, I. L. (2021). MITGARD: an automated pipeline for mitochondrial genome assembly in eukaryotic species using RNA-seq data. Briefings in Bioinformatics, 22(5), bbaa429.

Nachtigall, P. G., Rautsaw, R. M., Ellsworth, S. A., Mason, A. J., Rokyta, D. R., Parkinson, C. L., & Junqueira-de-Azevedo, I. L. (2021). ToxCodAn: a new toxin annotator and guide to venom gland transcriptomics. Briefings in Bioinformatics, 22(5), 1-16.

Nachtigall, P. G., Freitas-de-Sousa, L. A., Mason, A. J., Moura-da-Silva, A. M., Grazziotin, F. G., & Junqueira-de-Azevedo, I. L. (2022). Differences in PLA2 constitution distinguish the

venom of two endemic brazilian mountain lanceheads, *Bothrops cotiara* and *Bothrops fonsecai*. Toxins, 14(4), 237.

Negrin, A., Morais, V., Carreira, S., & Tortorella, M. N. (2019). Mordedura de *Phalotris lemniscatus (*Duméril, Bibron & Duméril, 1854) (Squamata, Dipsadidae) en Uruguay. Acta toxicológica argentina, 27(2), 65-71.

Nogueira, C. C., Argôlo, A. J., Arzamendia, V., Azevedo, J. A., Barbo, F. E., Bérnils, R. S., ... & Martins, M. (2019). Atlas of Brazilian snakes: verified point-locality maps to mitigate the Wallacean shortfall in a megadiverse snake fauna. South American Journal of Herpetology, 14(sp1), 1-274.

Oliveira, L. D., Grazziotin, F. G., Sánchez-Martínez, P. M., Sasa, M., Flores-Villela, O., Prudente, A. L. D. C., & Zaher, H. (2023). Phylogenetic and morphological evidence reveals the association between diet and the evolution of the venom delivery system in Neotropical goo-eating snakes. *Systematics and Biodiversity*, *21*(1), 2153944.

Patthy, L. (1999). Genome evolution and the evolution of exon-shuffling—a review. Gene, 238(1), 103-114.

Phillips, S.J., Anderson, R.P., Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190(3-4):231–259.

Pla, D., Sanz, L., Whiteley, G., Wagstaff, S. C., Harrison, R. A., Casewell, N. R., & Calvete, J. J. (2017). What killed Karl Patterson Schmidt? Combined venom gland transcriptomic, venomic and antivenomic analysis of the South African green tree snake (the boomslang), *Dispholidus typus*. Biochimica et Biophysica Acta (BBA)-General Subjects, 1861(4), 814-823.

Portillo, F., Stanley, E. L., Branch, W. R., Conradie, W., Rödel, M. O., Penner, J., ... & Greenbaum, E. (2019). Evolutionary history of burrowing asps (Lamprophiidae: Atractaspidinae) with emphasis on fang evolution and prey selection. PLoS One, 14(4), e0214889.

Powell, R. L., Reyes, S. R., & Lannutti, D. I. (2006). Molecular barcoding, DNA from snake venom, and toxinological research: Considerations and concerns. Toxicon, 48(8), 1095-1097.

Puorto, G. & França, F. O. S. (2009). Serpentes não peçonhentas e aspectos clínicos dos acidentes. Pp. 124-129. In: Cardoso, J. L. C., França, F. O. S., Wen, F. H., Maláque, C. M. S., Haddad-Jr., V. Animais peçonhentos no Brasil: Biologia, Clínica e Terapêutica dos Acidentes. 2 Ed. Sarvier.

Puorto, G., & Ferrarezzi, H. (1993). Uma nova espécie de *Phalotris* Cope, 1862, com comentários sobre o grupo *bilineatus* (Serpentes: Colubridae: Xenodontinae). Memórias do Instituto Butantan, 55(1), 39-46.

Trevine, V. C., Grazziotin, F. G., Giraudo, A., Sallaberry‐Pincheira, N., Vianna, J. A., & Zaher, H. (2022). The systematics of Tachymenini (Serpentes, Dipsadidae): An updated classification based on molecular and morphological evidence. Zoologica Scripta, 51(6), 643-663.

Tioyama, E. C., Bayona-Serrano, J. D., Portes-Junior, J. A., Nachtigall, P. G., de Souza, V. C., Beraldo-Neto, E., ... & Freitas-de-Sousa, L. A. (2023). The venom composition of the snake tribe Philodryadini: ‘Omic’ techniques reveal intergeneric variability among South American racers. Toxins, 15(7), 415.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology, 67(5), 901-904.

Rokyta, D.R., Lemmon, A.R., Margres, M.J., Aronow, K., 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). BMC Genomics 13, 1–23.

Rokyta, D.R., Margres, M.J., Ward, M.J., Sanchez, E.E., 2017. The genetics of venom ontogeny in the Eastern Diamondback Rattlesnake (*Crotalus adamanteus*). PeerJ 5, e3249.

Rosenberg, H. I. (1992). An improved method for collecting secretion from Duvernoy's gland of colubrid snakes. Copeia, 1992(1), 244-246.

Salomão, M.D.G., Albolea, A.B.P, Santos, S.M.A., (2003). Colubrid snakebite: a public health problem in Brazil. Herpetological Review 34, 307–312.

Savitzky, A. H. (1980). The role of venom delivery strategies in snake evolution. Evolution, 1194-1204.

Schlegel, H., (1828). Untersuchung der Speicheldrusen bei den Schlangen mit gefurchten Zahnen, im Vergleich mit denen der Giftlosen und Giftigen. Nova Acta Acad. Caes. Leopoldino-Carolinae Nat. Curiosorum vol. 14

Scrocchi, G. J., Giraudo, A. R., & Nenda, S. J. (2022). Taxonomic notes on the *Phalotris bilineatus* group (Serpentes: Dipsadidae: Elapomorphini), with the description of a new species from northwestern Argentina. Cuadernos de Herpetología, 36(1), 47-63.

Silva, A., Hifumi, T., Sakai, A., Yamamoto, A., Murakawa, M., Ato, M., ... & Kuroda, Y. (2014). *Rhabdophis tigrinus* is not a pit viper but its bites result in venom-induced consumptive coagulopathy similar to many viper bites. Journal of intensive care, 2(1), 1-3.

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210-3212.

Smeets, R. E. H., Melman, P. G., Hoffmann, J. J. M. L., & Mulder, A. W. (1991). Severe coagulopathy after a bite from a 'harmless' snake (*Rhabdophis subminiatus*). Journal of internal medicine, 230(4), 351-354.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313

Suvilesh, K. N., M. Yariswamy, M. N. Savitha, V. Joshi, A. N. Nanjaraj Urs, A. P. Urs, M. Choudhury, D. Velmurugan, B. S. Vishwanath. 2017. Purification and characterization of an anti-hemorrhagic protein from *Naja naja* (Indian cobra) venom. Toxicon 140:83–93.

Terrat, Y., Sunagar, K., Fry, B. G., Jackson, T. N., Scheib, H., Fourmy, R., ... & Ducancel, F. (2013). *Atractaspis aterrima* toxins: the first insight into the molecular evolution of venom in side-stabbers. Toxins, 5(11), 1948-1964.

Tioyama, E. C., Bayona-Serrano, J. D., Portes-Junior, J. A., Nachtigall, P. G., de Souza, V. C., Beraldo-Neto, E., ... & Freitas-de-Sousa, L. A. (2023). The venom composition of the snake tribe Philodryadini:'Omic'techniques reveal intergeneric variability among South American racers. Toxins, 15(7), 415.

Uetz, P., Freed, P, Aguilar, R. & Hošek, J. (eds.) (2021) The Reptile Database, http://www.reptile-database.org, accessed 30 December 2021.

Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi‐gene datasets with character set and codon information. Cladistics, 27(2), 171-180.

Vidal, N. (2002). Colubroid systematics: evidence for an early appearance of the venom apparatus followed by extensive evolutionary tinkering. Journal of Toxicology: Toxin Reviews, 21(1-2), 21-41.

Vivas, J., C. Ibarra, A. M. Salazar, A. G. C. Neves-Ferreira, E. E. Sánchez, J. Perales, A. Rodríguez-Acosta, B. Guerrero. 2016. Purification and characterization of tenerplasminin-1, a serine peptidase inhibitor with antiplasmin activity from the coral snake (Micrurus tener tener) venom. Comp. Biochem. Physiol. C Toxicol. Pharmacol. 179:107–15.

Weinstein, S. A., Warrell, D. A., & Keyler, D. E. (2011). "Venomous Bites from Non-Venomous Snakes: A Critical Analysis of Risk and Management of "Colubrid Snake Bites. Elsevier Insights.

Weinstein, S.A., Kardong, K.V., 1994. Properties of Duvernoy's secretions from opisthoglyphous and aglyphous colubrid snakes. Toxicon 32, 1161–1185.

Wüster, W., & Thorpe, R.S. (1989). Population affinities of the Asiatic cobra (Naja naja) species complex in south-east Asia: reliability and random resampling. Biological Journal of the Linnean Society, 36(4), 391-409.

Zhang, C., Sayyari, E., & Mirarab, S. (2017). ASTRAL-III: increased scalability and impacts of contracting low support branches. RECOMB International Workshop on Comparative Genomics, 2017, 53–75.

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30, 614–620.

Supplementary Files

Entiauspe-Neto, O.M., Nachtigall, P.G., Borges-Martins, M., Junqueira-de-Azevedo, I., Grazziotin, F.G. (2023) Supplemental Material from: "*Highly conserved and extremely variable: the paradoxical pattern of toxin expression revealed by comparative venom-gland transcriptomics of* Phalotris *(Serpentes: Dipsadidae)*" Figshare Dataset (under embargo until 20 January 2023).

Supplementary Table 1. Overview of transcriptome results for *Phalotris reticulatus* (.tsv). [Link: https://drive.google.com/file/d/1Rio_uUPFttUfFiCtAUD7kT1B7fhlNCQi/view?usp=sharing].

Supplementary Table 2. Overview of transcriptome results for *Phalotris lemniscatus* (.tsv). [Link: https://drive.google.com/file/d/10mChZAbjYCrjZ-EktmCAv2ahKyqvcGDl/view?usp=sharing].

Supplementary Table 3. Overview of transcriptome results for *Phalotris mertensi* (.tsv). [Link: https://drive.google.com/file/d/1UwFV_0NIIMdXjMIKPCj74OQcBjcFIySH/view?usp=drive_link]

Figures

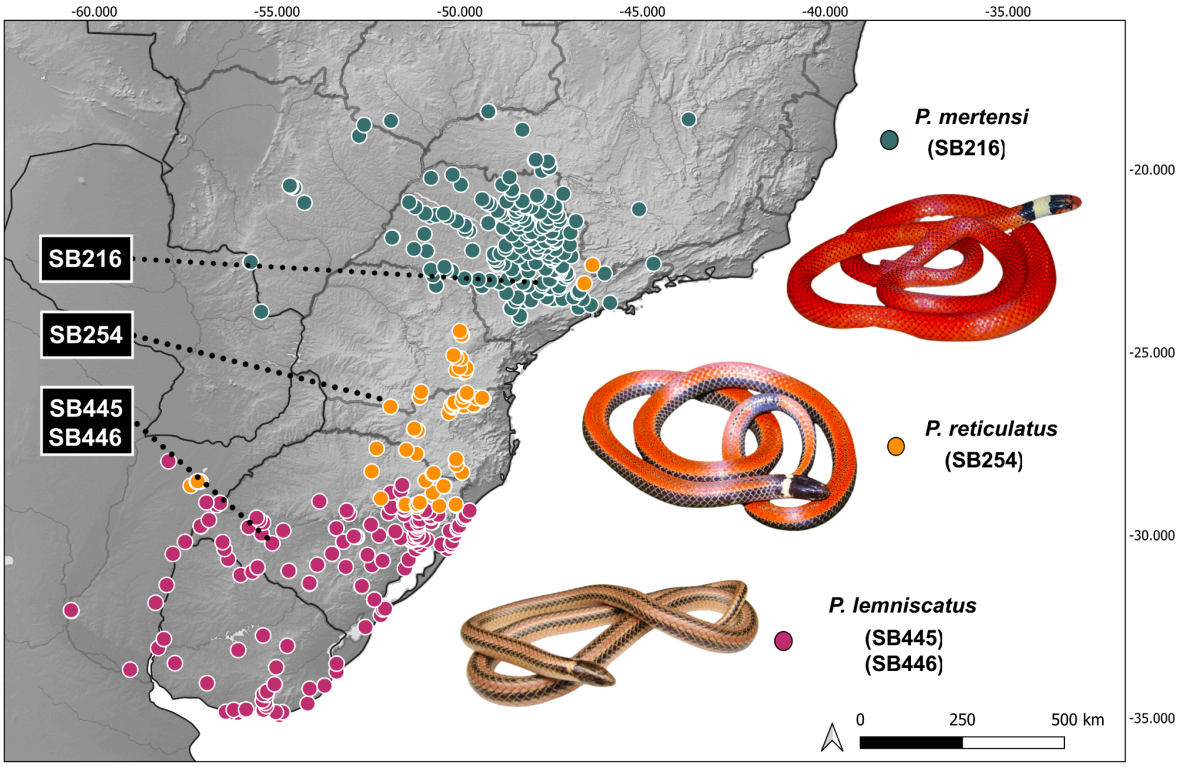
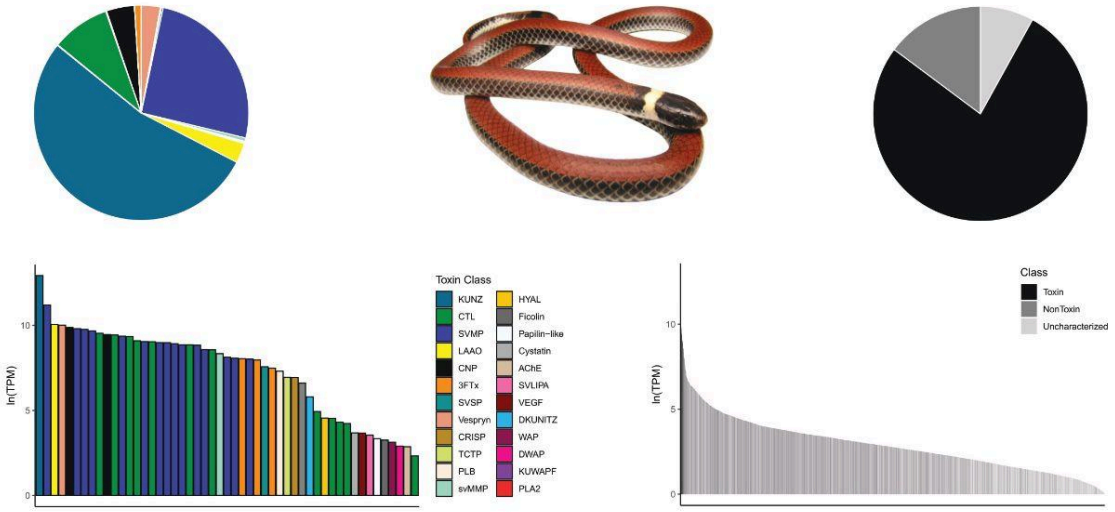
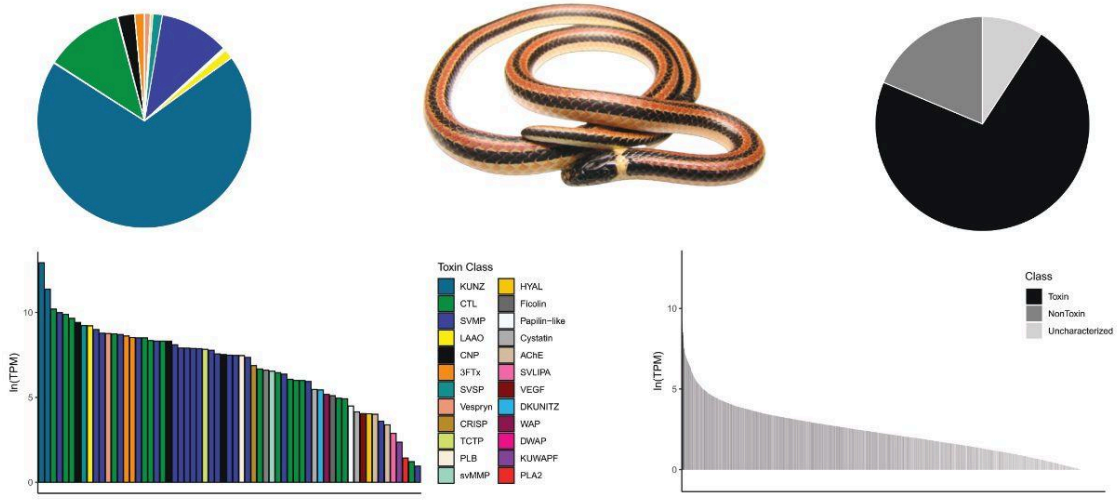


Figure 1. Sampling and approximate geographic distribution of evaluated *Phalotris* species in South America: *Phalotris mertensi* (SB0216, blue), Cerrado, *Phalotris reticulatus* (SB0254, orange), Araucaria Pine forests, *Phalotris lemniscatus* (SB0445, SB0446, purple), Pampa Grasslands. Circles indicate examined specimens. Inset photographs: Arthur Abegg (*P. reticulatus*), Márcio Borges-Martins (*P. lemniscatus*), and Marcelo Duarte (*P. mertensi*). [Print in color]

P. reticulatus



P. lemniscatus



P. mertensi

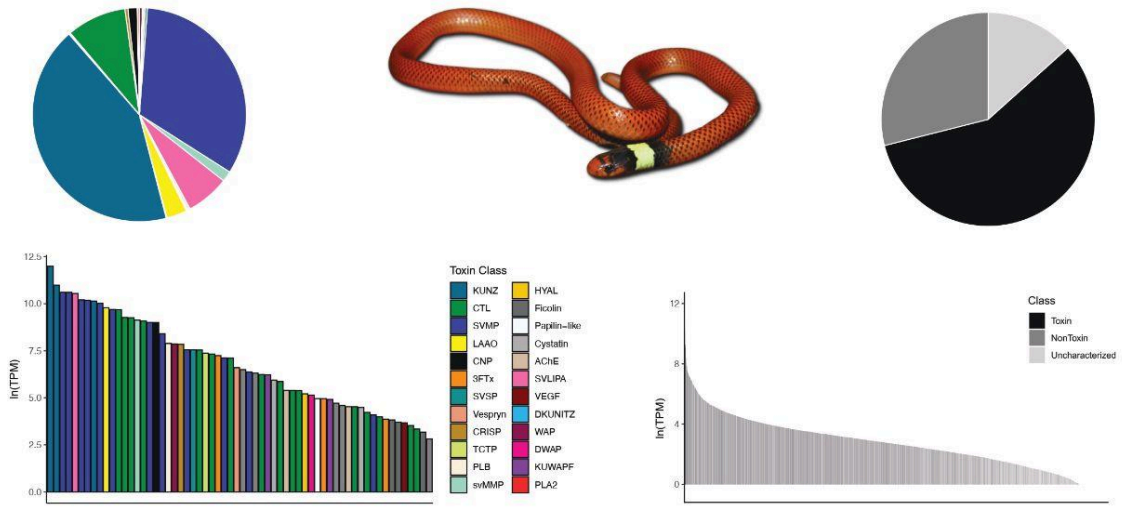


Figure 2. Expression of the venom gland transcriptome of *Phalotris lemniscatus* (SB445, 446), *Phalotris reticulatus* (SB0254), and *Phalotris mertensi* (SB216). Upper right: Pie chart of proportional toxin gene expression by toxin family; Bottom right: Bar chart of each toxin transcript; Upper left: Pie chart of proportional toxin, nontoxin, and uncharacterized transcripts; Bottom left: toxin and nontoxin gene expression in the venom's gland. Inset photographs: Márcio Borges-Martins (*P. lemniscatus*, *P. reticulatus*), Michel Alves Passos (*P. mertensi*).
[Print in color]

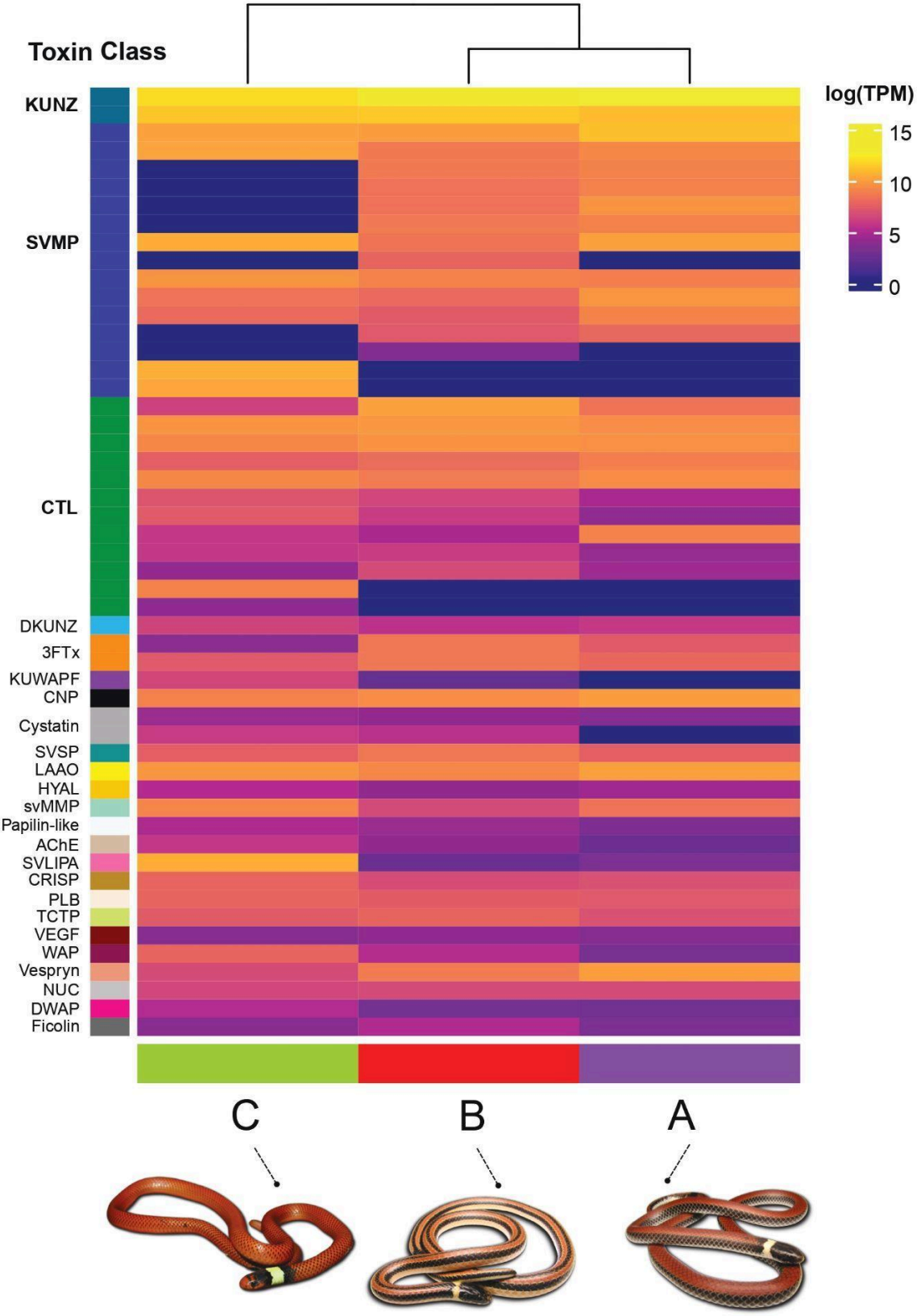

Toxin Class
KUNZ
SVMP
CTL
DKUNZ
3FTx
KUWAPF
CNP
Cystatin
SVSP
LAAO
HYAL
svMMP
Papilin-like
AChE
SVLIPA
CRISP
PLB
TCTP
VEGF
WAP
Vespryn
NUC
DWAP
Ficolin
log(TPM)
15
10
5
0
C
B
A

Figure 3. Heatmap showing the toxin expression level in the venom gland transcriptome of *Phalotris reticulatus* (A), *Phalotris lemniscatus* (B), and *Phalotris mertensi* (C) species. The expression level is shown in log-scale of the TPM value (Transcripts per Million). Toxins are clustered per class (left), whereas samples were clustered by expression profile of toxins (top). Inset photographs: Márcio Borges-Martins, Diego J. Alvares, Michel Alves Passos.

[Print in color]

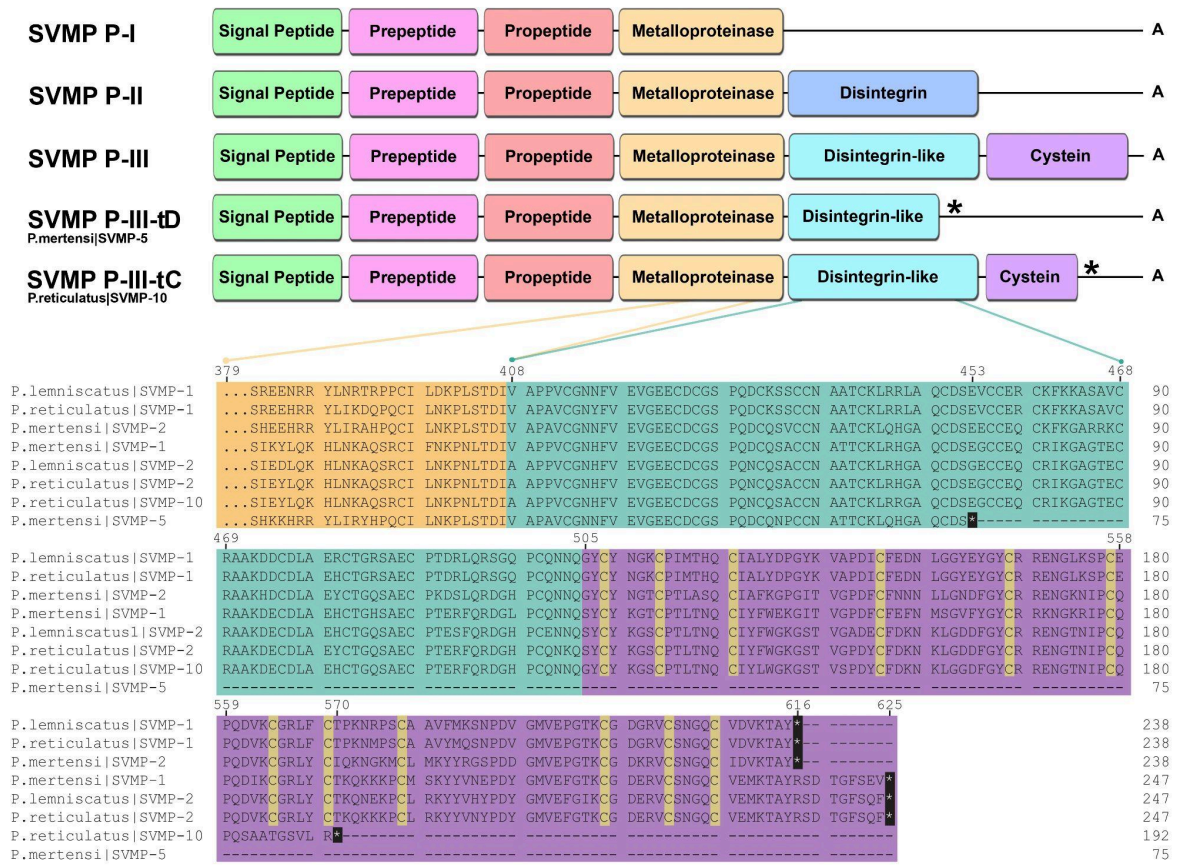
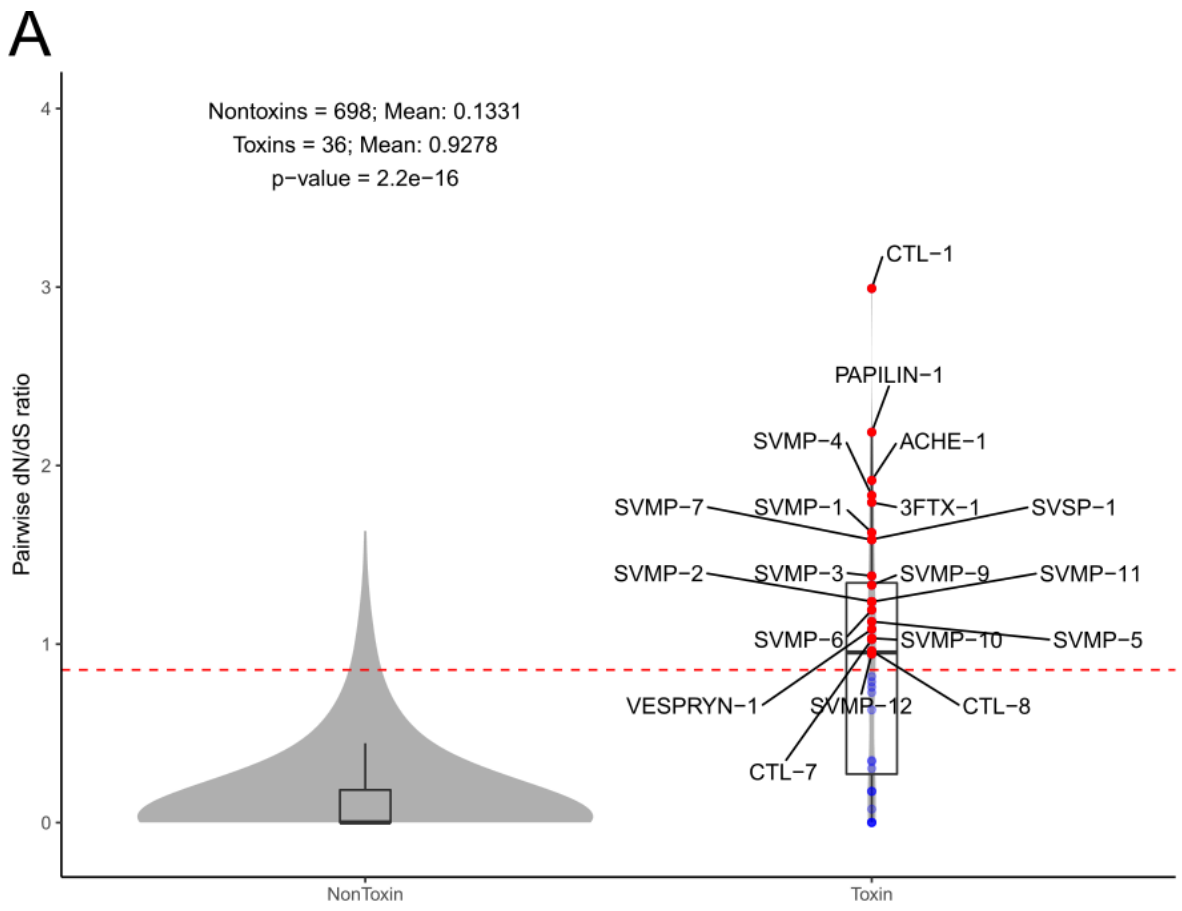


Figure 4. Snake venom metalloproteinases (SVMPs) detected in *Phalotris* venom. Top: Schematic organization of SVMP domain types following Dawson et al. (2021), the "truncated" SVMP type III of Campos et al. (2016) and also detected in the present work (herein named as SVMP-III tD), and the truncated SVMP type III identified in the present work (herein named as

SVMP-III tC); Bottom: Alignment showing in detail the catalytic domain and disintegrin-like domain sequences of the *Phalotris* SVMP toxins identified in the present work.

[Print in grayscale]


A
Nontoxins = 698; Mean: 0.1331
Toxins = 36; Mean: 0.9278
p−value = 2.2e−16
Pairwise dN/dS ratio
4
3
2
1
0
CTL−1
PAPILIN−1
SVMP−4
ACHE−1
SVMP−7
SVMP−1
3FTX−1
SVSP−1
SVMP−2
SVMP−3
SVMP−9
SVMP−11
SVMP−6
SVMP−10
SVMP−5
VESPRYN−1
SVMP−12
CTL−8
CTL−7
NonToxin
Toxin

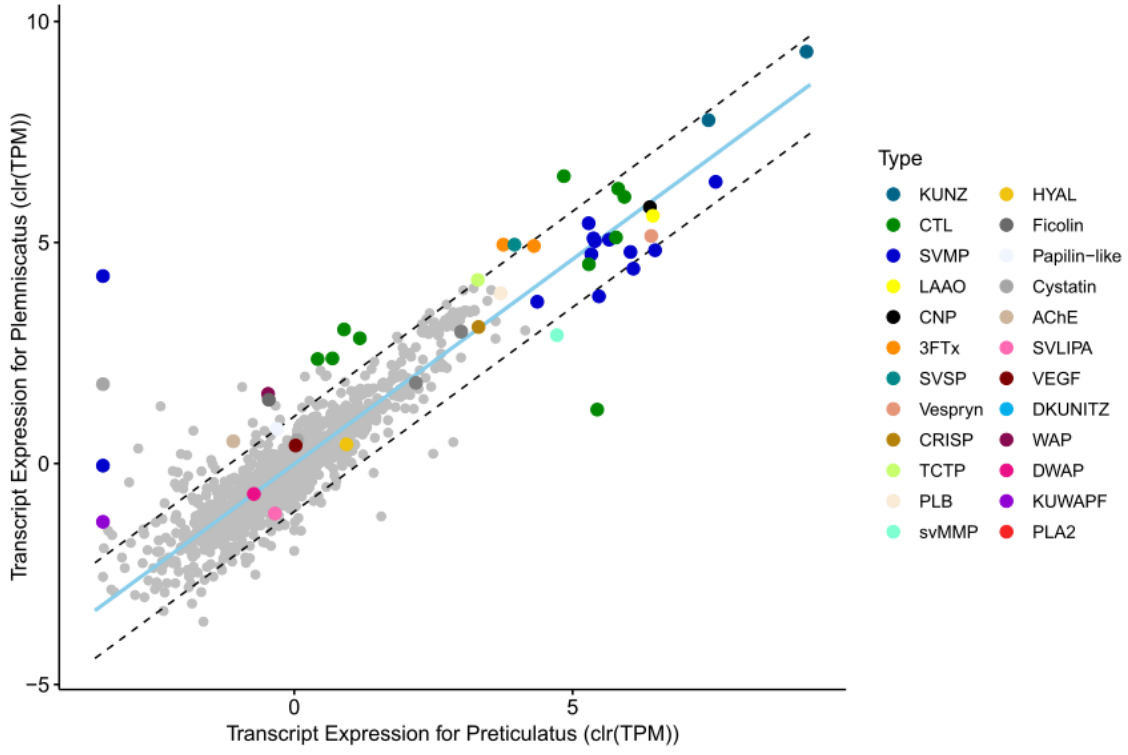
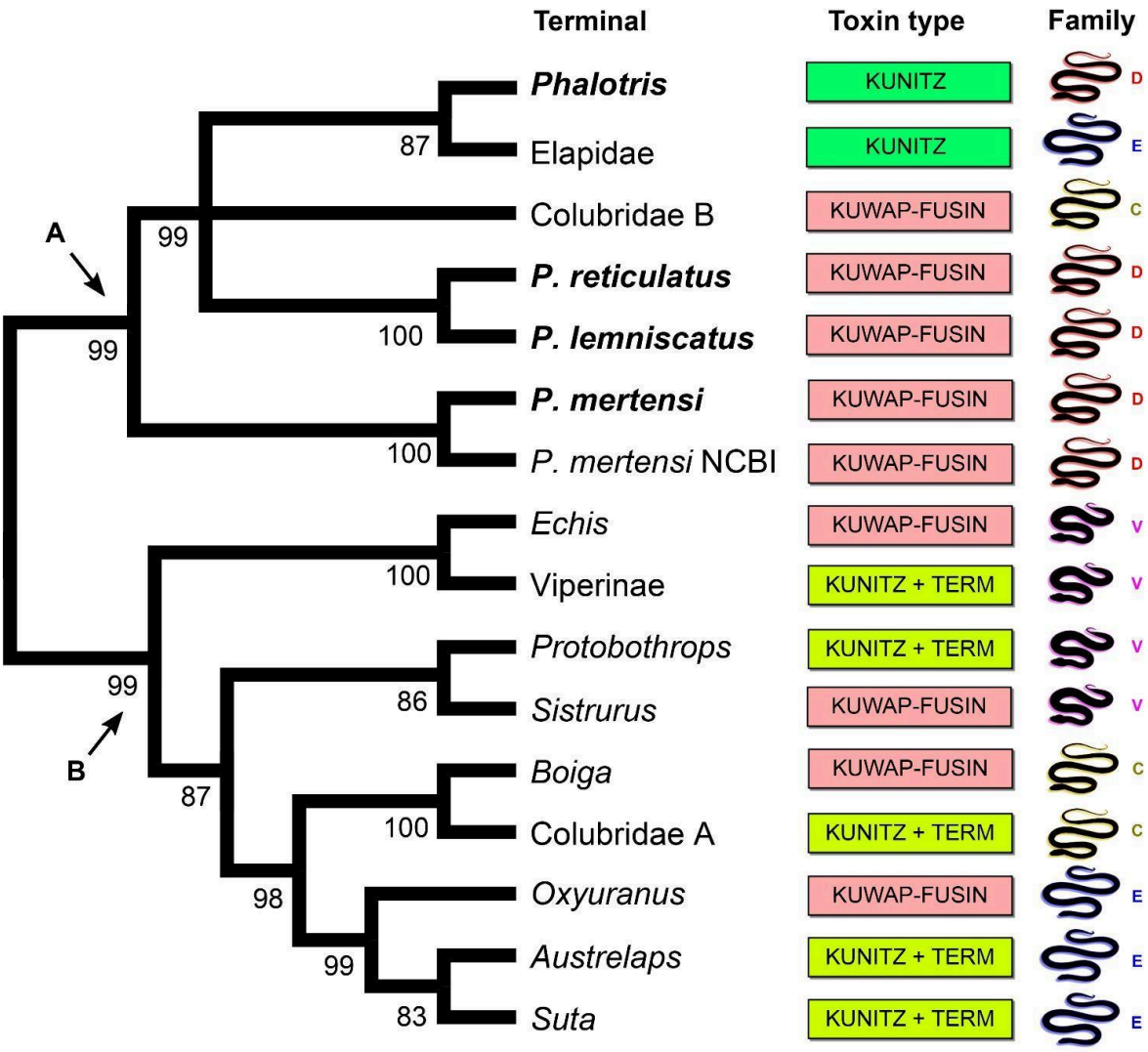

B
Transcript Expression for Plemniscatus (clr(TPM))
10
5
0
−5
0
5
Transcript Expression for Preticulatus (clr(TPM))
Type
KUNZ
CTL
SVMP
LAAO
CNP
3FTx
SVSP
Vespryn
CRISP
TCTP
PLB
svMMP
HYAL
Ficolin
Papilin−like
Cystatin
AChE
SVLIPA
VEGF
DKUNITZ
WAP
DWAP
KUWAPF
PLA2

Figure 5. Distribution of pairwise dN/dS (number of nonsynonymous substitutions per non-synonymous site [pN] to the number of synonymous substitutions per synonymous site [pS]) ratios and nontoxins orthologous coding sequences (A) and scatterplots of interspecific variation in transcription expression (B) of the venom gland transcriptome of *Phalotris lemniscatus* and *Phalotris reticulatus*, in TPM (Transcripts Per Million). For all plots, the data was centered-log transformed to account for their compositional nature. Dashed lines in scatter plots denote the 99% confidence interval of nontoxin expression and the light blue line shows the line of best fit based on orthogonal residuals.

[Print in grayscale]


Terminal
Toxin type
Family
Phalotris
Elapidae
Colubridae B
P. reticulatus
P. lemniscatus
P. mertensi
P. mertensi NCBI
Echis
Viperinae
Protobothrops
Sistrurus
Boiga
Colubridae A
Oxyuranus
Austrelaps
Suta
KUNITZ
KUNITZ
KUWAP-FUSIN
KUWAP-FUSIN
KUWAP-FUSIN
KUWAP-FUSIN
KUWAP-FUSIN
KUNITZ + TERM
KUNITZ + TERM
KUWAP-FUSIN
KUWAP-FUSIN
KUNITZ + TERM
KUWAP-FUSIN
KUNITZ + TERM
KUNITZ + TERM
D
E
C
D
D
D
D
V
V
V
V
C
C
E
E
E
A
B
99
99
100
100
100
99
86
87
100
98
99
83
87

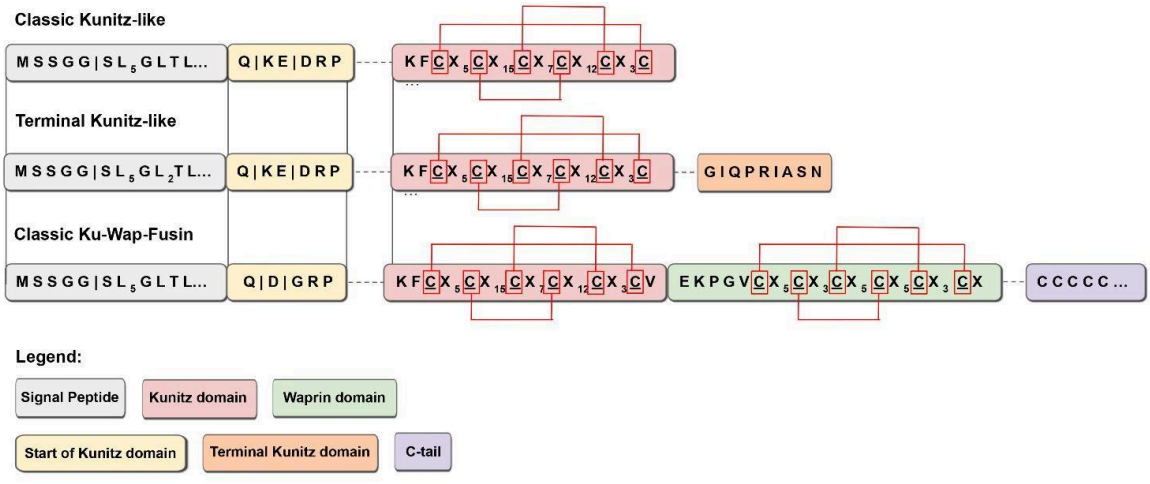

Classic Kunitz-like
MSSGG|SL5GLTL...
Q|KE|DRP
KFCX5CX18CX7CX12CX3C
Terminal Kunitz-like
MSSGG|SL5GL2TL...
Q|KE|DRP
KFCX5CX18CX7CX12CX3C
GIQPRIASN
Classic Ku-Wap-Fusin
MSSGG|SL5GLTL...
Q|D|GRP
KFCX5CX18CX7CX12CX3CV
EKPGVCX5CX3CX5CX5CX3CX
CCCCC...
Legend:
Signal Peptide
Kunitz domain
Waprin domain
Start of Kunitz domain
Terminal Kunitz domain
C-tail

Figure 6. Phylogenetic tree of Kunitz-type inhibitor toxin family evolution among Caenophid snakes, as inferred using maximum likelihood framework (top) and a schematic representation of the amino acid composition of major kunitz-type inhibitor toxin types (bottom). Inset figures refer to snake families, sequences in bold were generated in this study, and bootstrap values are available in each node.

[Print in grayscale]

TABLES

Table 1. Specimens of *Phalotris* used in this study for transcriptome sequencing and associated data.

Field #	Voucher	OTU	Sex	SVL (mm)	Locality	Collection date	SRA Accession
SB0445	IBSP 90137	B.1	Male	257	Rosário do Sul, RS, Brazil	11/11/2017	[INSERT]
SB0446	IBSP 90138	B.2	Female	404	Rosário do Sul, RS, Brazil	11/13/2017	[INSERT]
SB0254	IBSP 89723	A.1	Female	610	Água Doce, SC, Brazil	1/18/2018	[INSERT]
SB0216	IBSP 89655	C.1	Male	387	Rio Claro, SP, Brazil	13/12/2017	[INSERT]

ARTIGO II

Testing species boundaries within the *Phalotris bilineatus* (Serpentes: Dipsadidae: Xenodontinae) species group

Vertebrate Zoology

Testing species boundaries within the *Phalotris bilineatus* (Serpentes: Dipsadidae: Xenodontinae) species group

Omar M. Entiauspe-Neto[1],[2]; Fabiane de Oliveira Noronha[1]; Claudia Koch[3]; Arthur Tiutenko[4]; Jayme Massim[5]; Felipe G. Grazziotin[2]; Márcio Borges-Martins[1]

[1] Programa de Pós-graduação em Biologia Animal, Departamento de Zoologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, CEP 91501-970, Porto Alegre, RS, Brazil.

[2] Laboratório de Coleções Zoológicas, Instituto Butantan , 05503-90, Av. Vital Brazil, 1500, Butantã, São Paulo, SP, Brazil.

[3] Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 160, 53113, Bonn, Germany

[4] Friedrich-Alexander-Universität Erlangen-Nürnberg, Schloßplatz 4, 91054, Erlangen, Germany.

[5] Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga, 6681, 90619-900, Porto Alegre, RS, Brazil.

Abstract: *Phalotris* is a speciose genus of small to medium-sized Dipsadid snakes, encompassing 19 species that occur in open areas from northern Brazil to southern Argentina. The *Phalotris bilineatus* species group contains at least 12 available nomina, with an extensive taxonomic history, and disagreements among authors regarding the validity of its species. In this work, we provide an integrative revision for the *P. bilineatus* species group, reviewing available vouchered specimens and designating operational taxonomic units (OTUs), which were examined for external and internal morphology, as well as phylogenetic relationships based on molecular data. Our inferred phylogenetic and haplotype network inferences recover widespread paraphyletism among *P. bilineatus*, *P. lemniscatus*, *P. trilineatus* and *P. reticulatus*, with clades corresponding to geographical regions, rather than morphological species. Our morphological multivariate analyses also failed to distinguish among species of the *P. bilineatus* species group. We propose allocating *P. trilineatus* and *P. reticulatus* as junior synonyms of *P. lemniscatus*, and provide comments into other species of the *P. bilineatus* group.

Keywords: Caenophidia; Taxonomy; Molecular Phylogeny; Systematics.

Introduction

The delimitation and testing of species remains one of the paramount objectives for biologists, with deep implications on our understanding of biodiversity. Historically, several concepts and criteria have been used for species conceptualization and delimitation, with increasingly complex methodology and applications, which have contributed to both significant advances in reducing subjectivity in description of new taxa, as well as deleterious effects, with persistent confusion between both concepts of delimitation and conceptualization. This had led to the proliferation of

poorly defined species, dissatisfactory taxonomic applications, and incorrect decisions (Mayden, 1997; de Queiroz, 2007; Entiauspe-Neto et al. 2019).

The unified species concept, for which de Queiroz (2007) considers species as a separately evolving metapopulation lineage with relevant diagnostic characters, has been widely used in recent taxonomic applications; it has also been associated with the proposal of an integrative taxonomy species delimitation practice, for which Dayrat (2005) suggests the combination of several lines of evidence, such as comparative morphology, population genetics, ecology, and behavior, in order to reduce subjectivity in the delimitation of species. Together, these two concepts would (ideally) be employed to recognize species as an evolutionary lineage, with cohesion in diagnostic characters and lines of evidence, as well as recent common ancestry, in order to avoid taxonomic instability (Kaiser et al. 2013; Freitas et al. 2020; Maddison et al. 2023).

The Elapomorphini tribe, which encompasses approximately 45 species of fossorial and cryptozoic Dipsadid snake species, has been subject to long taxonomic instability and controversy (Ferrarezzi 1993; Harvey 1999; Entiauspe-Neto et al. 2019, 2020a, 2021). Five genera are currently known, *Apostolepis* Cope, 1861, *Coronelaps* Lema and Hofstadler-Deiques, 2010, *Elapomorphus* Wiegmann (in Fitzinger), 1843, *Parapostolepis* Amaral, 1930, and *Phalotris* Cope, 1862, which share as putative synapomorphies an frontoparietal suture in U-shape, dentigerous process of the dentary short, frontal bones dorsally included by the anterolateral process of the parietal, few supralabial scales (5–7), and second supralabial contacting the eye, undivided nasal plate, and other morphological characters (Savitzky 1979; Ferrarezzi 1993, 1994; Zaher 1994; Zaher et al. 2009). These genera have been historically

diagnosed based largely on morphology and have not yet been tested under a robust molecular phylogenetic framework (for a summary of generic diagnoses, see Entiauspe-Neto et al. 2021a).

The genus *Phalotris* contains 19 small to medium sized species, that range from northeastern Brazil and Bolivia into Argentina (Ferrarezzi et al. 1993; Entiauspe-Neto et al. 2021; Uetz et al. 2023). These species have been allocated in three groups, proposed by Ferrarezzi (1993) in an unpublished dissertation: *P. bilineatus* group, diagnosed by wide dark dorsal stripes and black ventrals; *P. nasutus* group, diagnosed by a projected rostral scale, second and third temporals possibly fused; *P. tricolor* group, diagnosed by uniformly colored dorsal pattern and wide nuchal collars. Of these, the *P. bilineatus* group likely has the most extensive and complicated taxonomic history, with five species usually recognized until recently: *Phalotris bilineatus* Duméril, Bibrón, & Duméril, 1854, *Phalotris lemniscatus* Duméril, Bibrón, & Duméril, 1854, *Phalotris multipunctatus* Ferrarezzi & Puorto, 1993, *Phalotris reticulatus* (Peters, 1860), and *Phalotris normanscotti* Cabral & Cacciali, 2015 (Ferrarezzi & Puorto, 1993; Giraudo & Scrocchi, 2002; Cabral & Cacciali, 2015; Nogueira et al., 2019; Guedes et al., 2023). Conversely, alternative taxonomic arrangements have been proposed, for which several nomina lie in the synonymy or are relegated as subspecies of *P. lemniscatus*, and *P. bilineatus* considered as invalid, which would be an intergrade specimen between *Phalotris spegazzinii* Boulenger, 1913 and *Phalotris suspectus* Amaral, 1924 (Lema, 1979), or even where authors considered *P. bilineatus* as a junior synonym of *P. lemniscatus* (Cabral & Cacciali, 2015). A recent taxonomic revision of the *P. bilineatus* species group to Argentina proposed the revalidation of *P. spegazzinii* and *P. suspectus*, recognition of *P. bilineatus*, *P. lemniscatus*, *P. reticulatus*, *P. multipunctatus* and *P. normanscotti* as valid species, with the description of a new species, *Phalotris illustrator* Scrocchi, Giraudo & Nenda, 2022 (Scrocchi et al. 2022). Currently, there

are 12 available nomina associated with the *P. bilineatus* species group, with varying degrees of acceptance among authors (Uetz et al. 2023).

Considering these taxonomic issues, we provide an integrative revision for the *P. bilineatus* species group, reviewing available vouchered specimens and designating operational taxonomic units (OTUs), which were examined for external and internal morphology, as well as phylogenetic relationships based on molecular data. We examined specimens attributed to the nomina *P. bilineatus*, *P. lemniscatus*, *P. reticulatus*, *P. illustrator*, *P. multipunctatus*, *P. suspectus*, *P. spegazzinii* and *P. normanscotti*, based on the integration of molecular, internal and external morphology, inferring its phylogenetic relationships with other congeners, which were used to delimit and test species boundaries within the *P. bilineatus* species group.

Materials and Methods

Molecular Analyses

We extracted DNA from muscle tissue using the DNeasy DNA Extraction kit (ThermoFisher, MA, USA). We generated sequence data for 5 previously unsampled species, targeting the mitochondrial genes 12S rRNA (*12S ribosomal RNA*), 16S rRNA *(16S ribosomal RNA)*, and CYTB (*Cytochrome B*), and the nuclear gene BDNF (*Brain-derived neurotrophic factor*). Sequence fragments for four genes were amplified via polymerase chain reaction (PCR). Primers are described in Table 1. PCRs were performed using standard protocols, with addition of 0.4% of Triton X 100 to increase the efficiency of amplification. We used an annealing temperature of 54 ºC for 12S rRNA and 16S rRNA, 56º C for BDNF, a touchdown cycle of 50–60 ºC, with final annealing of 54 ºC for CYTB, JUN, NT3 and CMOS. PCR products were purified with shrimp alkaline phosphatase and exonuclease I (GE healthcare, Piscataway, NJ, USA), and the

sequences were processed using the DYEnamic ET Dye Terminator Cycle Sequencing Kit in a MegaBACE 1000 automated sequencer (GE healthcare, Piscataway, NJ, USA) following manufacturer's protocols. Both strands were checked, and when necessary edited manually. The consensus of both strands was generated using Geneious software (v. 7.1.8.; Kearse et al., 2012).

The phylogenetic relationships of *Phalotris* specimens and other Elapomorphini species were tested with an extended molecular data set for Caenophidia, modified from Zaher et al. (2018), for a total of four gene fragments, covering 56 terminals. This matrix is available in Supplementary File 1. Sequences were aligned using MAFFT 1.3.6 software (Katoh et al., 2013) plugin in Geneious (v. 7.1.8.; Kearse et al., 2012). The gene fragments 12S rRNA and 16S rRNA were aligned with an E-INS-i algorithm, and other gene fragments with a G-INS-i algorithm. We employed default parameters for gap opening and extension. We visually checked protein-coding genes with Geneious to verify the correct reading frame. Mitochondrial and nuclear gene fragments were concatenated with SequenceMatrix (v. 1; Vaidya et al. 2011).

We used PartitionFinder (v. 2; Lanfear et al., 2016) to choose the combined sets of partitioning schemes and models of molecular evolution, based on the Akaike Information Criterion with correction (AICc). We treated the two rRNA genes (12S and 16S) as separate partitions and partitioned protein coding genes by codon positions. We used the greedy search option in PartitionFinder, allowing only the selection of models of molecular evolution implemented in RAxML 8.2.3 (Stamatakis, 2014) without any correction for proportion of invariant sites, as recommended in the RAxML's manual.

We inferred phylogenetic relationships under Maximum Parsimony (MP) and Maximum Likelihood (ML) frameworks. The MP analysis was performed in TNT v1.6 software (Goloboff et al., 2008) considering gaps either as a fifth state or as missing data, using equally weighted

parsimony, traditional search, tree bisection reconnection algorithm, and 1,000 nonparametric bootstrap replicates. The ML analysis was computed using RAxML software (Stamatakis, 2014) in CIPRES Science Gateway (available at https://www.phylo.org/), with a mixed partition model from PartitionFinder 2, searching the most likely tree 100 times and conducting 1,000 nonparametric bootstrap replicates. The run was performed with the GTR + Γ model for all partitions. Due to the comparatively better resolution of polytomic nodes, we consider our resulting ML topology as a preferred hypothesis of phylogenetic relationships.

Uncorrected genetic distances (*p*-distances) were calculated for the CYTB gene fragment using MEGA software (v. 11; Tamura et al., 2021), using a d parameter (Transitions + Transversions), while assuming uniform rates among sites and a homogeneous pattern among lineages. The *p*-distance calculation was made based on the proportional (*p*) differences among nucleotide sites in which two compared sequences differ, as inferred through the division of nucleotide differences by the total nucleotides (Tamura et al., 2021). The ML topology was used to infer branch lengths and patristic distances (absolute time and mutation rate, to which patristic distances represent the sum of branch lengths used to link the terminal nodes of two species in a tree) as reinforcement proxies of genetic distance, following Montingelli et al. (2020) and Entiauspe-Neto et al. (2021c). We inferred patristic distances with the *ape* package (Paradis & Schliep, 2019) in R environment (v. 4.3; R Core Team, 2021).

In order to assess putatively interspecific variation and evolution in diagnostic coloration characters, we built a state matrix for two characters: white nuchal collar, as a continuous character, in which we assessed presence or absence, and width (measured to the nearest 0.01 mm within an individual), and the number of dorsal stripes, as a discrete character, mapped as two or three stripes as character states, based on the analysis of physical individuals

(Supplementary File 1). The continuous characters were evaluated with the *Contmap* function, which projects the observed and reconstructed values of a continuous trait onto the edges of a tree using a color gradient, estimating states at internal nodes using Maximum Likelihood and interpolating the states along each edge using equation [2] of Felsenstein (1985), transforming these values into maximum likelihood estimates under a Brownian evolutionary process (Revell, 2023). The discrete characters were evaluated with the *anc.ML* function, which estimates the evolutionary parameters and marginal ancestral states for Brownian evolution in nodes of a tree using Maximum Likelihood. Discrete characters were fitted with Equal Rates (ER, equal rates for all permitted transitions), Symmetric (SYM, symmetric backward & forward rates for all permitted transitions), and All Rates Different (ARD, all-rates-different for permitted transitions) Markov K (Mk) evolutionary models, which were evaluated with log Likelihood [log(L)], Akaike Information Criterion (AIC), and weighted Akaike Information Criterion score (AICw) statistical values. All analyses were conducted using the *phytools* R package (v. 1.5.1; Revell, 2012) in the R environment (R Core Team, 2023).

Selection of Analytical Units

We partitioned our sample into Operational Taxonomic Units (OTU, hereafter), in order to test for differential hypotheses of species limits. An overview of valid nomina associated with the *P. bilineatus* species group is referred to here (Figure 1–10). This method has been modified from Wüster & Thorpe (1992), considering it also accounts for the inclusion of juvenile specimens. We conducted the following OTU hypotheses: (1) Morphological-based, upon the congruence of ten diagnostic characters, based largely upon morphological characteristics and geographic distribution; (2) Molecular-based, in which we used our previous haplotype network as a

framework to delimit haplogroups, and subsequent OTUs; (3) Geographical-based, in which we used major geographic ecoregions to delimit OTUs. Individuals that could not be assigned to these aforementioned OTU's, with congruence for at least five of the diagnostic characters, were considered as outliers, and will be treated separately. Due to its small sample size and unique morphology, we consider the taxon *Phalotris multipunctatus* Puorto & Ferrarezzi, 1994 as *incertae sedis*, and refrain from including it in our analyses. Its taxonomic status and phylogenetic relationships will be addressed elsewhere (Entiauspe-Neto, O.M. pers. obs.). This species can be diagnosed from all other species of the *Phalotris bilineatus* species group by lacking a black cloacal collar.

Morphological Operational Taxonomic Units

For this approach, we evaluated for congruence in ten characters of external morphology (i.e. taxonomical approach) that could be associated to a geographic distribution pattern (n = 243). We compared our pooled morphological characters and geographic distribution to type specimens, in order to include available nomina to each candidate OTU.

OTU 1 (n = 209): head dorsal background coloration mostly black or dark brown, white nuchal collar usually present (up to three rows wide), black nuchal collar usually present (up to two rows wide), two medium-sized lateral black stripes (up to three rows wide), two large-sized red stripes (up to four rows wide), black vertebral stripe present (up to two rows wide), paraventral region uniformly cream, dorsal cloacal black blotch present, ventral coloration with medium-sized lunar black blotches, recorded to Pampas in northeastern Argentina, southern Brazil, and Uruguay. The nomina *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854,

Phalotris melanopleurus Cope, 1885, *Elapomorphus trilineatus* Boulenger, 1889, and *Elapomorphus lemniscatus divittatus* Lema, 1984 are available and assigned to this OTU.

OTU 2 (*n* = 68): head dorsal background coloration uniformly black, white nuchal collar present (up to three rows wide), black nuchal collar present (up to four rows wide), two large-sized lateral black stripes (up to four rows wide), background dorsal coloration in life uniformly red, vertebral stripe absent or reduced, paraventral region black, dorsal cloacal black blotch present, ventral coloration uniformly black, recorded to Araucaria forests in southern Brazil. The nomina *Elapomorphus reticulatus* Peters, 1860 and *Elapomorphus iheringi* Strauch, 1884 are available and assigned to this OTU.

OTU 3 (*n* = 28): head dorsal background coloration mostly black or dark brown, white nuchal collar usually absent, black nuchal collar usually absent, two small-sized lateral black stripes (up to a single row wide), background dorsal coloration light brown or gray, black vertebral stripe usually absent or vestigial, paraventral region uniformly cream, dorsal cloacal black blotch present, ventral coloration with small-sized lunar black blotches, recorded to Pampas in northwestern Argentina, central Paraguay, and southern Brazil. The nomina *Elapomorphus bilineatus* Duméril, Bibrón, & Duméril 1854, *Elapomorphus suspectus* Amaral, 1924 and *Phalotris normanscotti* Cabral & Cacciali, 2015 are available and assigned to this OTU.

OTU 4 (*n* = 26): head uniformly dark brown, white nuchal collar usually reduced, incomplete or absent (up to two rows wide), black nuchal collar usually absent, two medium-sized lateral black stripes (up to three rows wide), two large-sized brown or dark orange stripes (up to four rows wide), black vertebral stripe present (up to a single row wide), paraventral region uniformly cream or black, dorsal cloacal black blotch present, ventral

coloration with small-sized lunar black blotches, recorded to northern, central, southeastern, and southern Argentina. The nomina *Elapomorphus bollei* Mertens, 1954, *Elapomorphus spegazzinii* Boulenger, 1913, *Phalotris illustrator* Scrocchi, Nenda & Giraudo 2022 are available and assigned to this OTU.

Molecular Operational Taxonomic Units

In order to evaluate allele sharing among lineages and haplogroup structure among OTUs, we inferred a gene tree and haplotype network for a Mitochondrially Encoded Cytochrome B (*CYTB*) gene fragment alignment, of 567 bp. The gene tree was inferred under a Maximum Likelihood framework implemented in Randomized Axelerated Maximum Likelihood (RAxML HPC, v. 8.0; Stamatakis et al. 2014), using the same softwares and specifications described above. The haplotype network was inferred using a Median neighbor-joining method, as implemented in POPART (v.1; Leigh & Bryant, 2015). Based on the resulting haplotype network, we directly assigned OTUs based on geographical haplogroups (n = 232), and conducted a Minimum Convex Polygon based on georeferenced individuals of each haplogroup, to indirectly include sampled populations to these OTUs. We also assigned available nomina to individuals based on morphological identification, following Scrocchi et al. (2022).

OTU 1 (n = 132): individuals directly assigned from "Haplogroup 1", occuring in coastal and southern Rio Grande do Sul, Brazil, sampled from the localities of Arambaré, Bagé, Cachoeira do Sul, Caçapava, Eldorado do Sul, Dom Feliciano, Palmares do Sul, Rosário do Sul, and São José do Norte. The nomina *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854 and *Elapomorphus trilineatus* Boulenger, 1889 are assigned to this OTU.

OTU 2 (*n* = 18): individuals directly assigned from "Haplogroup 2", occurring in southwestern Rio Grande do Sul and northern Uruguay, sampled from the localities of Alegrete, Rosário do Sul, Quaraí, and Santana do Livramento. The nomen *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854 is assigned to this OTU.

OTU 3 (*n* = 79): individuals directly assigned from "Haplogroup 3", occurring in Santa Catarina, northern Rio Grande do Sul, and northwestern Argentina, sampled from the localities of Água Doce, Bom Jesus, Caxias do Sul, Colonia Carlos Pellegrini, Lagoa dos Esteves, and Manoel Viana. The nomina *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854, *Elapomorphus bilineatus* Duméril, Bibrón, & Duméril 1854, *Elapomorphus trilineatus* Boulenger, 1889 and *Elapomorphus reticulatus* Peters, 1860 are assigned to this OTU.

Geographical Operational Taxonomic Units

In order to test for geographical structure and fixation of morphological characters, we arbitrarily delimited OTUs based on major ecoregions (sensu Olson et al. 2001) and geographical barriers (*n* = 331). Examined individuals were georeferenced, overlaid to a geographic map with drawn geographical limits, and then assigned to individual OTUs. We also assign available nomina to individuals based on morphological identification, following Scrocchi et al. (2022).

OTU 1 (*n* = 277): individuals georeferenced upon the Dry Chaco, Espinal, Monte, Humid Chaco, Humid Pampas, and Yungas ecoregions in Argentina and Paraguay, at west of the Uruguay and Paraná rivers. The nomina *Elapomorphus bilineatus* Duméril, Bibrón, & Duméril 1854, *Elapomorphus suspectus* Amaral, 1924, *Elapomorphus bollei* Mertens, 1954, *Elapomorphus spegazzinii* Boulenger, 1913, *Phalotris normanscotti* Cabral & Cacciali, 2015,

and *Phalotris illustrator* Scrocchi, Nenda & Giraudo 2022 are available and assigned to this OTU.

OTU 2 (*n* = 52): individuals georeferenced upon the Atlantic Forest and Dry Pampas ecoregions in southern Brazil and Uruguay, at east of the Uruguay and Paraná rivers. The nomina *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854, *Phalotris melanopleurus* Cope, 1885, *Elapomorphus trilineatus* Boulenger, 1889, *Elapomorphus lemniscatus divittatus* Lema, 1984, *Elapomorphus reticulatus* Peters, 1860 and *Elapomorphus iheringi* Strauch, 1884 are available and assigned to this OTU.

External Morphology Analyses

We examined a total of 337 specimens of *Phalotris* (Appendix 1). Throughout this paper, we use the museum acronyms of Sabaj (2020). Scale counts follow Dowling (1951) and Peters (1964). Sex determination was done with a ventral incision in the base of the tail. An emended diagnosis is based on the nomenclature used by Entiauspe-Neto *et al*. (2021). We measured head length (from the center of the rostral to the corner of mouth), smallest head width (between nostrils), largest head width (at the corner of mouth), and other head measurements to the nearest 0.01 mm using a dial caliper; snout-vent length (SVL, from the center of the rostral to the posterior margin of the cloacal scale), total length (TTL, from the center of the rostral to tail tip), tail length (TL, from the posterior margin of the cloacal scale to the distal tip of the terminal scale) to the nearest 1 mm using a flexible ruler. We measured scales on the right side of head and defined our measurement within the description when appropriate.

In order to reduce subjectivity in identification of *Phalotris* specimens, we generated a dataset (Dataset 1; Supplementary File 1; Appendix I) of quantitative and qualitative diagnostic

morphological characters to be evaluated in a series of examined specimens of the *P. bilineatus* species group, which underwent sex determination: (1) snout-vent length (SVL), measured from the center of the rostral to the posterior margin of the cloacal scale; (2) tail length, from the posterior margin of the cloacal scale to the distal tip of the terminal scale; (3) head length, from the center of the rostral tip to the corner of the mouth, ventrally; (4) largest head width, measured at the suture of the supralabials with the infralabials; (5) smallest head width, measured at the nostril; (6) eye diameter; (7) rostral length; (8) rostral width; (9) internasal suture length; (10) prefrontal length; (11) prefrontal width; (12) prefrontal length; (13) prefrontal width; (14) parietal length; (15) temporal scales; (16) postocular scales; (17) supralabial scales; (18) infralabial scales; (19) infralabial scales contacting the anterior pair of chinshields; (20) ventral scales; (21) subcaudal scales; (22) white nuchal collar length, dorsally; (23) black nuchal collar width, dorsally; (24) vertebral stripe width; (25) lateral stripe width. In order to account for allometry, we follow Burbrink (2001) and log-transformed morphometric values, and additional categories include ratios of the aforementioned meristic and morphometric values.

Considering this dataset combines categoric, discrete and continuous variables in our multivariate analyses, we mainly used distribution-free statistical methods to analyze morphological data, following Barbo et al. (2022). To further ensure our data adhered to model assumptions, we identified outliers, collinearity, zero-variance in characters, and possible problems in our dataset through the inspection of Quantile-Quantile plots, correlograms, histograms, and box plot graphs. We used a Principal Component Analysis (PCA), as a dimensionality reduction technique in order to find the underlying structure of a dataset by identifying patterns in the data and expressing the data in a new, lower-dimensional space, as an exploratory approach to visually test for morphological differences among groups. We evaluated

intraspecific presence of sexual dimorphism within OTUs with a Student *t*-test (*t*) for morphometric (continuous) and a Wilcoxon test for meristic (discrete) characters, with *P*-value 0.05, after evaluating the assumptions of univariate normality by using a Shapiro-Wilk test, and homoscedasticity through Levene's test (Zar, 1999). Ranges are reported followed in parentheses by mean ± standard deviation, and sample size. All statistical analyses were conducted using built-in functions in R Environment (R Core Team, 2021).

Internal Morphology Analyses

We evaluate the skull morphology of specimens attributed to *P. illustrator* (MCN 6734*), *P. suspectus* (MCN 9726*), *P. spegazzinii* (MCN 9050, 3683, 8884), *P. lemniscatus* (UFRGS 6112, 6113, ZFMK 10219), *P. trilineatus* (UFRGS 6237*), and *P. reticulatus* (UFRGS 4965, 6486, ZFMK 102639) based on a high-resolution micro-CT scan, performed with a Bruker SkyScan 1173 (Bruker, Kontich, Belgium) at the IPR/PUCRS in Brazil. The scan was recorded over a 180° rotation with a frame averaging of 2, an X-ray beam with 50 kV source voltage, 160 μA current, an exposure time of 800 ms per projection, without the use of a filter. Rotation steps of 0.33° degrees were used, resulting in 727 projections, an isotropic voxel size of 7.86 μm, and a total scan duration of 38:40 min. We reconstructed the CT-dataset using N-Recon software version 1.7.4.6 (Bruker MicroCT, Kontich, Belgium) and rendered the images in three dimensions through the aid of Amira visualization software (FEI, Thermo Fisher Scientific). Segmentation to separate and color the bones was also performed using Amira. We use the osteological terminology of Bullock & Tanner (1966) and Cundall & Irish (2008), and the description of the skull follows Entiauspe-Neto et al. (2020a, 2021b).

We analyzed for differences in skull shape by setting landmarks over dorsal photographs of reconstructed high-resolution micro-CT scans of *Phalotris* specimens. Scanned individuals marked with an asterisk were the single representatives of their species, and were duplicated in our analyses. On each photograph, we digitized a configuration of 35 unilateral land-marks to characterize dorsal head shape (Supplementary Figure 1). We digitized the landmark coordinates using tpsDig version 2.17 (Rohlf, 2015). In order to maintain shape variation between landmark coordinates, we manually removed effects of scale, location and orientation by applying a generalized Procrustes analysis (Rohlf & Slice, 1990), employed with the R package geomorph (Baken et al. 2021). We calculated head size as a centroid size, considered to be the square root of the sum of the squared distance of every landmark to the centroid or "center" of the landmark configuration, as described by Esquerré et al. (2017). We analyzed landmark structure by employing a Principal Component Analysis (PCA), as a dimensionality reduction technique, over the generalized Procrustes analysis output.

Results

Taxonomic Summary

The taxonomic history of species that have been allocated to the *Phalotris bilineatus* species group is long and complicated. Therefore, it will be approached in detail here. As seen in several other Elapomorphini taxa (e.g. Entiauspe-Neto et al. 2019, 2020a-d, 2021b), most descriptions of *Phalotris* were based upon small series and diagnosed largely on characters of external morphology. In this section, original descriptions were translated and approached in detail for some key taxa, whenever valuable for comparisons.

Duméril, Bibrón & Duméril (1854:839) described *Elapomorphus bilineatus* based on a single specimen (MNHN-RA-0.3667), collected by Alcides D'Orbignyi in an unknown locality of Corrientes, northeastern Argentina. The description of *P. bilineatus* is particularly informative, for which the authors diagnose it from congeners based on "*Le dessus du corps avec deux lignes longitudinales noires; les gastrotèges noires au milieu. Huit plaques sus-céphaliques seulement; la pré-frontale n'étant pas divisée en deux pièces. Deux post-oculaires. Deux ou trois squammes temporales entre la plaque pariétale et les quatrièmoe, cinquièmae et sixième sus-làbiâles. Bout dé la queue non pointu*."[En: Upper body with two black longitudinal lines; black ventrals in the middle. Eight supracephalic plaques only; the prefrontal is not divided into two parts. Two postoculars. Two or three temporal scales between the parietal plate and the fourth, fifth and sixth supralabials. Point of tail blunt.]. It is also mentioned that the specimen bears 15 dorsal rows without reduction, 218 ventrals, a single cloacal scale, 22 subcaudals, and 5+2 maxillary teeth. At the time of our analysis, this specimen was severely dehydrated, and was seemingly damaged during the Second World War (T. Lema, pers. comm).

Duméril, Bibrón & Duméril (1854:840) also described *Elapomorphus lemniscatus*, based on a specimen (MNHN-RA-0.3668) from "Amérique du Sud' (South America), collected by Charles Darwin in the vicinities of Montevideo, Uruguay. This description is also very informative, as the authors stated "*Le dessus de la tête noir, le tronc partagé en six bandes longitudinales, trois blanches et trois noires. Tout le dessous du corps d'une couleur noire. Huit plaques sus-céphaliques seulement, la pré-frontale n'étant pas divisée en deux pièces. Deux post-oculaires. Deux squames temporales entre la plaque pariétale et les quatrième, cinquième et sixième sus-labiales; bout de la queue pointu*." [En: Top of the head black, trunk divided into six longitudinal bands, three white and three black. The entire underside of the body of a black

color. Eight supracephalic plates only, the prefrontal not being divided into two parts. Two postoculars. Two temporal scales between the parietal plate and the fourth, fifth and sixth supralabials; pointy tail tip.]. It is also mentioned that this specimen bears 15 dorsal rows without reduction, 192 ventrals, a single cloacal scale, and 27 subcaudals.

Shortly after, Peters (1860:518) described *Elapomorphus reticulatus* based on a specimen (ZMB 3811) from an unknown locality in Brazil. Although not as detailed as the previous descriptions, the author provided a general characterization of the specimen: "*Kopf, Bauchseite (mit Ausnahme des Afterschildes) und Seiten-theile schwarzbraun, Bauch-, Schwanzschilder und Seitenschuppen weifsgerändert, Halsband und Rückenseite weifs, mit Aus-nähme einer schwarzbraunen am Halse keulenförmig beginnenden schwarzbraunen Längslinie, welche bis zur etwas zusammengedrückten Schwanzspitze geht, und der Basis des Schwanzes, welche ebenfalls schwarzbraun ist. Hinterer gefurchter Oberkieferzahn mäßig groß, Nasenschild einfach, zwei Internasalschilder, ein einziges Praefrontalschild, 1 Praeoculare, 2 Postocularia, sechs Supralabialschilder, von denen das 2te und 3te in den Augen-sechs Supralabialschilder, von denen das 2te und 3te in den Augenring treten, das 4te an das untere Postorbitale stößt, das ote mit zwei Temporal Schildern in Berührung steht, von denen das vorderste länger aber weniger breit als das folgende ist.*" [En: Head, ventral side (with the exception of the cloacal shield) and lateral portions blackish brown, ventral and tail shields and side scales bordered in white, collar and back side white, with the exception of a black-brown, black-brown longitudinal line that begins in a club-shaped manner on the neck and extends to the slightly compressed tip of the tail and the base of the tail, which is also black-brown. Posterior grooved upper jaw tooth moderately large, nasal shield simple, two internasal shields, a single prefrontal shield, 1 preocular, 2 postocular, six supralabial shields, of which the 2nd and 3rd

enter the orbit, the 4th contacts the lower postorbital, two temporal shields are in contact, of which the foremost is longer but less broader than the anterior.]. The author also mentioned the specimen bears dorsals in 15 rows without reduction, ventrals 197, and subcaudals in 32 pairs.

This was followed by the taxonomic arrangement of Cope (1862), who described the genera *Apostolepis* and *Phalotris* for species previously allocated in *Elapomorphus*, allocating *E. bilineatus* and *E. lemniscatus* to *Phalotris* based on their cephalic scutelation. Strauch (1884:581) [1885] considered the genera erected by Cope (1862) as synonymous with *Elapomorphus*, and proposed that these should be recognized as groups within the former genus, also describing *Elapomorphus iheringi*, based on a specimen (ZISP 6955) from Taquara, Rio Grande do Sul, Brazil.

Cope (1885:189), possibly unaware of the proposed changes of Strauch (1884), described *Phalotris melanopleurus* based on two specimens (ANSP 11213, 11214) from Rio Grande do Sul, Brazil, for which no diagnosis is given, and only a brief morphological characterization of both specimens is made. In the same year, Boulenger (1885:321) published an account on the variation of specimens housed in the British Museum, which he identified as *E. lemniscatus*, and proposed that *E. iheringi* should be considered a junior synonym of it. Later, Boulenger (1889:265) described *Elapomorphus trilineatus* based on a single specimen (BMNH 1946.1.2.90) from Camaquã, Rio Grande do Sul, Brazil; this description is particularly brief, and bears no diagnosis from any congeners. Boulenger (1913:49) described another new taxon, *Elapomorphus spegazzinii*, based on a single specimen (MSNG 30651) from La Plata, Argentina, that is only briefly diagnosed in this work. Amaral (1924) described *Elapomorphus suspectus* based on a specimen (USNM 48939) from Pilár, Córdoba, Argentina; this description

contains only a small morphological section, and a mention that "*This species is closely allied to* E. lemniscatus, *from which it differs greatly in coloration*" (Amaral, 1924:202).

The revision of Amaral (1929) proposed the synonymization of several species; although this is an extensive work, the author only briefly commented on his reasoning for the allocation of *E. lemniscatus*, *E. spegazzinii*, *E. suspectus*, and *E. trilineatus* as junior synonyms of *E. bilineatus*, for which he attributed morphological differences to "*individual variations*" (Amaral, 1929:49). Mertens (1954) described *Elapomorphus bollei* based on a specimen (SMF 42004) from Tandil, Buenos Aires, Argentina. In a series of works, Thales de Lema proposed a series of taxonomic arrangements for this complex. Lema (1970) proposed the revalidation of *E. lemniscatus* and *E. reticulatus* as subspecies of *E. bilineatus*. Later, Lema (1978a), suggested the revalidation of *E. spegazzinii* as another subspecies of *E. bilineatus*, while also proposing that *E. bollei* was a junior synonym of *E. bilineatus*. Lema (1978b), proposed the revalidation of *E. suspectus* as a subspecies of *E. bilineatus*. Oddly enough, Lema (1979) suggested that the holotype of *E. bilineatus* was a hybrid between *E. bilineatus suspectus* and *E. bilineaus spegazzinii*, and this name should be considered synonymous with *E. lemniscatus*, while considering the nomina *E. reticulatus*, *E. spegazzinii*, and *E. suspectus* as subspecies of *E. lemniscatus*. Later, Lema (1984) proposed the recognition of *Phalotris* as a subgenus of *Elapomorphus*, and considered *E. trilineatus* and *E. iheringii* as subspecies of *E. lemniscatus*. The author also described a new subspecies, *Elapomorphus* (*Phalotris*) *lemniscatus divittatus*, based on a specimen (MCN 3784) from Passo do Atalho, Canguçu, Rio Grande do Sul, Brazil, distinguished from *E. lemniscatus lemniscatus* for lacking a vertebral black stripe.

In an unpublished Master dissertation, Ferrarezzi (1993) proposed the recognition of *Phalotris lemniscatus* and *Phalotris reticulatus* as valid species, considering *E. bilineatus*

synonymous with *P. lemniscatus* and *E. iheringi* synonymous with *P. reticulatus*. Ferrarezzi & Puorto (1993) presented a brief reappraisal on the taxonomic history of the group, which recognized only *P. bilineatus* and *P. lemniscatus* as valid taxa, and described *Phalotris multipunctatus* as a new species, based on specimens from São Paulo and Mato Grosso do Sul, which was assigned to the *P. bilineatus* group. Shortly after, Lema (1994) would argue for the recognition of *P. lemniscatus divittatus*, *P. lemniscatus trilineatus*, and *P. lemniscatus iheringi* as valid subspecies. Later, Cabral & Cacciali (2015) described *Phalotris normanscotti* based on specimens from the Chaco of Paraguay, and assigned it to the *P. bilineatus* group.

Recently, Scrocchi et al. (2022) provided a taxonomic revision for the species of *P. bilineatus* in Argentina, proposing the revalidation of two nomina, *P. suspectus* and *P. spegazzinii*, the description of *Phalotris illustrator* Scrocchi, Nenda & Giraudo 2022, and recognizing *P. bilineatus*, *P. lemniscatus*, *P. reticulatus*, *P. multipunctatus* and *P. normanscotti* as valid species. For these changes, the authors propose that *P. illustrator* can be diagnosed from other congeners by a black dorsal head coloration, absent nuchal collars, uniformly black venter, lateral black coloration, black cloacal collar present, and a rostral separated from prefrontal by the internasals (Scrocchi et al., 2022:50); *P. bilineatus* can be diagnosed from other congeners by a black dorsal head coloration, white rostral and supralabials, nuchal collars absent, circular ventral black blotches over a white background, narrow lateral stripes, rostral scales contacting or close to contacting prefrontal (Scrocchi et al., 2022:50); *P. lemniscatus* can be diagnosed from other congeners by a black dorsal head coloration, white rostral and supralabials, white and black nuchal collars present, circular ventral black blotches over a white background, black cloacal collar present, rostral well separated from prefrontal (Scrocchi et al., 2022:54); *P. reticulatus* can be diagnosed from other congeners by a uniformly black dorsal head coloration, with a single dot

on 3rd-4th supralabials, white and black nuchal collars present, black ventral coloration, wide lateral black stripes, rostral well-separated from prefrontals (Scrocchi et al., 2022:54); *P. spegazzinii* can be diagnosed from other congeners by a uniformly black dorsal head coloration, nuchal collars absent, ventral coloration uniformly black, rostral well separated from prefrontals (Scrocchi et al., 2022:56); *P. suspectus* can be diagnosed from other congeners by a uniformly black head, white and black nuchal collars present, circular ventral black blotches over a white background, rostral well separated from prefrontals (Scrocchi et al., 2022:56). There have been no recent changes in taxonomy for this group since then, with continuous disagreements between authors (see Cabral & Cacciali, 2015; Nogueira et al. 2019; Uetz et al. 2023; Entiauspe-Neto et al. 2022).

Phylogenetic Relationships and Phylogeography

Our concatenated alignment totalized 2980 base pairs (526 bp for 12S, 590 bp for 16S, 1092 bp for CYTB, and 769 for BDNF). The resulting ML (LnL = -26182.966570) and MP topologies (TBR = 5260) recovered a well-supported *Phalotris bilineatus* species complex (bootstrap support < 95%, Figure 11). In both analyses, we recover the terminals *P. bilineatus*, *P. trilineatus*, *P. lemniscatus*, and *P. reticulatus* as reciprocally paraphyletic. Both topologies largely agree with the major phylogenetic relationships among the sampled terminals within the *P. bilineatus* species group. Five major clades are supported by our ML and MP inferences: C.1, containing *P. bilineatus*, *P. lemniscatus P. trilineatus*, and *P. reticulatus* terminals, ranging from eastern Corrientes, in northeastern Argentina, to northern Rio Grande do Sul and southern Santa Catarina, in southern Brazil; C.2, containing *P. lemniscatus* and *P. trilineatus*, ranging from central to southeastern Rio Grande do Sul, in southern Brazil; C.3, containing *P. trilineatus*,

ranging in coastal Rio Grande do Sul, in southern Brazil; C.4, containing *P. lemniscatus*, ranging in central and western Rio Grande do Sul; C.5, containing *P. lemniscatus*, ranging in western Rio Grande do Sul, at the border with Uruguay.

Our haplotype network reveals three major allele haplogroups (Figure 12): Haplogroup 1, with seven haplotypes, composed of *P. lemniscatus* and *P. trilineatus*; Haplogroup 2, with seven haplotypes, composed exclusively of *P. lemniscatus*; Haplogroup 3, with 11 haplotypes, composed of *P. bilineatus*, *P. lemniscatus*, *P. trilineatus*, and *P. reticulatus*. These haplogroups reveal marked geographic segregation, with Haplogroup 1 widely distributed in northwestern, western, southern and eastern Rio Grande do Sul, Haplogroup 2 restricted to northwestern and central Rio Grande do Sul, and Haplogroup 3 distributed to northern and northwestern Rio Grande do Sul, Santa Catarina and northeastern Argentina. Haplogroup 3 appears to be restricted to north of the Ibicuí river, while Haplogroup 2 appears to be restricted to south of this river. For the CYTB mitochondrial gene fragment, our inferred uncorrected *p*-distances ranged from 0–0.09% among all sampled *Phalotris bilineatus* species group terminals, while patristic distances range from 0.002–0.08.

Multivariate Analyses of Operational Taxonomic Units

Our exploratory analyses found a largely homogeneous or invariable distribution for meristic characters associated to cephalic scutellation (number of postocular scales, anterior, medial, and posterior temporals, supralabials) along the sampled range. These characters were considered to be largely uninformative and were removed from our multivariate analyses. After our exploratory analyses, the final dataset consisted of 23 log-transformed morphometric measurements, ratios, and meristic scale counts.

In the morphological OTUs hypothesis (H1), the first two components of the principal component analysis explained 62.1% (46.3% and 15.3%, respectively) of the total variance for males and females (Fig. 13). No groups associated with external morphology can be distinguished, and a total overlap can be seen among OTUs 01-04 for the analysis combining males and females, while OTUs 01-02 can be partially distinguished for analyses discriminating between males and females, while still bearing overlap. Among all H1 OTUs, we encountered sexual dimorphism for rostral length ($t^{327} = 0.008$), smallest head width ($t^{296} = 0.004$), and rostral width ($t^{297} = 0.0006$).

In the molecular OTUs hypothesis (H2), the first two components of the principal component analysis explained 65% (48.8% and 16.2%, respectively) of the total variance for males and females (Fig. 14). Similarly, no groups associated with molecular haplogroups can be diagnosed by external morphology, and a total overlap was observed for OTUs 01-03, in analysis combining and discriminating between males and females. Among all H2 OTUs, we encountered sexual dimorphism for rostral length ($t^{232} = 0.005$), smallest head width ($t^{231} = 0.009$), head length ($t^{231} = 0.0001$), rostral width ($t^{231} = 0.0008$), and infralabials on the right side of head ($W^{231} = 0.03$).

In the geographical OTUs hypothesis (H3), the first two components of the principal component analysis explained 62.7% (47.7% and 15%, respectively) of the total variance for males and females (Fig. 15). As for our other hypothesis, no groups associated with geographical limits can be diagnosed for external morphology, and a total overlap was observed for OTUs 01-02. Among all H3 OTUs, we encountered sexual dimorphism for rostral length ($t^{325} = 0.008$), smallest head width ($t^{291} = 0.003$), frontal length ($t^{311} = 0.0001$), rostral width ($t^{295} = 0.0006$),

infralabials on the right side of head ($t^{331} = 0.03$), and infralabials on the left side of head ($t^{330} = 0.03$).

Our PCA analysis of skull shape based on procrustes analyses of landmarks provides little support for osteological distinction of morphological-based species in the *P. bilineatus* species group (Figure 16). We uncovered three major non-overlapping groups associated with morphology: 1) composed by *P. suspectus*; 2) composed by *P. illustrator*, *P. lemniscatus*, *P. reticulatus*, and *P. spegazzinii*; 3) composed by *P. trilineatus*. While there is no overlap between *P. suspectus*, *P. trilineatus* and other sampled species, it should be noted that both these species are represented by duplicated landmarks over a single photograph. Furthermore, the sampled skull of *P. trilineatus* appears to be damaged, which could explain its difference from other groups. Considering the relatively small sample available (total $n = 1–3$ scan per species), this analytical approach should be seen as provisional, as it does not account for geographical or ontogenetic variation. No major qualitative osteological character was found to distinguish morphological species.

The aforementioned diagnostic characters of coloration (white nuchal collar presence, stripe size, ventral coloration) and pholidosis (rostral to prefrontal contact) are shown here to be polymorphic (Supplementary Table 1). For instance, our continuous character mapping reveals that the white nuchal collar was independently lost at least 14 times within terminals of the *P. bilineatus* species group (Figure 17). The discrete diagnostic character, the number of dorsal stripes, is shown to have changed in at least four instances (ER AICw = 0.11; ARD AICw = 0.77; SYM AICw = 0.11; Figure 18). Other characters, such as the rostral to prefrontal contact, are recovered as absent ($n = 259$) or present ($n = 17$) in *P. lemniscatus*, and therefore shown to be prone to intraspecific variation. Considering the lack of diagnostic characters in external and

internal morphology, and reciprocal paraphyletism, we propose to include *Elapomorphus trilineatus* Boulenger, 1889 and *Elapomorphus reticulatus* Peters, 1860 as junior synonyms of *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854.

Species Account

***Phalotris lemniscatus* (Duméril, Bibrón, & Duméril, 1854)**

Synonymy:

Elapomorphus lemniscatus Duméril, Bibrón, & Duméril, 1854:840. Holotype adult male (MNHN-RA 3668, Figure 18), collected by Charles Darwin, from Montevideo, Departamento Montevideo, Uruguay.

Elapomorphus reticulatus Peters, 1860:518. Holotype adult male (ZMB 3811, Figure 19), from "Brazil", estimated here to be likely from Santa Catarina, Brazil. **New synonymy.**

Elapomorphus iheringi Strauch, 1885:571. Holotype adult female (ZISP 6255, Figure 19), from Mundo Novo, Taquara, Rio Grande do Sul, Brazil. **New synonymy.**

Phalotris melanopleurus Cope, 1885:189. Syntype adult female (ANSP 11213), collected by Edward D. Cope, from São João do Monte Negro, Rio Grande do Sul, Brazil.

Elapomorphus trilineatus Boulenger, 1889:265. Holotype adult female (BMNH 1946.1.2.90, Figure 20), from Rio Camaquã, São Lourenço do Sul, Rio Grande do Sul, Brazil. **New synonymy.**

Elapomorphus (Phalotris) lemniscatus divittatus Lema, 1984:58. Holotype adult female (MCN 3784), from Passo do Atalho, Canguçu, Rio Grande do Sul, Brazil. Paratypes (MCN 4458) from Bagé, Rio Grande do Sul, Brazil; (MCN 4476, 1772) from Pelotas, Rio Grande do Sul, Brazil.

Diagnosis: A species of *Phalotris* with the following characters: (1) 15/15/15 dorsal scales, without apical pits; (2) preocular present, usually contacting nasal; (3) loreal absent; (4) temporals 1+1+2; (5) supralabials six, rarely five or seven, 2nd–3rd entering orbit; (6) infralabials six, rarely seven, 1st–4th contacting anterior chinshields; (7) ventrals 177–216 (177–208 in males, 185–216 in females); (8) subcaudals 16–39 (21–39 in males, 16–34 in females); (9) dorsal background coloration uniformly red, orange or yellow, with three or two longitudinal black or dark brown stripes; (10) ventral background coloration uniformly cream, with semilunar black markings ranging from small to covering entire ventral scale; (11) white nuchal collar present or absent, ranging from 0.5–5 dorsal scales; (12) black nuchal collar present or absent, ranging from 0.5–3 dorsal scales; (13) cloacal black collar present, reaching up to eight dorsal scales; (14) SVL 118–714 mm, TL 9–59 mm.

Redescription of the holotype: Adult male (MNHN-RA 3668). Total length 268 mm; SVL 245 mm; tail length 23 mm (8.57% of total length, 9.39% of SVL). Head length 8 mm (2.99% of total length, 3.27% of SVL); largest head width 4.6 mm (57.5% of head length); interorbital distance 3.2 mm (69.5% of largest head width); rostro-orbital distance 6.3 mm; naso-orbital distance 2.1 mm (33.3% of rostro-orbital distance). Cervical constriction indistinct; head slightly distinct from neck, rounded in dorsal view, slightly narrow anteriorly, arched in lateral view. Pupil indistinct. Rostral rounded, 0.6 mm long, slightly projected over lower jaw, length of portion visible in dorsal view slightly smaller than distance to external anterior edge of prefrontal. Internasals present, paired, square-shaped, 0.53 mm long, 0.7 mm wide; each internasal contacting rostral, prefrontal, and nasals. Prefrontal present, single, rectangular-shaped, 1.34 mm long, 1.2 mm wide; prefrontal contacts internasals, nasals,

preocular, supraocular, and frontal. Frontal hexagonal, 1.92 mm long, 1.6 mm wide, contacting prefrontal, supraocular, and parietal. Supraocular trapezoidal, longer than wider, in contact with posterior edges of prefrontal, lateral edges of frontal, superior edge of preocular, superior edge of upper postocular, and anterior edges of parietals. Parietals paired, right parietal 3.7 mm long, 2.4 mm at its largest width, contacting supraocular, frontal, temporals 1–3. Occipitals square-shaped, enlarged; each occipital contacts a single interoccipital and dorsals. Interoccipitals three, slightly smaller than vertebral and paravertebral rows of dorsals. Nasal triangular, undivided, twice as long as wide, contacting rostral, prefrontal, first and second supralabials, and preocular. Nostril in anterior third of nasal, slightly visible from above. Prefrontal pentagonal, contacting second supralabial, nasal, internasal, and supraocular. Eye diameter 0.9 mm. Postoculars two, square-shaped, contacting supraocular, third and fourth supralabial, and parietal. Temporals 1+1+2, contacting fourth, fifth, and sixth supralabial, postoculars, and parietal. Six supralabials, second and third entering orbit; fifth contacting postocular and parietal, sixth contacting parietal; fifth supralabial largest length and height. Mental triangular, as long as wide. Two pairs of chinshields, second pair longer. Six infralabials, 1–4 in contact with anterior chinshields, 5–6 in contact with posterior chinshields, first pair in contact with each other behind mental, fourth and fifth infralabials largest, equal in size. Dorsals smooth, in 15/15/15 rows. Ventrals 192. Subcaudals paired, 27/27. Terminal scale acuminate, curved ventrally.

Head uniformly dark brown dorsally and laterally, dark brown pigmentation covering up to occipital scales. Rostral uniformly dark brown on dorsal surface, with diffuse white pigmentation on its dorsal surface and anterior edges, reaching up to internasals. Supralabial blotch indistinct, restricted to small diffuse white pigmentation on third supralabial. White nuchal collar present, starting at first dorsal scale row, covering up to three dorsal rows in width,

smallest width at vertebral scale row. Black nuchal collar present, starting at third dorsal scale row, covering up to four dorsal rows in width. Diffuse dark brown pigmentation on outer edges of infralabials and chinshields. Black gular collar present, covering up to four rows of gular scales. Dorsal background coloration cream, with paired paraventral cream stripes. Dorsal stripes present, three, dark brown colored; vertebral stripe covering vertebral row and half of adjacent dorsal rows, lateral stripes paired, covering up to three rows each. Ventral background coloration cream, with lunar shaped dark brown blotches, gradually increasing in size from first body third up to tail. Subcaudal coloration uniformly dark brown. Cloacal black collar present, dark brown, covering up to 11 dorsal scale rows.

Discussion

Our comparative analyses of mitochondrial and nuclear gene fragments provide no support for the recognition of different monophyletic taxons within the *Phalotris lemniscatus* species complex. The evaluated taxa, *P. bilineatus*, *P. reticulatus*, *P. lemniscatus*, and *P. trilineatus*, were unambiguously recovered as reciprocally paraphyletic or polyphyletic in most analyses, which appear to suggest no support for the current morphological recognition of species in the *P. bilineatus* group. This relationship is also supported by our uncorrected *p*-distances and phylogeographic analysis of mitochondrial gene fragments, in which small distances and no taxon specific genetic or geographic clade structure was recovered.

The multivariate morphological analyses also provide no ground for the recognition of distinct taxa within the *P. bilineatus* species group. While the limited available sample for some species (e.g. *P. normanscotti*, known from three individuals) and the numerical unevenness of some evaluated taxa (*e.g. P. lemniscatus* with 136 individuals and *P. illustrator* with six

individuals) represents a significant statistical constraint to our analyses, we found no unambiguous morphological diagnosis supporting the current taxonomy for the *P. bilineatus* species group, with significant overlap in all evaluated meristic and morphometric characters. The coloration pattern characters used for diagnosis of species within the group are also shown to be polymorphic (Figure 16–17), representing merely coloration phases loosely associated with geography, and not supported by morphology or molecular datasets.

Our analyses also reinforce the need for integration of molecular data in species delimitation, which might uncover paraphyletic relationships or incomplete lineage sorting in otherwise morphologically conservative taxons. While speciation is not conditional to reciprocal monophyletism (Avise, 2000), the delimitation of species based exclusively upon external morphology is likely to overestimate biological diversity, which appears to occur in the *Phalotris bilineatus* species complex. The current species delimitation for *P. bilineatus*, *P. lemniscatus*, *P. trilineatus*, and *P. reticulatus*, which are diagnosed based on different coloration patterns and geographic ranges, appear to fit an isolation-by-distance intraspecific variation model, representing a single, widespread and phenotypically diverse species.

Several *Phalotris* species remain diagnosed based exclusively upon characters of external morphology, such as coloration and pholidosis (Cacciali & Cabral, 2014; Atkinson et al. 2018; Entiauspe-Neto et al. 2021; Smith et al. 2022; Scrocchi et al. 2022). Due to the lack of comprehensive morphological series and molecular data for *P. bilineatus*, *P. illustrator*, *P. normanscotti*, *P. spegazzinii*, and *P. suspectus*, we refrain from proposing synonymies among these taxa, although it should be noted that their current diagnoses are based upon largely uninformative data. Further sampling of individuals for their morphology, added to sequencing

of DNA sequences, should allow for the pairwise testing of these taxonomic entities, and clarify their taxonomic validity.

Acknowledgements

We are deeply indebted to Renata Perez and Geiza Pontes Esteves (UFRGS) for kindly allowing use of generated data for this work. We would like to thank Diego J. Alvares (UFRGS) for providing photographs of *Phalotris* specimens in life. We would also like to thank the curators A. Argôlo (UESC), A.L. Prudente (MPEG), A. Kluge, D. Rabosky (UMMZ), C.J. Franklin, E.N. Smith, J. Campbell (UTA), D.M. Borges-Nojosa (CHUFC), D. Klingberg Johansson, P.R. Møller (ZMUC), D. Meneghelli (UFRO-H), D. Santana (ZUFMS), E.M.X. Freire (MUFAL), F. Curcio, C. Strüssmann (UFMT-R), F.G. Grazziotin, G. Puorto, V. J. Germano (IBSP), G. Skuk (in memorian, MUFAL), G. Cotta (FUNED), G. Köhler (SMF), H. Zaher (MZUSP), L. Gonzales, S. Reichle (MNKR), L. Vohnnahme, D. Kizirian, L.P. Nuñez (AMNH), F. Werneck (INPAH), J. Mata (FMNH), J. Martínez (MCZ), J.C. Moura-Leite (MHNCI), M. Calderón, G.M. Fabián-Rangel (ICN), M. Moura (MZUFV), M.-O. Röddel (ZMB), N. Gilmore (ANSP), N. Vidal (MNHN), P. Cacciali (MHNP), P. Manzani (ZUEC), P. Passos (MNRJ), P. Campbell (BMNH), R. Lira-da-Silva (MZUFBA), R. Brown (KU), and S. Schweiger (NMW), kindly allowed and assisted with the examination of specimens under their care. OME-N thanks Fundação de Amparo à Pesquisa do Estado de São Paulo for his research grant (FAPESP, 2021/13671-7) and the Scales of Biodiversity Project (Escalas da Biodiversidade, FAPESP, 2016/13469-5, 2016/50127-5) for logistical support. MBM thanks the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for funding the the project “Taxonomia e Sistemática de Lagartos e

Serpentes (Lepidosauria, Squamata) no Bioma Pampa no sul do Brasil e Uruguai" (Process 562355/2010-3) in the "Brazilian Program in Taxonomy - MCT/CNPq/MEC/CAPES/PROTAX (Edital nº 52/2010)".

References

Amaral, A. D. (1924). New genus and species of South American snakes contained in the United States National Museum. Journal of the Washington Academy of Sciences, 14(9), 200-202.

Amaral, A. D. (1929). Valor systematico de várias formas de ophidios neotropicos. Estudo sobre os ophidios neotropicos XVII. Memórias do Instituto Butantan, 4, 3-67.

Atkinson, K., Smith, P., Dickens, J. K., & Catherine, L. Z. (2018). Rediscovery of the 'lost' snake *Phalotris multipunctatus* (Serpentes: Dipsadidae) in Paraguay with behavioral notes and reference to the importance of Rancho Laguna Blanca for its conservation. Current Herpetology, 37(1), 75-80.

Avise, J. C. (2000). Phylogeography: the history and formation of species. Harvard university press.

Baken, E. K., Collyer, M. L., Kaliontzopoulou, A., & Adams, D. C. (2021). geomorph v4. 0 and gmShiny: Enhanced analytics and a new graphical interface for a comprehensive morphometric experience. Methods in Ecology and Evolution, 12(12), 2355-2363.

Barbo, F. E., Booker, W. W., Duarte, M. R., Chaluppe, B., Portes-Junior, J. A., Franco, F. L., & Grazziotin, F. G. (2022). Speciation process on Brazilian continental islands, with the description of a new insular lancehead of the genus *Bothrops* (Serpentes, Viperidae). Systematics and Biodiversity, 20(1), 1-25.

Boulenger, G. A. (1885). IX.—Second list of reptiles and batrachians from the province Rio Grande do Sul, Brazil, sent to the Natural-History Museum by Dr. H. von Ihering. Journal of Natural History, 16(92), 85-88.

Boulenger, G. A. (1889). XXXIII.—Descriptions of a new snake and two new fishes obtained by Dr. H. von Ihering in Brazil. Journal of Natural History, 4(22), 265-267.

Boulenger, G. A. (1913). Descriptions of a new lizard and a new snake from South America. Annali del Museo cívico di storia naturale di Genova, 6, 49-50.

Bullock, R. E., & Tanner, W. W. (1966). A comparative osteological study of two species of Colubridae (*Pituophis* and *Thamnophis*). Brigham Young University Science Bulletin Biological Series, 8, 1–29.

Burbrink, F. T. (2001). Systematics of the eastern ratsnake complex (*Elaphe obsoleta*). Herpetological monographs, 1-53.

Cabral, H., & Cacciali, P. (2015). A new species of *Phalotris* (Serpentes: Dipsadidae) from the Paraguayan Chaco. Herpetologica, 71(1), 72-77.

Cope, E. D. (1862). Catalogues of the reptiles obtained during the explorations of the Parana, Paraguay, Vermejo and Uraguay Rivers, By Capt. Thos. J. Page, USN; and of those procured by Lieut. N. Michler, US Top. Eng., Commander of the expedition conducting the survey of the Atrato River. Proceedings of the Academy of Natural Sciences of Philadelphia, 346-594.

Cope, E. D. (1885). Twelfth contribution to the herpetology of tropical America. Proceedings of the American Philosophical Society, 22(118), 167-194.

Cundall, D., & Irish, F. (2008). The snake skull. In: Gans, C., Gaunt, A. S., & Adler, K. (eds) Biology of Reptilia. The Skull of Lepidosauria: 349–692. Society for the Study of Amphibian and Reptiles Press, USA.

Dayrat, B. (2005). Towards integrative taxonomy. Biological journal of the Linnean society, 85(3), 407-417.

De Queiroz, K. (2007). Species concepts and species delimitation. Systematic biology, 56(6), 879-886.

Dowling, H. G. (1951). A proposed standard system of counting ventrals in snakes. British Journal of Herpetology, 1, 97–99.

Duméril, A. M. C.; Bibron, G. & Duméril, A. H. A. (1854). Erpétologie générale ou histoire naturelle complète des reptiles. Tome septième. Deuxième partie, comprenant l'histoire des serpents venimeux. Librairie Encyclopédique de Roret, Paris, 1(12): 781-1536.

Entiauspe-Neto, O. M., Abegg, A. D., Koch, C., Nuñez, L. P., Azevedo, W. S., Moraes, L. J., ... & Loebmann, D. (2021). A new species of *Erythrolamprus* (Serpentes: Dipsadidae: Xenodontini) from the savannas of northern South America. Salamandra, 57(2): 196–218.

Entiauspe-Neto, O. M., de Sena, A., Tiutenko, A., & Loebmann, D. (2019). Taxonomic status of *Apostolepis barrioi* Lema, 1978, with comments on the taxonomic instability of *Apostolepis* Cope, 1862 (Serpentes, Dipsadidae). Zookeys, 841, 71.

Entiauspe-Neto, O. M., Koch, C., Guedes, T. B., & Tiutenko, A. (2020). Revisiting the taxonomic status of *Apostolepis sanctaeritae*, a forgotten Neotropical dipsadid snake. Salamandra, 56(4): 329–341.

Entiauspe-Neto, O. M., Koch, C., Harvey, M. B., Colli, G. R., & Guedes, T. B. (2021). Redescription of *Apostolepis ambiniger* (Peters, 1869)(Serpentes: Dipsadidae: Elapomorphini). Vertebrate Zoology, 71, 231-251.

Esquerré, D., Sherratt, E., & Keogh, J. S. (2017). Evolution of extreme ontogenetic allometric diversity and heterochrony in pythons, a clade of giant and dwarf snakes. Evolution, 71(12), 2829-2844.

Ferrarezzi, H. (1993). Sistemática filogenética de *Elapomorphus*, *Phalotris* e *Apostolepis* (Serpentes, Colubridae, Xenodontinae). MsC dissertation, Universidade de São Paulo, Brazil.

Ferrarezzi, H. (1994). Uma sinopse dos gêneros e classificação das serpentes (Squamata): II. Família Colubridae. In: Nascimento LB, Bernardes AT, Cotta GA (Eds) Herpetologia no Brasil 1. PUCMG, Belo Horizonte, 81–91.

Freitas, I., Ursenbacher, S., Mebert, K., Zinenko, O., Schweiger, S., Wüster, W., ... & Martínez-Freiría, F. (2020). Evaluating taxonomic inflation: towards evidence-based species delimitation in Eurasian vipers (Serpentes: Viperinae). Amphibia-Reptilia, 41(3), 285-311.

Giraudo, A.R. & Scrocchi, G.J. 2002. Argentinian Snakes: an annotated checklist. Smithsonian Herpetological Information Service, 132:1-53.

Goloboff, P. A., Farris, J. S., & Nixon, K. C. (2008). TNT, a free program for phylogenetic analysis. Cladistics, 24(5), 774-786.

Guedes, T. B., Entiauspe-Neto, O. M., & Costa, H. C. (2023). Lista de répteis do Brasil: atualização de 2022. Herpetologia Brasileira 12(1): 56-161.

Harvey, M.B. (1999). Revision of Bolivian *Apostolepis* (Squamata: Colubridae). Copeia 1999: 388–409.

Kaiser, H., Crother, B. I., Kelly, C. M., Luiselli, L., O'Shea, M., Ota, H., & Wüster, W. (2013). Best practices: in the 21st century, taxonomic decisions in herpetology are acceptable only when supported by a body of evidence and published via peer-review. Herpetological Review 44: 8–23.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647–1649.

Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017). PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular biology and evolution, 34(3), 772-773.

Leigh, J. W., & Bryant, D. (2015). POPART: full-feature software for haplotype network construction. Methods in ecology and evolution, 6(9), 1110-1116.

Lema, T. (1970). Sobre o status de *Elapomorphus bilineatus* Dumeril, Bibron & Dumeril, 1854, curiosa serpente subterranea. Iheringia, Ser. Zool. (38): 89-118.

Lema, T. (1978). O status de *Elapomorphus suspectus* Amaral 1924. (Ophidia: Colubridae). Comunicações do Museu de Ciências da Pontifícia Universidade Católica do Rio Grande do Sul, 16/17: 1-16.

Lema, T. (1978a). Invalidação de *Elapomorphus bollei* Mertens 1954 e o status de *Elapomorphus spegazzinii* Boulenger 1913 (Ophidia: Colubridae). Comunicações do Museu de Ciências da Pontifícia Universidade Católica do Rio Grande do Sul, 11−17.

Lema, T. (1979) Sobre a validade dos nomes *Elapomorphus bilineatus* Duméril, Bibron & Duméril, 1854 e *E. lemniscatus* Duméril, Bibron & Duméril, 1854 (Ophidia: Colubridae). Iheringia Série Zoologia, 54, 77−81.

Lema, T. (1984) Sobre o gênero *Elapomorphus* Wiegmann (Serpentes, Colubridae, Elapomorphinae). Iheringia Série Zoologia, 64, 53−86.

Lema, T. (1994) Lista comentada dos répteis ocorrentes no Rio Grande do Sul, Brasil. Comunicações do Museu de Ciências e Tecnologia da Pontifícia Universidade Católica do Rio Grande do Sul, 7, 41−150.

Maddison, W. P., & Whitton, J. (2023). The Species as a Reproductive Community Emerging From the Past. Bulletin of the Society of Systematic Biologists, 2(1), 1–35.

Mayden, R. L. (1997). A hierarchy of species concepts: the denouement in the saga of the species problem. In: Species: The units of biodiversity. London Chapman and Hall (pg. 381-424).

Mertens, R. (1954). Eine neue Natter der Gattung *Elapomorphus*. Senckenbergiana, 34: 183-185.

Montingelli, G. G., Barbo, F. E., Pereira Filho, G. A., Santana, G. G., França, F. G. R., Grazziotin, F. G., & Zaher, H. (2020). A second new species for the rare dipsadid genus *Caaeteboia* Zaher et al., 2009 (Serpentes: Dipsadidae) from the Atlantic Forest of northeastern Brazil. Cuadernos de Herpetología, 34.

Nogueira, C. C., Argôlo, A. J., Arzamendia, V., Azevedo, J. A., Barbo, F. E., Bérnils, R. S., ... & Martins, M. (2019). Atlas of Brazilian snakes: verified point-locality maps to mitigate the Wallacean shortfall in a megadiverse snake fauna. South American Journal of Herpetology, 14(sp1), 1-274.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V., Underwood, E. C., ... & Kassem, K. R. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. BioScience, 51(11), 933-938.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526–528.

Peters, J. A. (1964). Supplemental notes on snakes of the subfamily Dipsadinae (Reptilia: Colubridae). Studies on Neotropical Fauna and Environment, 4(1), 45–50.

Peters, W. C. H. (1860). Drei neue Schlangen des k. zoologischen Museums aus America und Bemerkungen über die generelle Unterscheidung von anderen bereits bekannten Arten. Monatsberichte der Königlichen Preussischen Akademie der Wissenschaften zu Berlin, 1860, 517-521.

Puorto, G., & Ferrarezzi, H. (1993). Uma nova espécie de *Phalotris* Cope, 1862, com comentários sobre o grupo bilineatus (Serpentes: Colubridae: Xenodontinae). Memórias do Instituto Butantan, 55(1), 39-46.

R Development Core Team. (2015). Team. R: a language and environment for statistical computing, Release 3.2. Available from http://cran.r-project.org [accessed 1 Aug. 2021].

Revell, L. (2012). “phytools: An R package for phylogenetic comparative biology (and other things).” Methods in Ecology and Evolution, 3, 217–223.

Rohlf, F. J. (2015). The tps series of software. Hystrix, 26(1), 9-12.

Rohlf, F. J., & Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. Systematic zoology, 39(1), 40-59.

Sabaj, M. H. (2020). Codes for natural history collections in ichthyology and herpetology. Copeia, 108(3), 593-669.

Savitzky, A. H. (1979). The origin of the New World proteroglyphous snakes and its bearing on the study of venom delivery systems in snakes. PhD thesis, University of Kansas, USA.

Scrocchi, G. J., Giraudo, A. R., & Nenda, S. J. (2022). Taxonomic notes on the *Phalotris bilineatus* group (Serpentes: Dipsadidae: Elapomorphini), with the description of a new species from northwestern Argentina. Cuadernos de Herpetología, 36: 47-63.

Smith, P., Brouard, J. P., & Cacciali, P. (2022). A new species of *Phalotris* (Serpentes, Colubridae, Elapomorphini) from Paraguay. Zoosystematics and Evolution, 98(1), 77-85.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 30(9), 1312-1313.

Strauch, A. (1884): Bemerkungen über die Schlangengattung *Elapomorphus* aus der Familie der Calamariden, 2nd Edition – Bulletin de l'Académie impériale des sciences de St.-Pétersbourg, 29: 541–590.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., & Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution, 30(12), 2725-2729.

Uetz, P., P. Freed & J. Hošek (2019): The Reptile Database. – Online available at: http://www.reptile-database.org, accessed May 5th, 2023.

Vaidya, G., Lohman, D. J., & Meier, R. (2011). SequenceMatrix: concatenation software for the fast assembly of multi‐gene datasets with character set and codon information. Cladistics, 27(2), 171-180.

Wüster, W., & Thorpe, R. S. (1992). Asiatic cobras: population systematics of the *Naja naja* species complex (Serpentes: Elapidae) in India and Central Asia. Herpetologica, 69-85.

Zaher, H. (1994). Phylogenie des pseudoboini et evolution des Xenodontinae sud-americains (Serpentes, Colubridae). PhD dissertation, Musée National D'Histoire Naturelle, France.

Zaher, H., Grazziotin, F. G., Cadle, J. E., Murphy, R. W., Moura-Leite, J. C. D., & Bonatto, S. L. (2009). Molecular phylogeny of advanced snakes (Serpentes, Caenophidia) with an emphasis on South American Xenodontines: a revised classification and descriptions of new taxa. Papéis Avulsos de Zoologia, 49(11), 115-153.

Zar, J. H. (1999). Biostatistical Analysis. 4th edition. Prentice Hill.

Figures

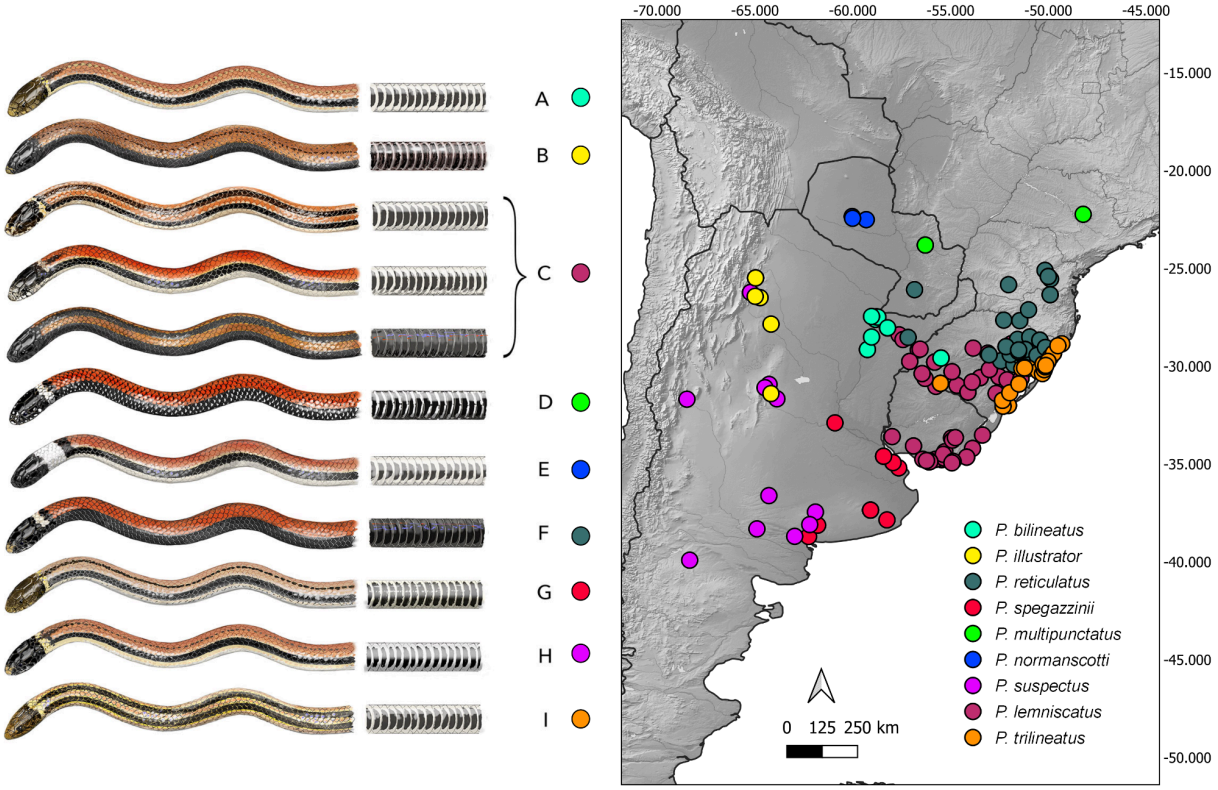


Figure 1. Overview of valid nomina within the *P. bilineatus* species group, sensu Scrocchi et al. (2023) and Uetz et al. (2023), and their geographic distribution of examined specimens. A = *P. bilineatus*; B = *P. illustrator*; C = *P. lemniscatus*; D = *P. multipunctatus*; E = *P. normanscotti*; F = *P. reticulatus*; G = *P. spegazzinii*; H = *P. suspectus*; I = *P. trilineatus*.

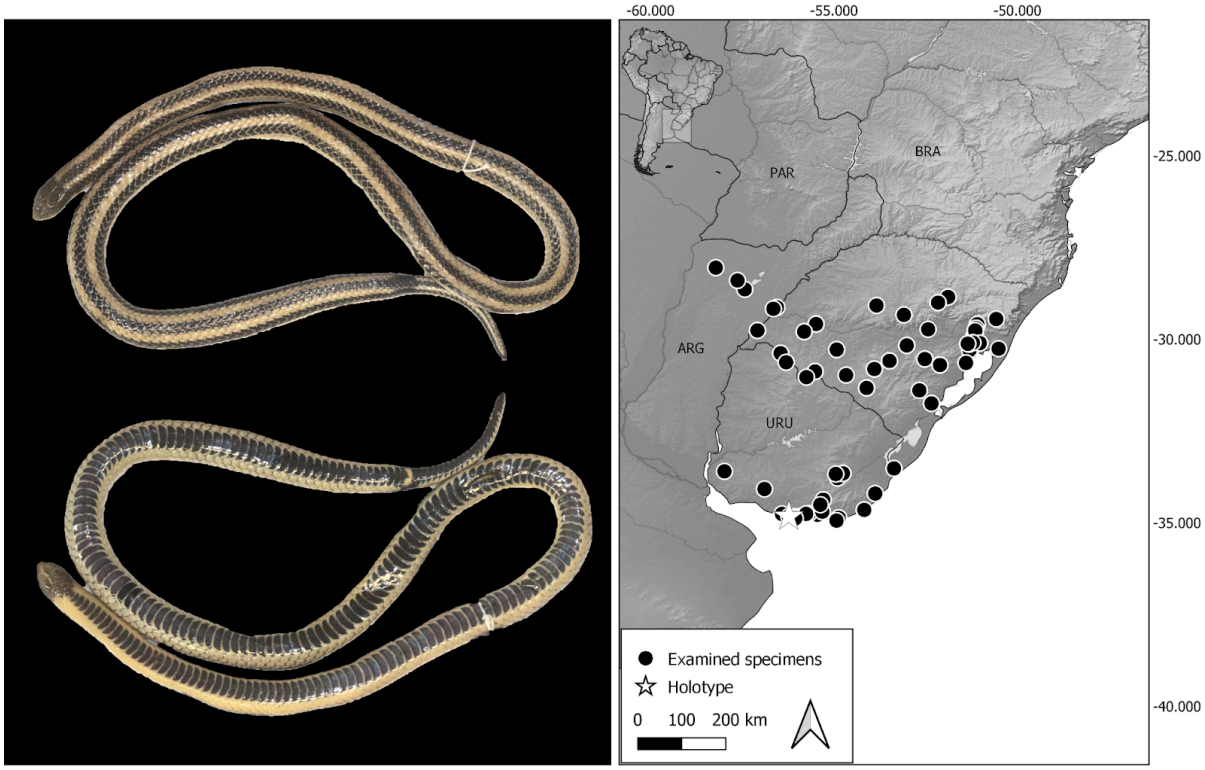


Figure 2. Overview of specimens attributed to the nomen *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris lemniscatus* (MCP13250) from Caçapava do Sul, Rio Grande do Sul, Brazil. Right: Geographic distribution of examined specimens (circles) and holotype of *Elapomorphus lemniscatus* (star).

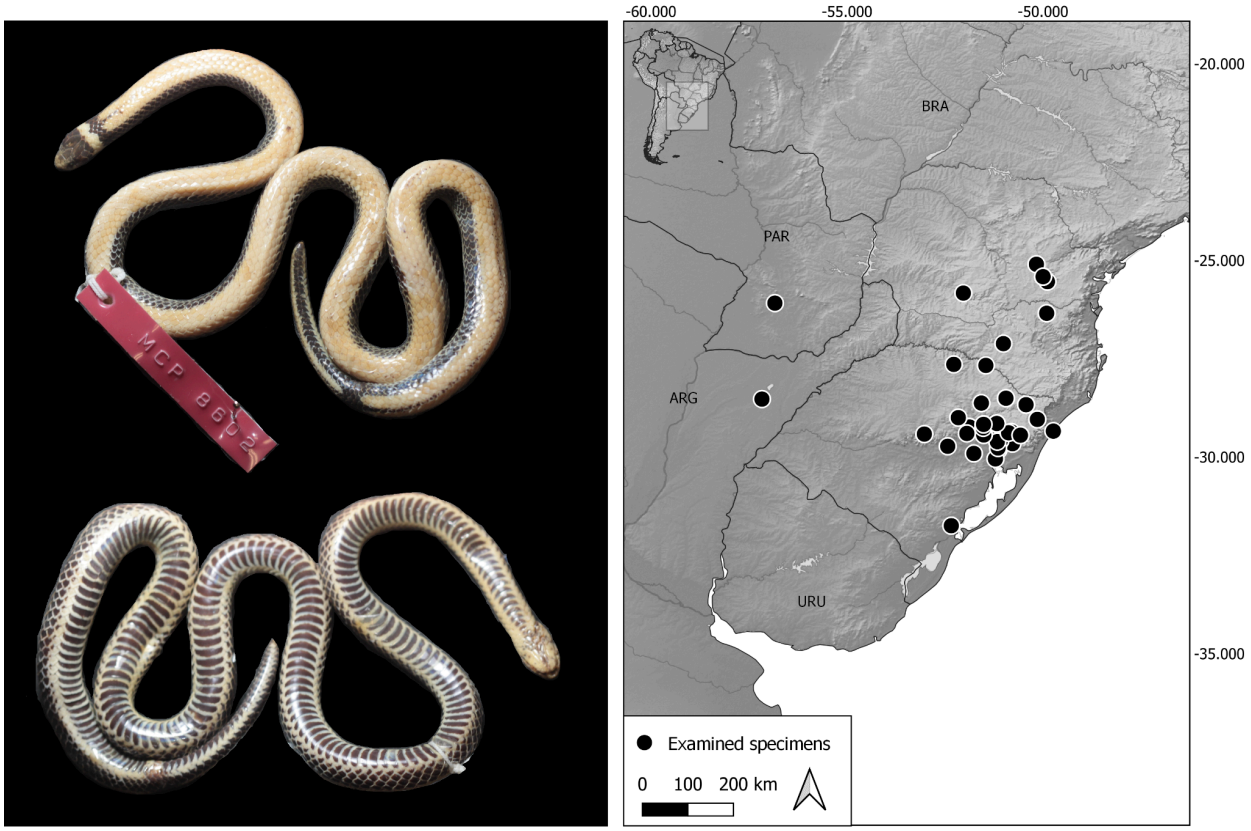


Figure 3. Overview of specimens attributed to the nomen *Elapomorphus reticulatus* Peters, 1860. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris reticulatus* (MCP 8602) from Dois Irmãos, Rio Grande do Sul, Brazil. Right: Geographic distribution of examined specimens (circles).

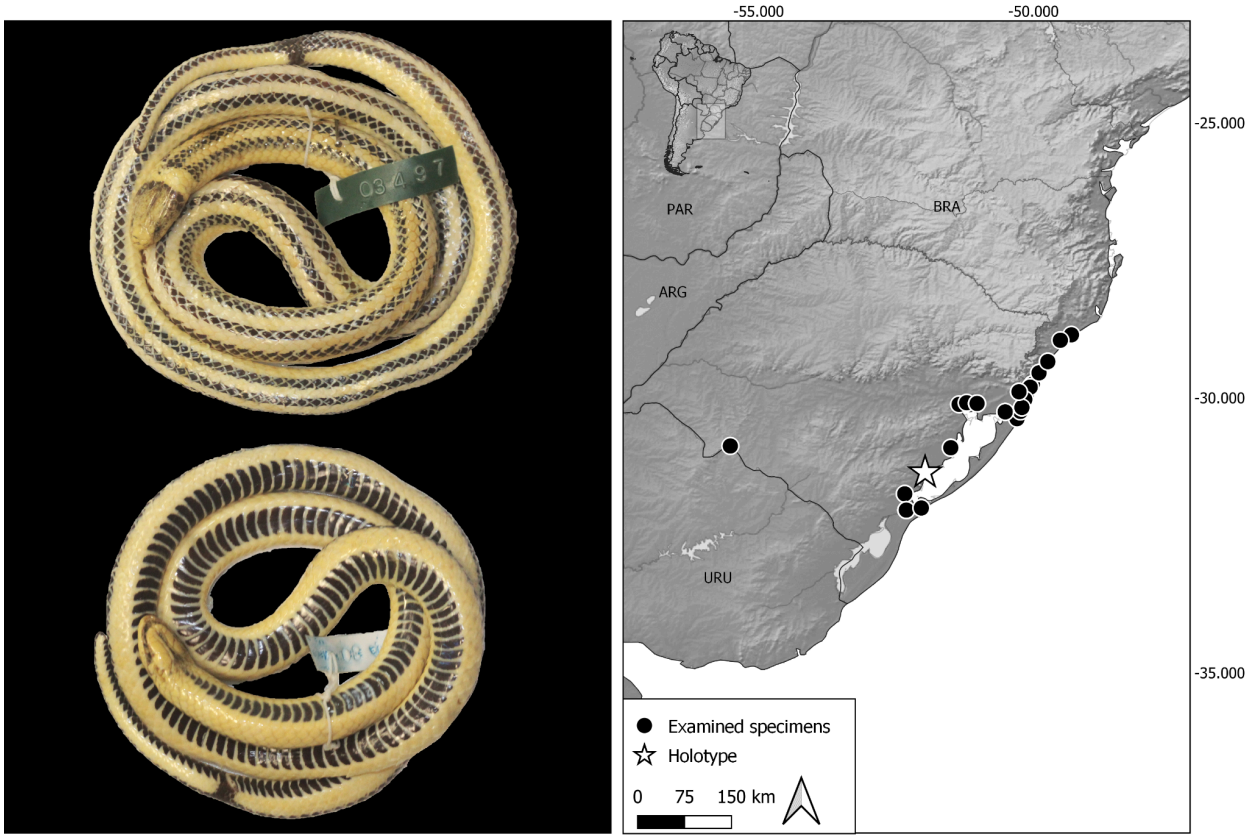


Figure 4. Overview of specimens attributed to the nomen *Elapomorphus trilineatus* Boulenger, 1889 is the earliest available. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris trilineatus* (UFRGS 3497) from Arambaré, Rio Grande do Sul, Brazil. Right: Geographic distribution of examined specimens (circles) and holotype of *Elapomorphus trilineatus* (star).

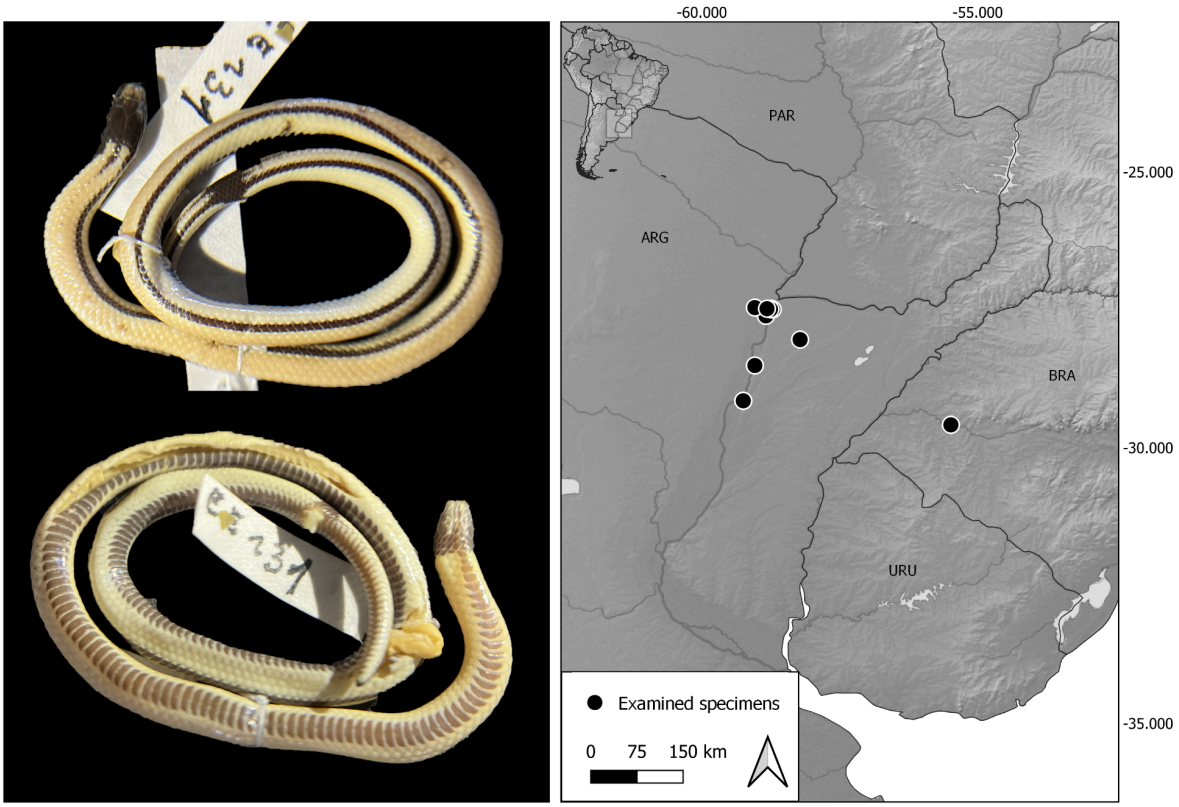


Figure 5. Overview of specimens attributed to the nomen *Elapomorphus bilineatus* Duméril, Bibrón, & Duméril 1854. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris bilineatus* (UNNEC 07678) from Laguna Brava, Corrientes, Argentina. Right: Geographic distribution of examined specimens (circles).

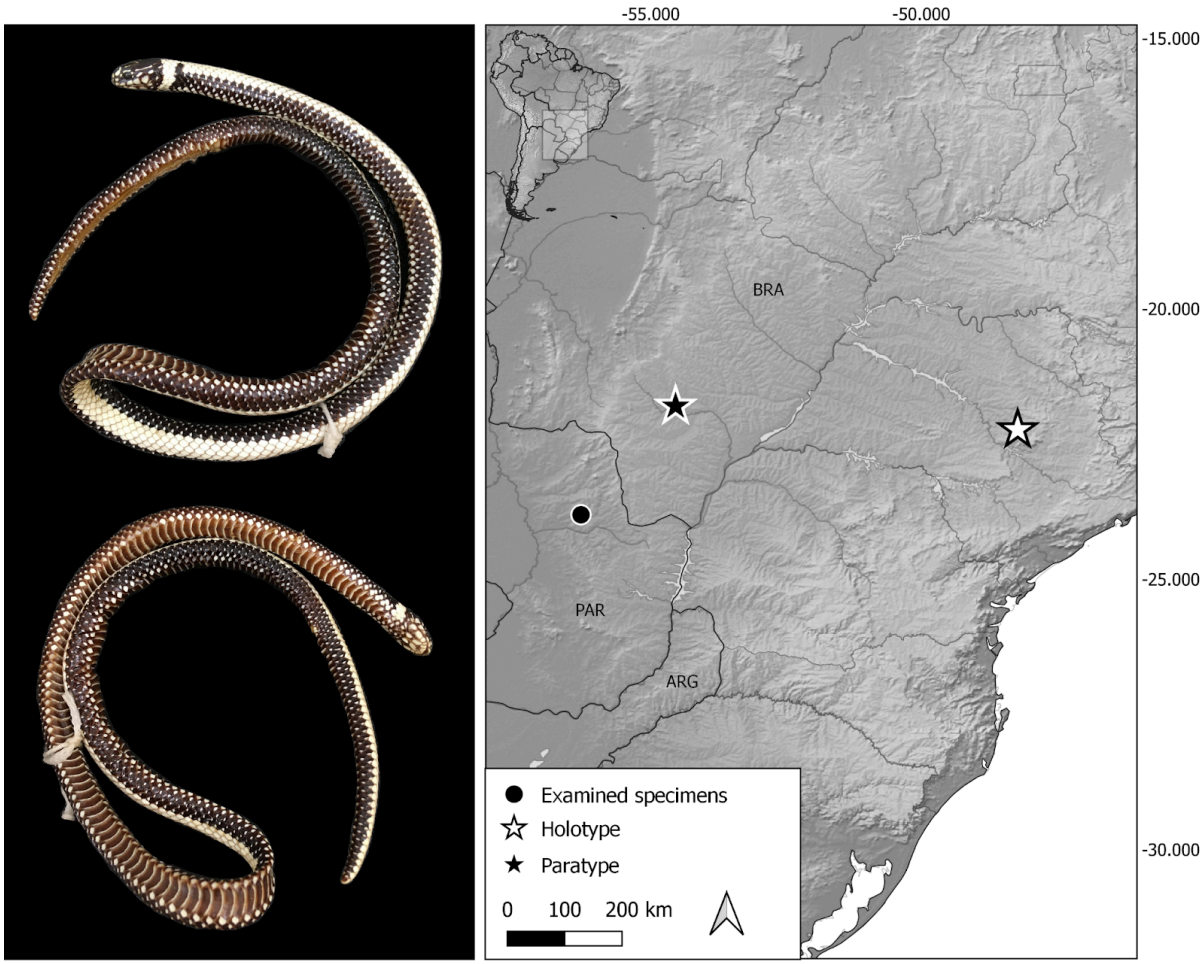


Figure 6. Overview of specimens attributed to the nomen *Phalotris multipunctatus* Puorto & Ferrarezzi, 1994. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris multipunctatus* (IBSP 43635) from Brotas, São Paulo, Brazil. Right: Geographic distribution of examined holotype (star).

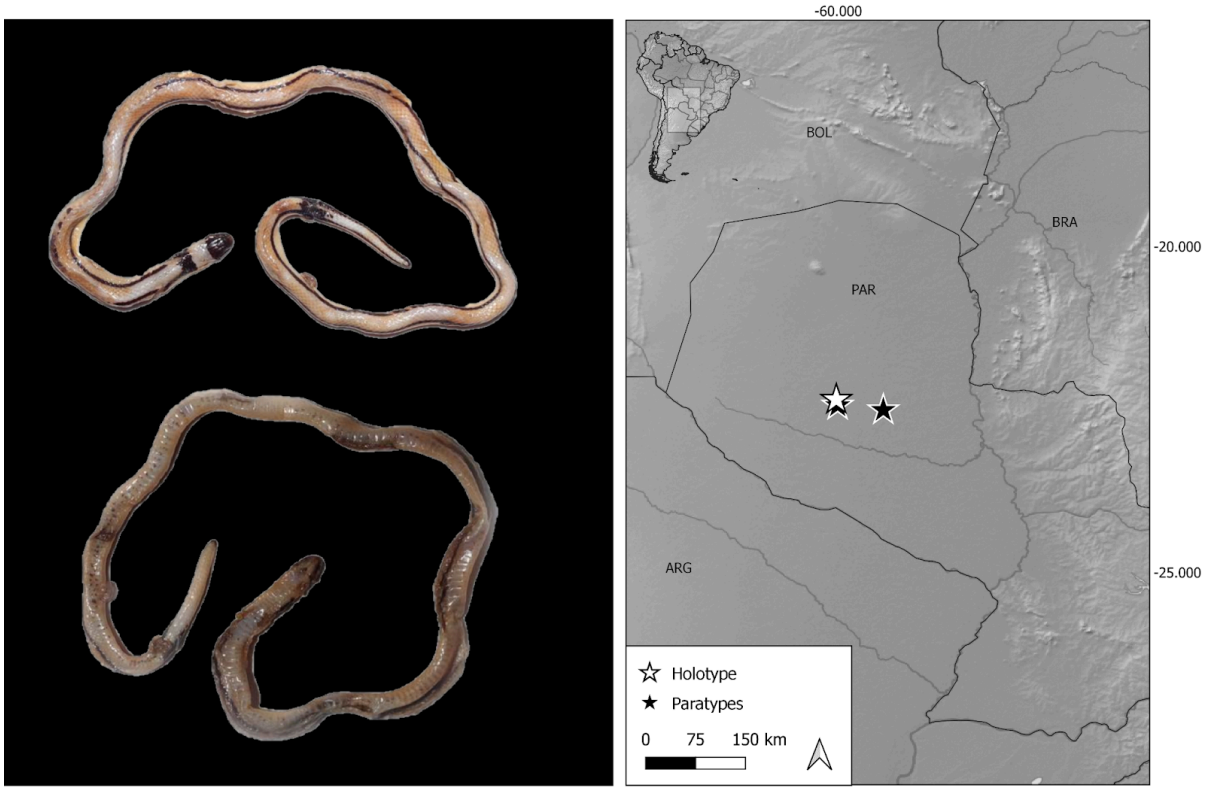


Figure 7. Overview of specimens attributed to the nomen *Phalotris normanscotti* Cacciali & Cabral, 2015. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris normanscotti* (MNHNP 5160) from Filadelfia, Boquerón, Paraguay. Right: Geographic distribution of examined holotype (star).

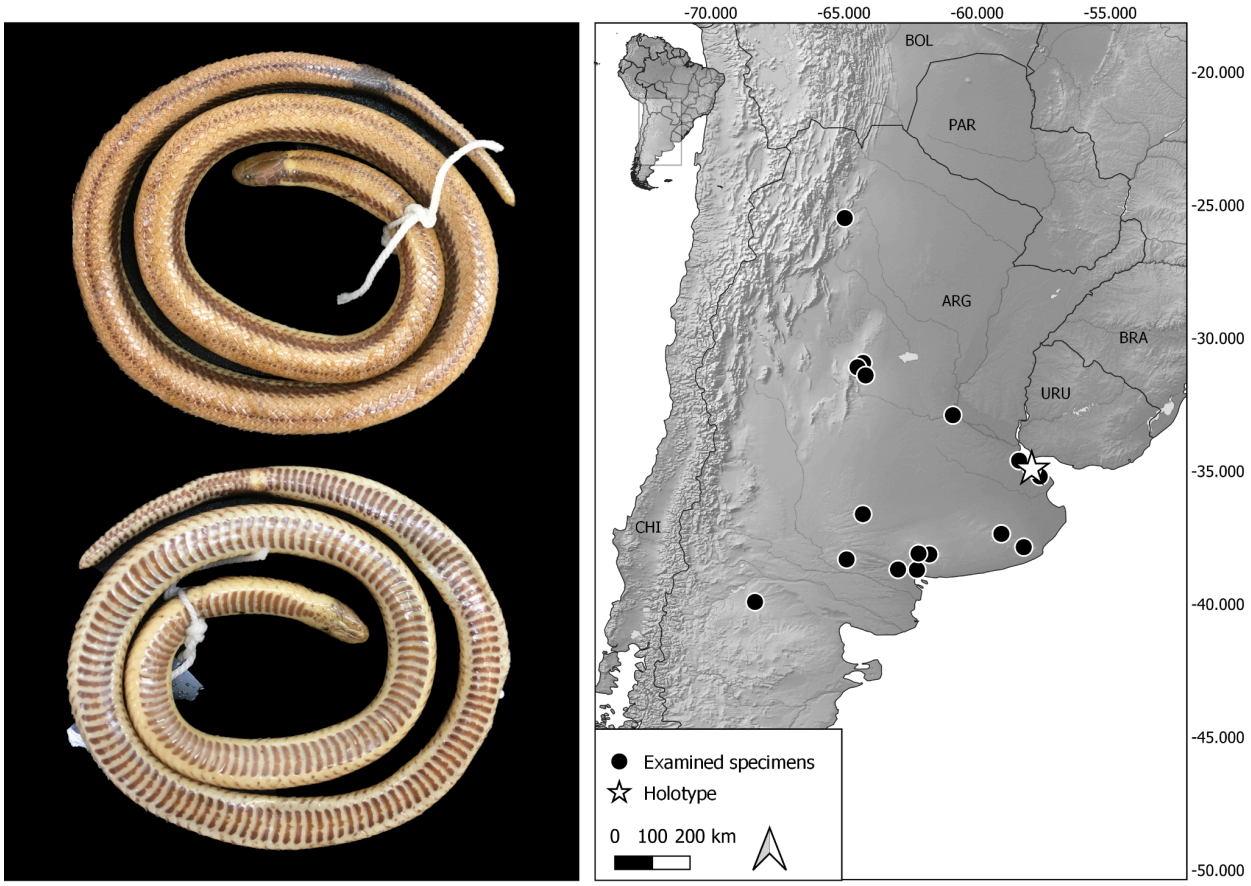


Figure 8. Overview of specimens attributed to the nomen *Elapomorphus spegazzinii* Boulenger, 1913. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris spegazzinii* (MCN 8884) from Buenos Aires, Buenos Aires, Argentina. Right: Geographic distribution of examined specimens (circles) and holotype of *Elapomorphus spegazzinii* (star).

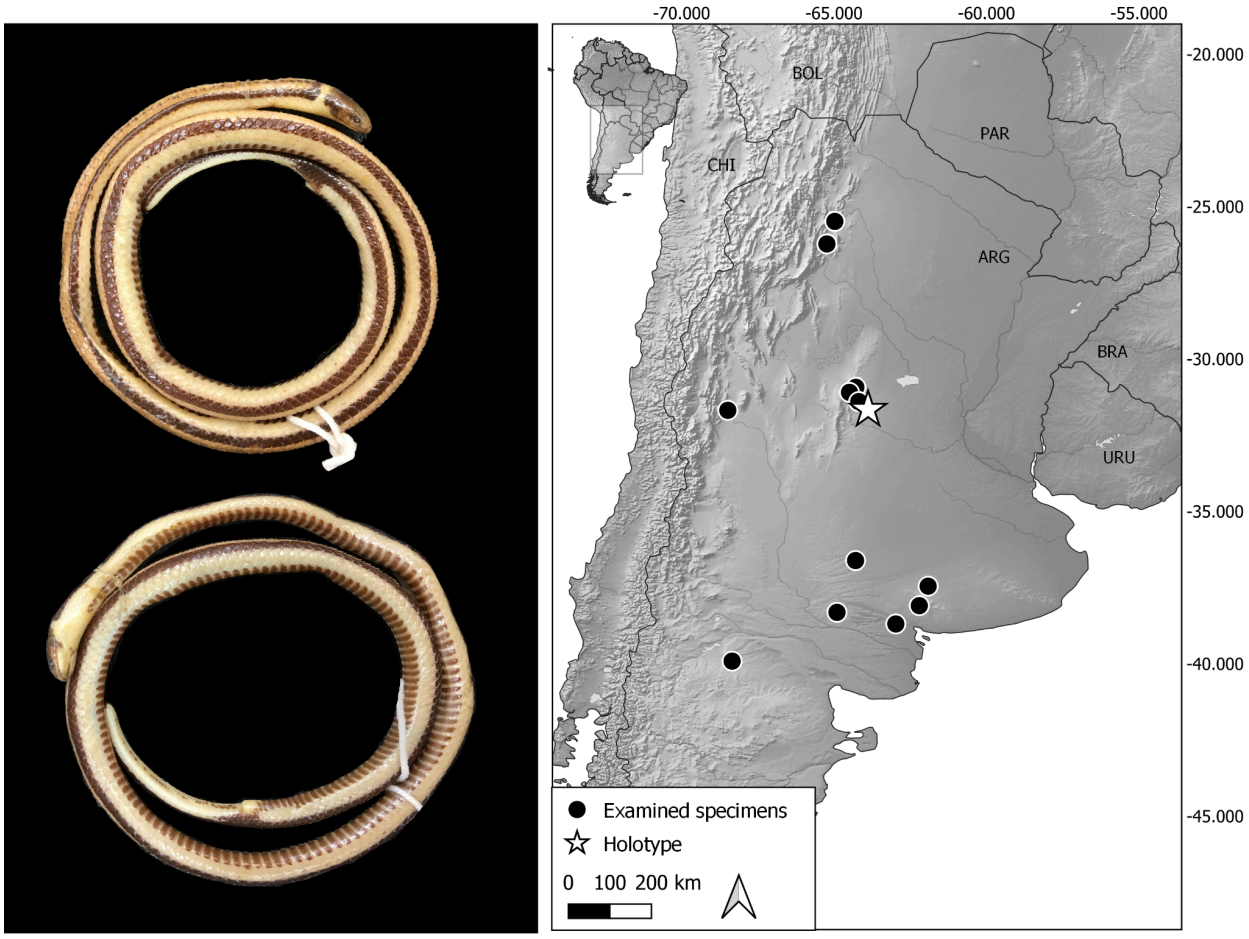


Figure 9. Overview of specimens attributed to the nomen *Elapomorphus suspectus* Amaral, 1924. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris suspectus* (MCN 7125) from Córdoba, Córdoba, Argentina. Right: Geographic distribution of examined specimens (circles) and holotype of *Elapomorphus suspectus* (star).

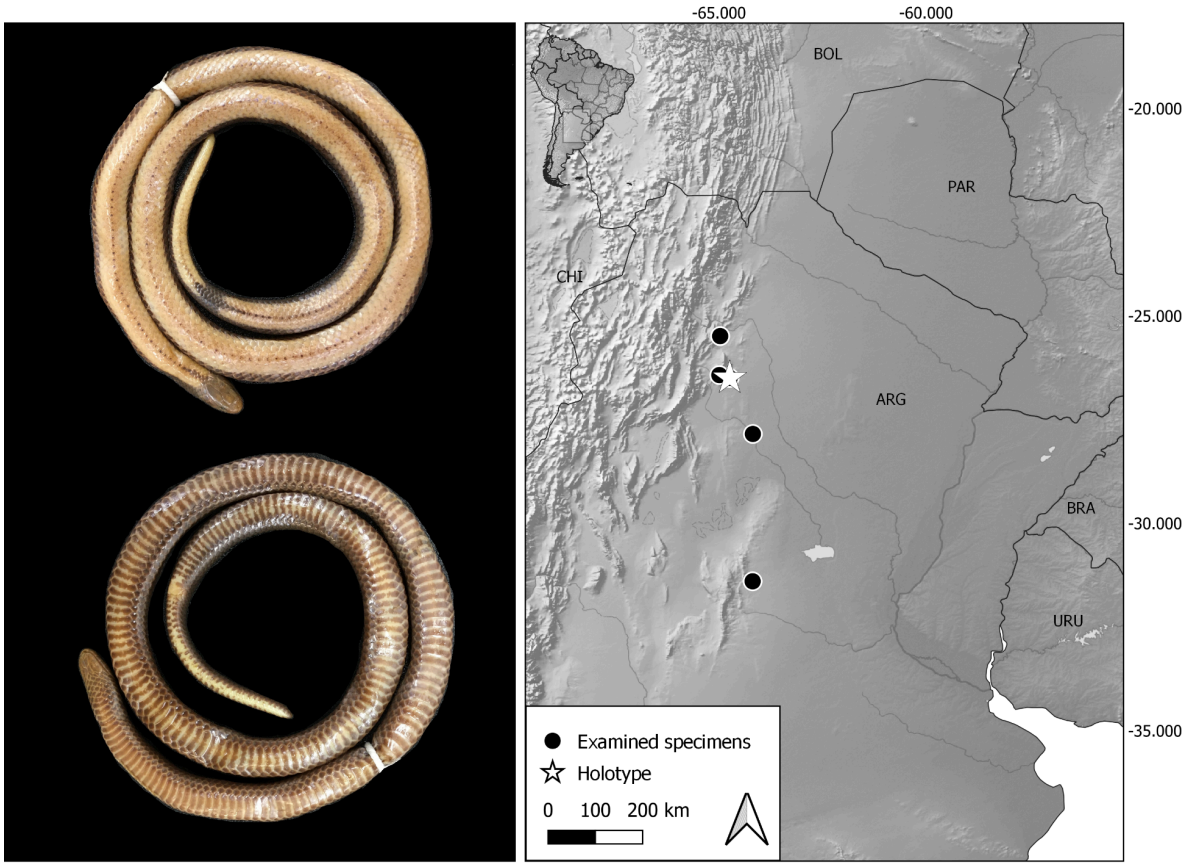


Figure 10. Overview of specimens attributed to the nomen *Phalotris illustrator* Scrocchi, Nenda & Giraudo, 2022 is the earliest available. Left: from top to bottom, dorsal and ventral views of adult female *Phalotris illustrator* (MCN 6734) from Córdoba, Córdoba, Argentina. Right: Geographic distribution of examined specimens (circles) and holotype of *Phalotris illustrator* (star).

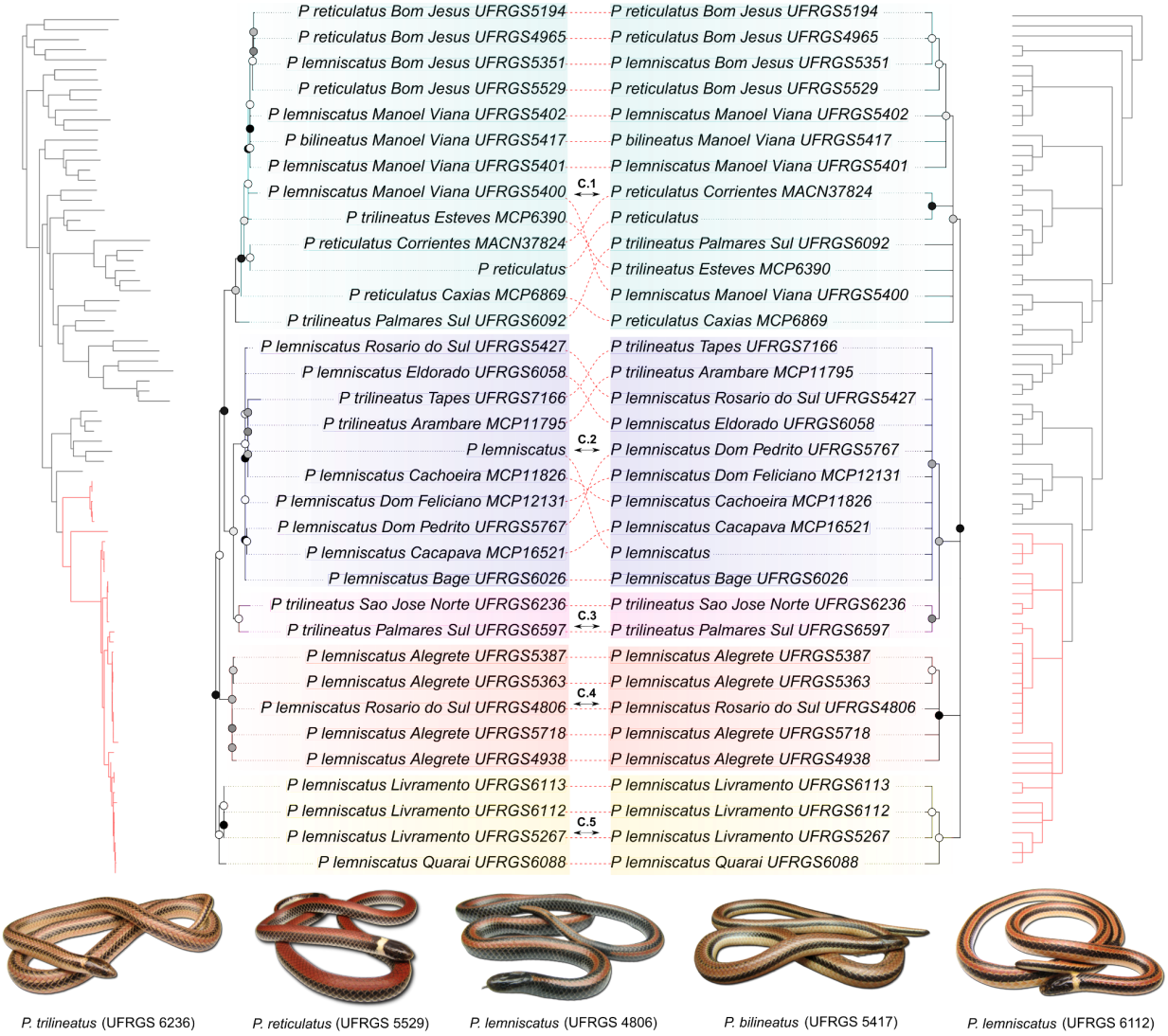


Figure 11. Phylogenetic relationships of *Phalotris bilineatus* species group, from four concatenated nuclear and mitochondrial DNA gene fragments, as inferred by a Maximum Likelihood (RAxML; left) and a Maximum Parsimony (TNT; right) frameworks. Node labels are: white (> 70% bootstrap support); gray (< 70% bootstrap support); black (100% bootstrap support). Inset photographs represent sampled *Phalotris bilineatus* species group specimens in life.

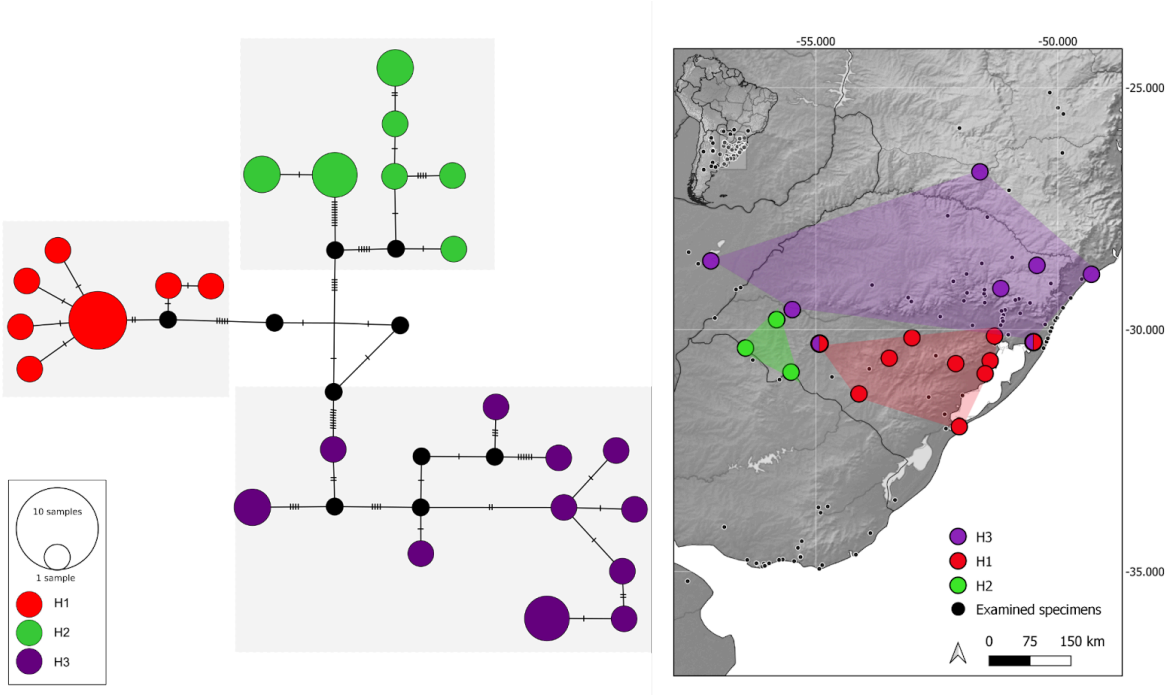
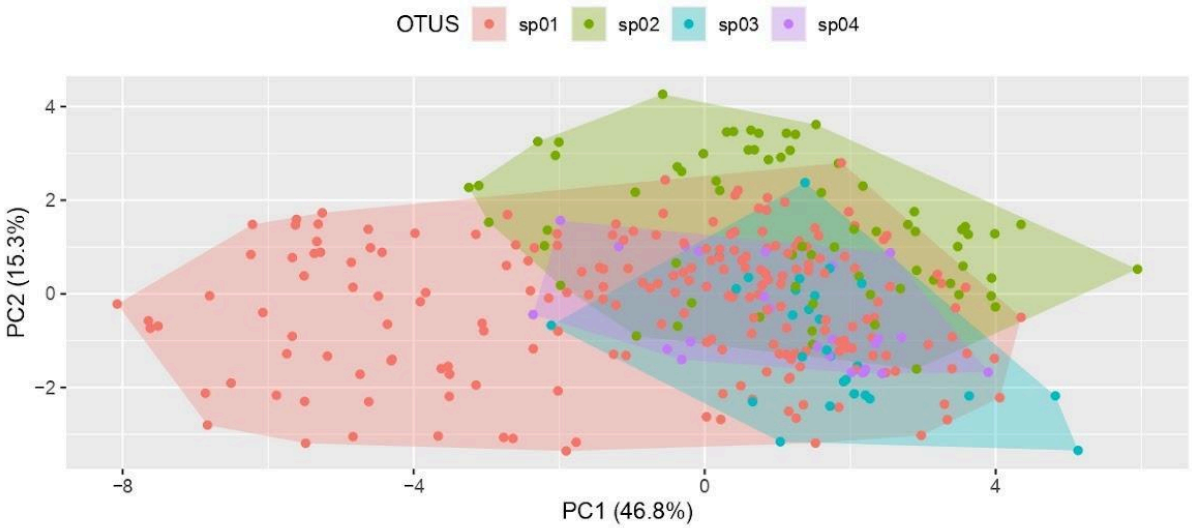
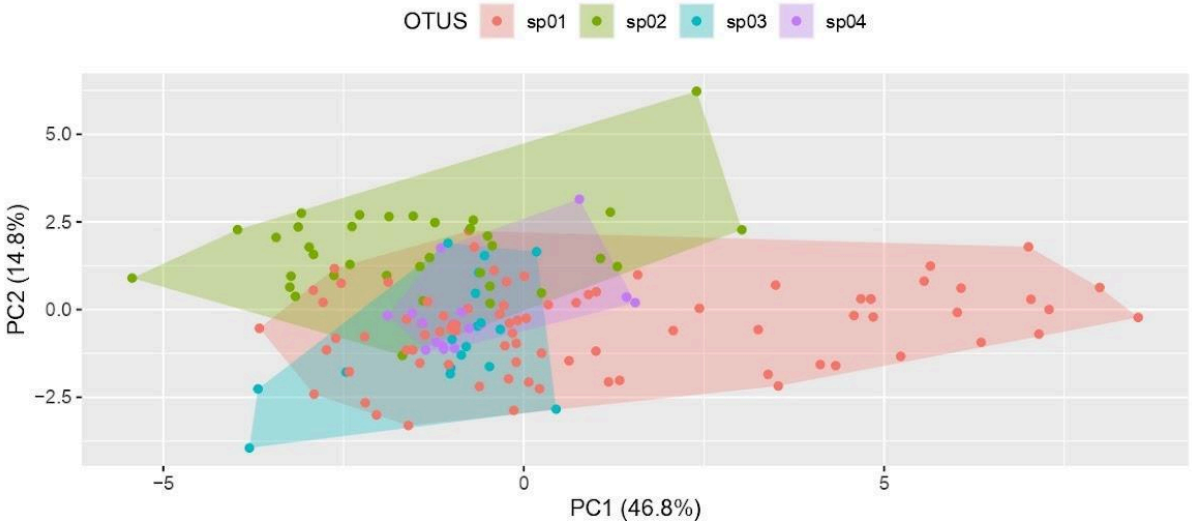


Figure 12. Phylogeography of the *Phalotris bilineatus* species group, based on the small Cytochrome b (*CYTB*) gene fragment, haplotype network (A), and geographic distribution of haplogroups by genetic samples (B). Hashes indicate mutations.


OTUS
sp01
sp02
sp03
sp04
PC2 (15.3%)
PC1 (46.8%)


OTUS
sp01
sp02
sp03
sp04
PC2 (14.8%)
PC1 (46.8%)

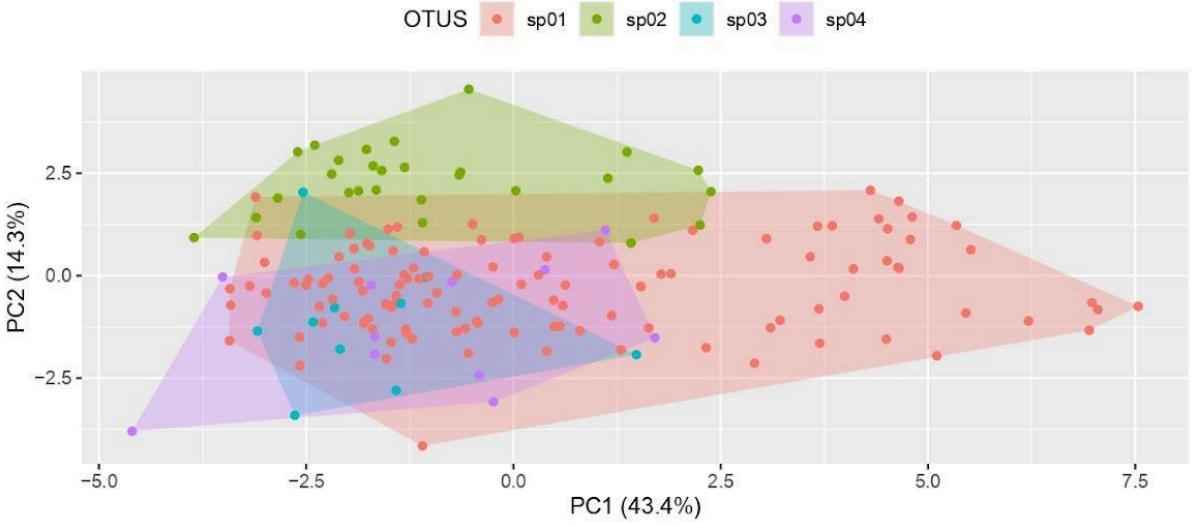
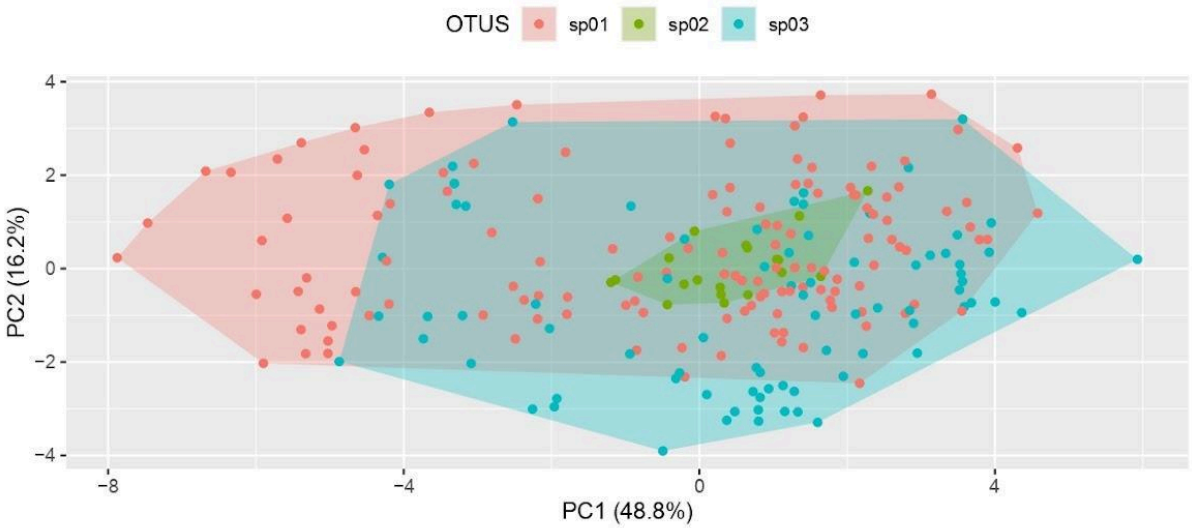
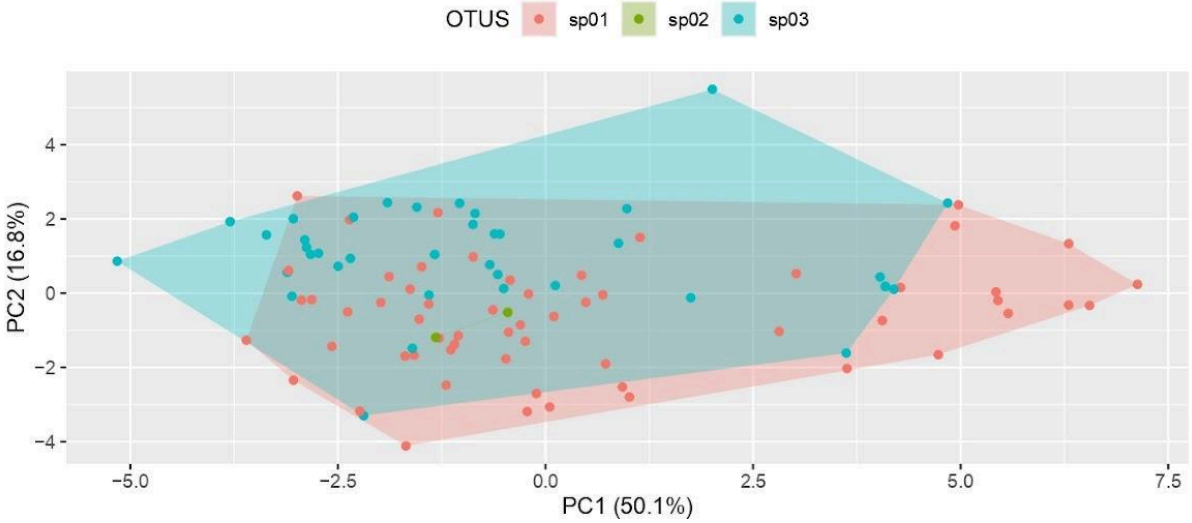

OTUS
sp01
sp02
sp03
sp04
PC2 (14.3%)
PC1 (43.4%)

Figure 13. Bidimensional scatterplots showing the results from the multivariate analyses of the *Phalotris bilineatus* species group (Morphological Operational Taxonomic Units). First two components of PCA (Principal Component Analysis) for all individuals (top), females (middle), and males (bottom).


OTUS
sp01
sp02
sp03
PC2 (16.2%)
PC1 (48.8%)


OTUS
sp01
sp02
sp03
PC2 (16.8%)
PC1 (50.1%)

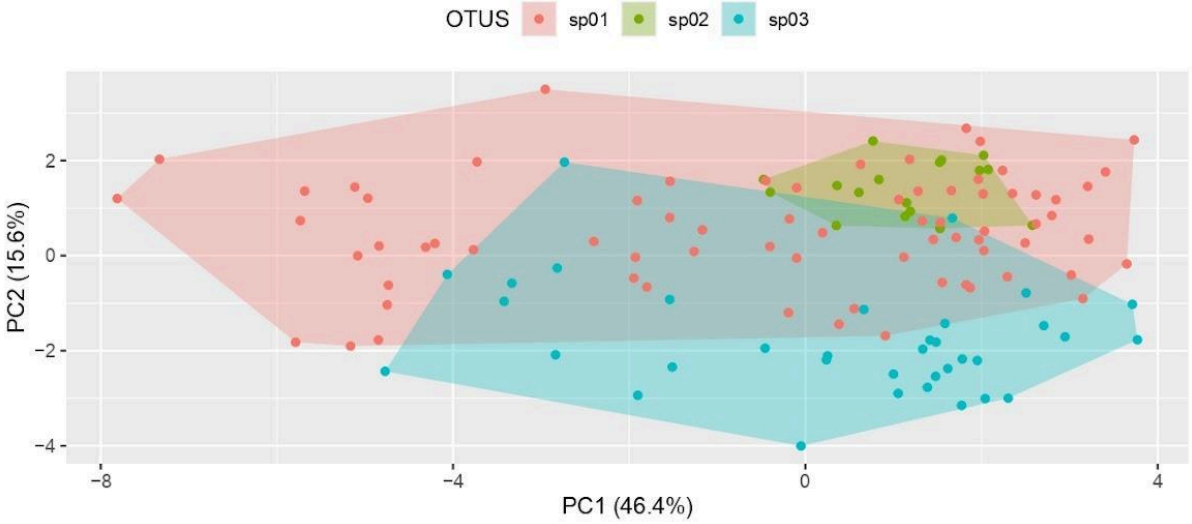
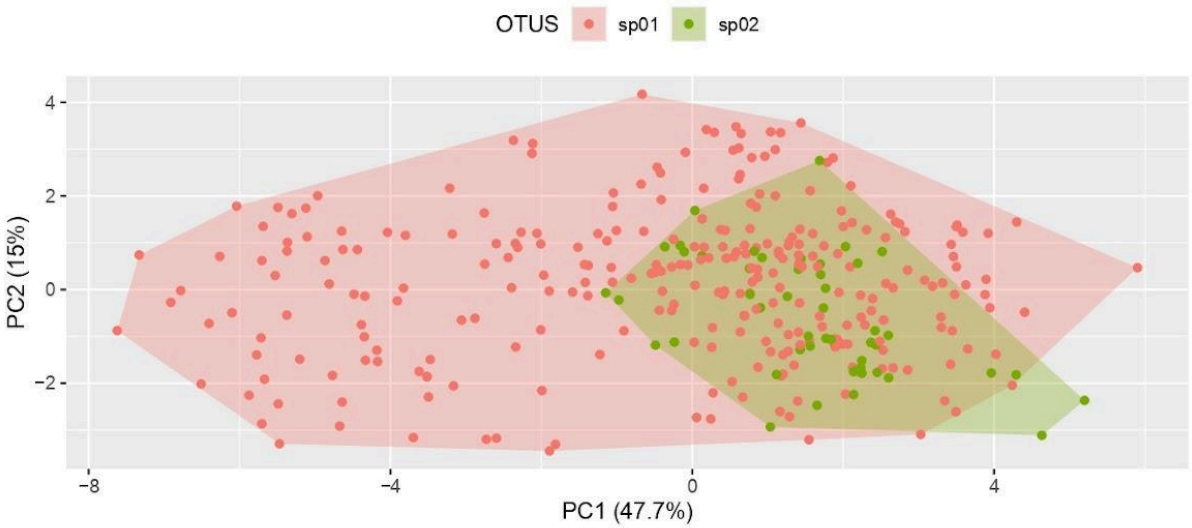
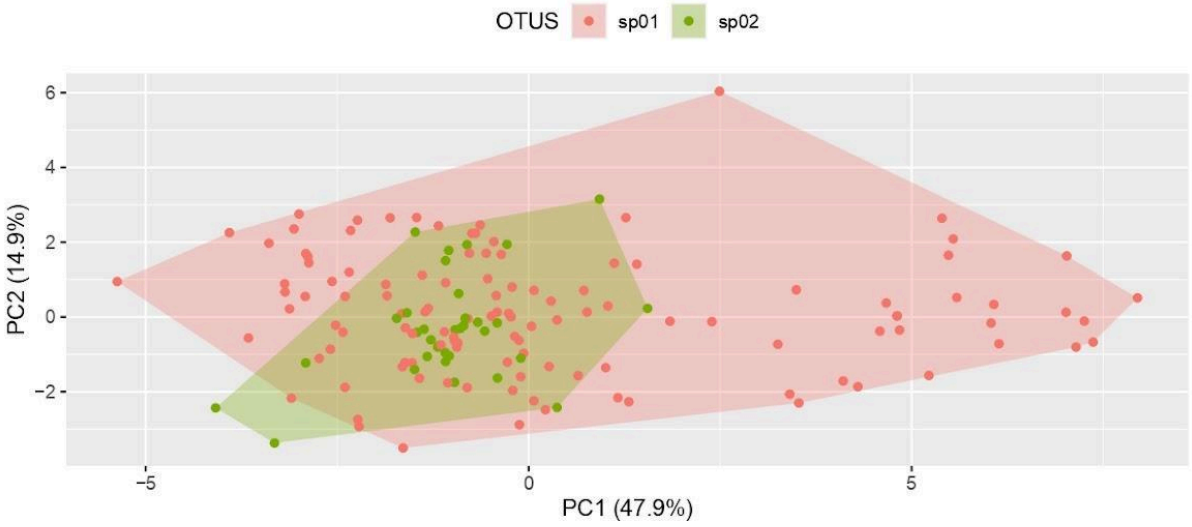

OTUS
sp01
sp02
sp03
PC2 (15.6%)
PC1 (46.4%)

Figure 14. Bidimensional scatterplots showing the results from the multivariate analyses of the *Phalotris bilineatus* species group (Molecular Operational Taxonomic Units). First two components of PCA (Principal Component Analysis) for all individuals (top), females (middle), and males (bottom).


OTUS
sp01
sp02
PC2 (15%)
PC1 (47.7%)


OTUS
sp01
sp02
PC2 (14.9%)
PC1 (47.9%)

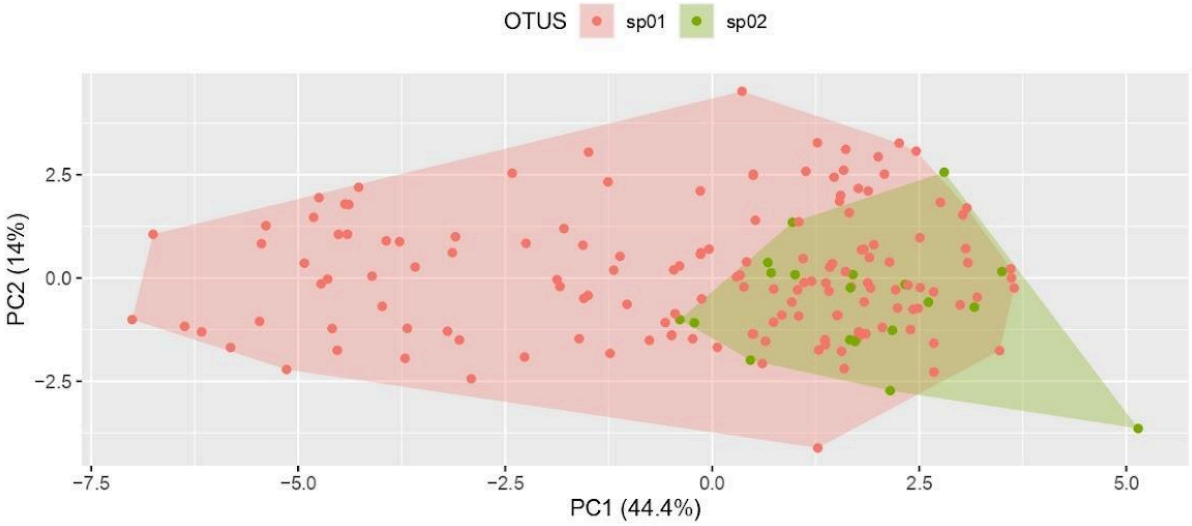

OTUS
sp01
sp02
PC2 (14%)
PC1 (44.4%)

Figure 15. Bidimensional scatterplots showing the results from the multivariate analyses of the *Phalotris bilineatus* species group (Geographical Operational Taxonomic Units). First two components of PCA (Principal Component Analysis) for all individuals (top), females (middle), and males (bottom).

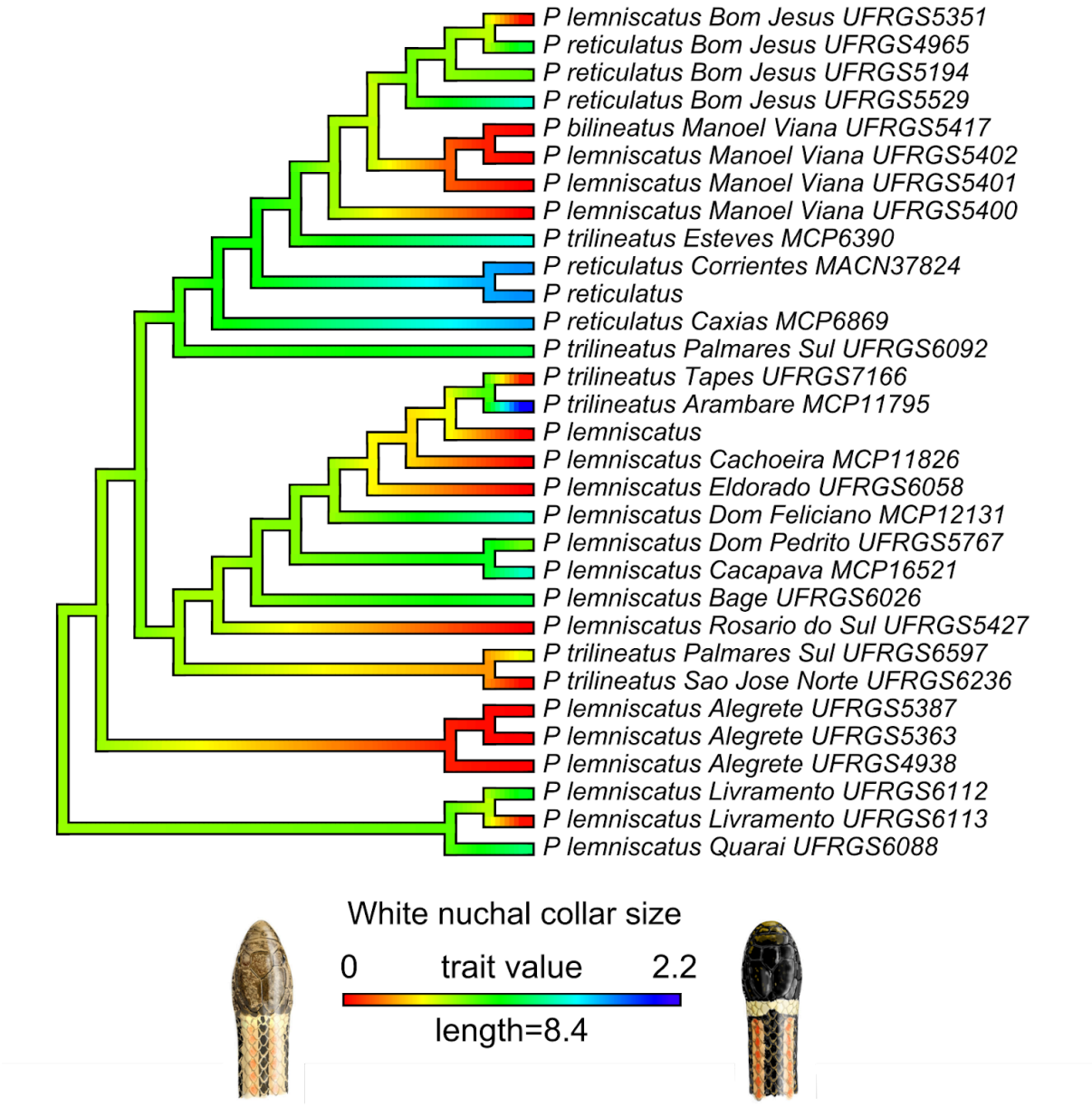
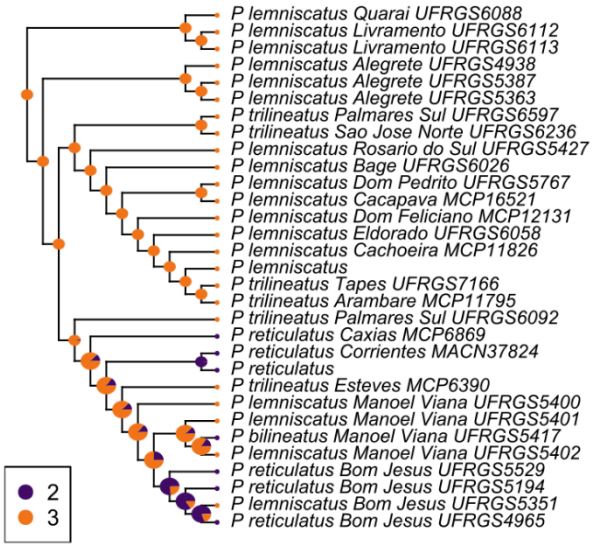
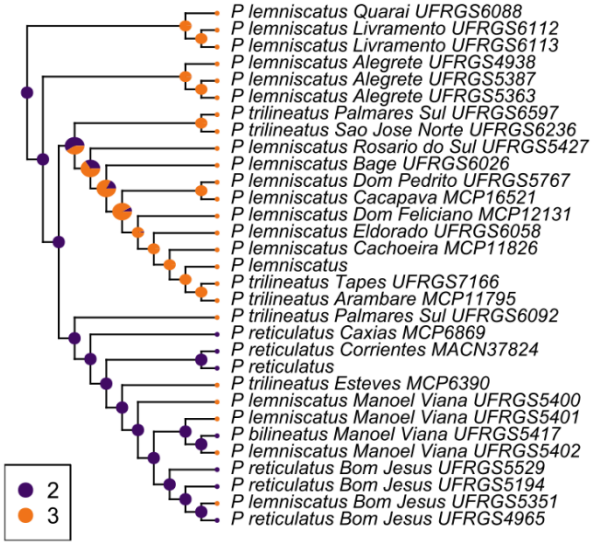


Figure 16. Continuous character mapping of white nuchal collar size (in mm) for the *Phalotris bilineatus* species group, over concatenated mitochondrial and nuclear ML topology.

a) ER model marginal ancestral states



b) ARD model marginal ancestral states



c) SYM model marginal ancestral states

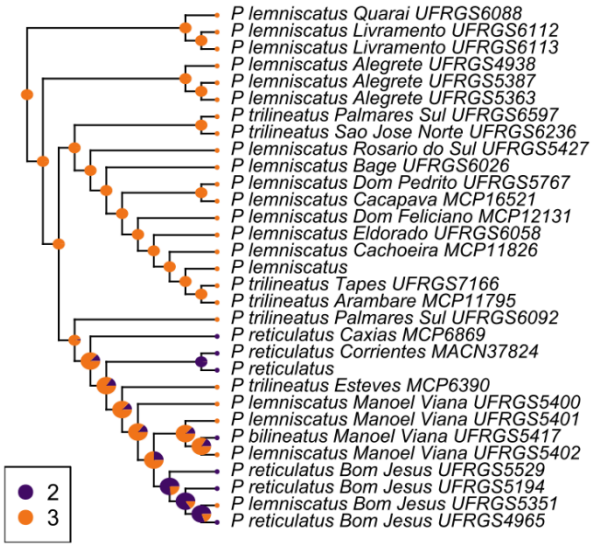


Figure 17. Marginal ancestral character estimation for number of dorsal stripes within the *Phalotris bilineatus* species group, over concatenated mitochondrial and nuclear ML topology. A) Equal Rates model; B) All Rates Different model; C) Symmetric model. Ambiguous nodes (LnL = ≥ 0.95) are represented as larger, and number of stripes are coded as colors: Two stripes (Blue); Three stripes (Orange).

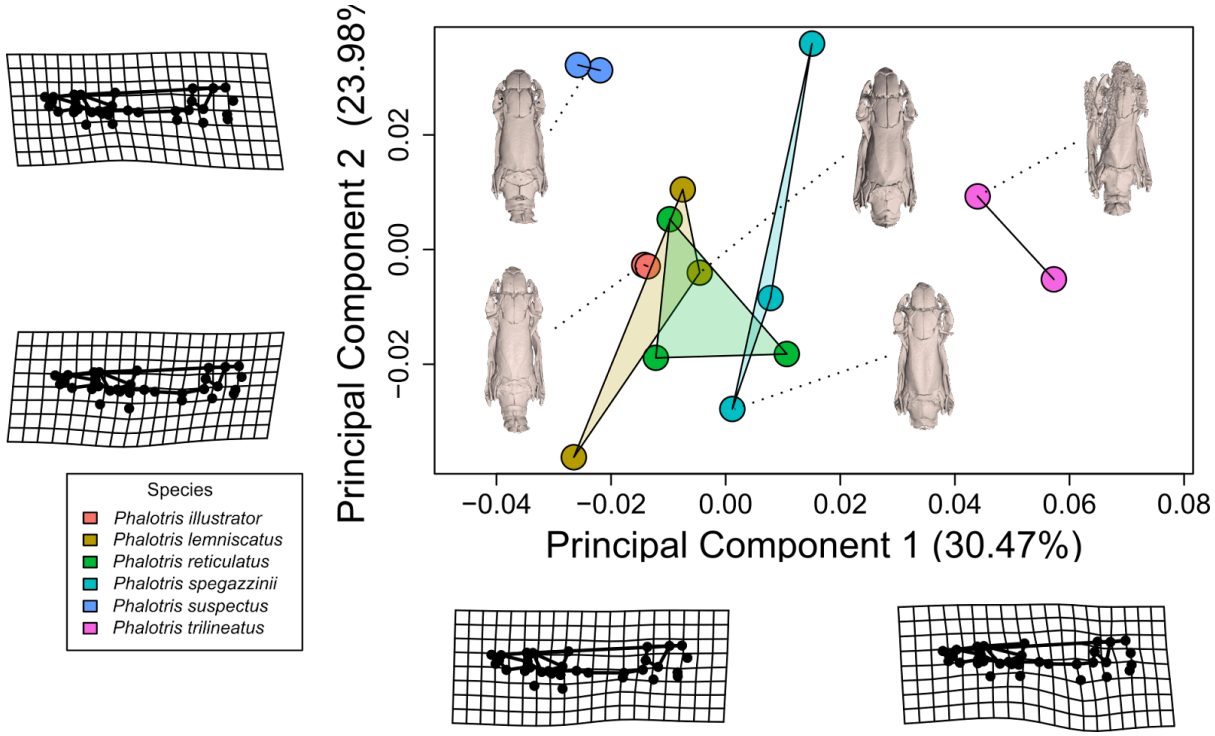


Figure 17. Principal component analysis of skull shape landmarks of the *Phalotris bilineatus* species group, as inferred over the generalized Procrustes analysis. Inset thin plate splines represent each maximum and minimum skull deformation in each principal component axis.

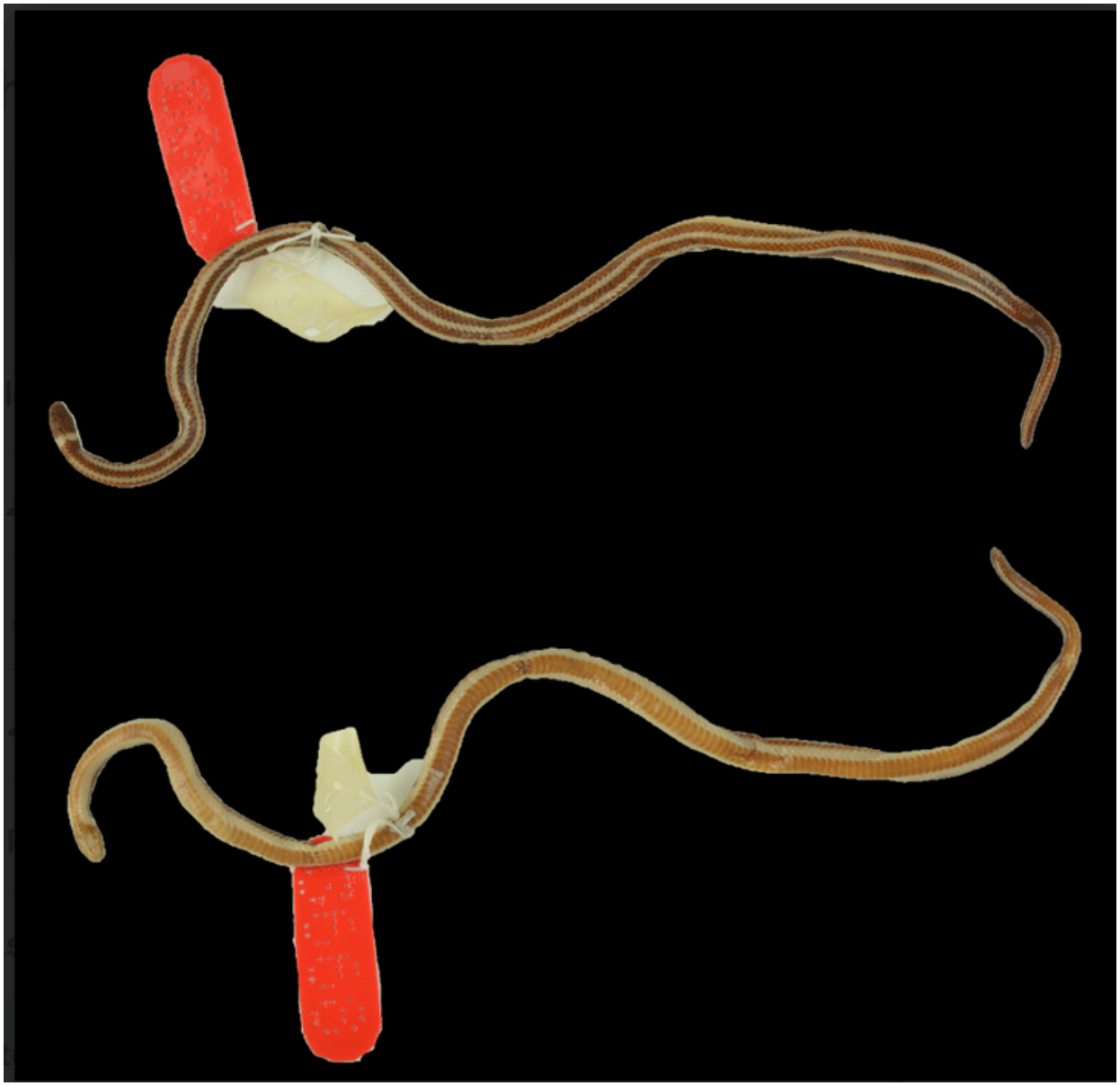


Figure 18. Holotype of *Elapomorphus lemniscatus* Duméril, Bibrón, & Duméril, 1854 (MNHN-RA 3668), from Montevideo, Departamento Montevideo, Uruguay. Top: dorsal view; bottom: ventral view.

Figure 19. Holotype of *Elapomorphus reticulatus* Peters, 1860 (ZMB 3811), from "Brazil", estimated here to be likely from Santa Catarina, Brazil. Top: dorsal view; bottom: ventral view.

Figure 20. Holotype of *Elapomorphus iheringi* Strauch, 1887 (ZISP 6255), from Mundo Novo, Taquara, Rio Grande do Sul, Brazil. Top: dorsal view; bottom: ventral view.

Figure 21. Holotype of *Elapomorphus trilineatus* Boulenger, 1889(BMNH 1946.1.2.90), from Rio Camaquã, São Lourenço do Sul, Rio Grande do Sul, Brazil. Top: dorsal view; bottom: ventral view.

Figure 22. Known morphological variation of *Phalotris lemniscatus* in lateral, dorsal, and ventral views. Top: “red phase”, morphological populations formerly attributed to *P. reticulatus*;

Middle: “striped phase”, morphological populations formerly attributed to *P. lemniscatus*;

Bottom: “orange phase”, morphological populations formerly attributed to *P. trilineatus*.

Tables

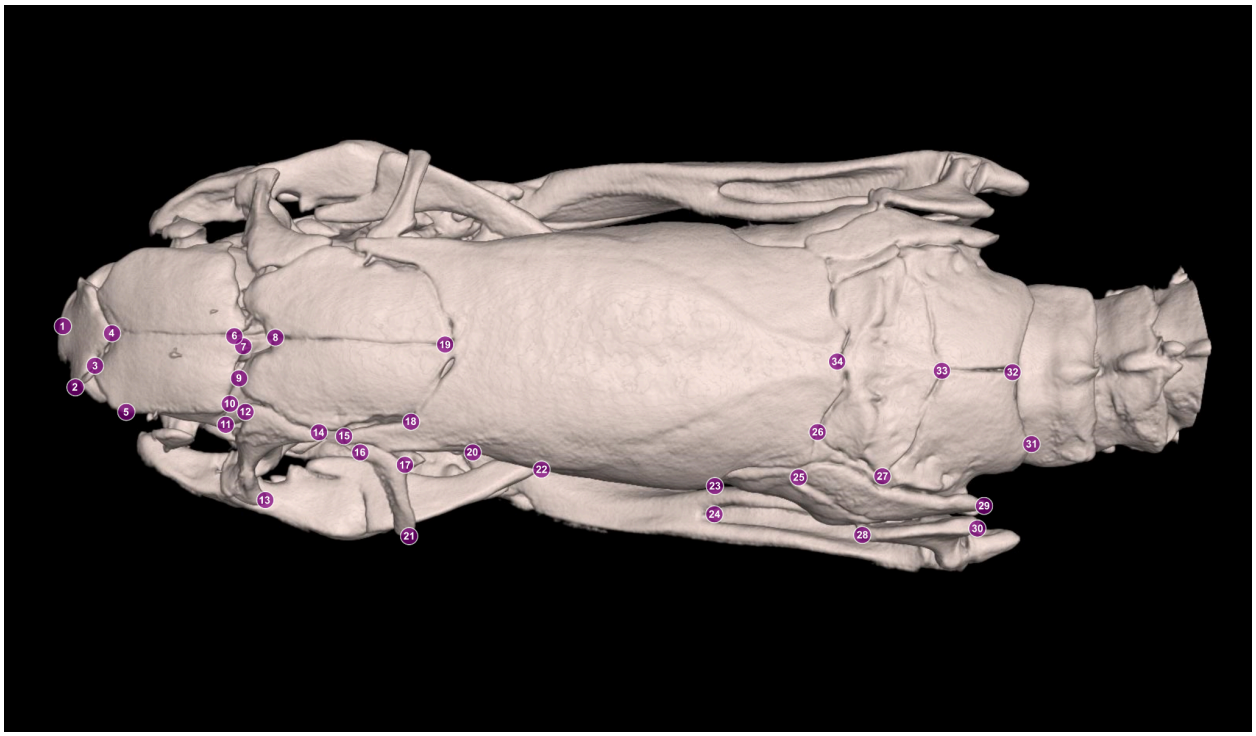
Table 1. Primers used in this study to amplify gene fragments for Sanger sequencing technique.

Gene	Primer	Sequences	Reference
12S rRNA	L1091mod	5’ CAAACTAGGATTAGAT ACCCTACTAT 3’	Kocher et al. (1989)
	H1557mod	5’ GTACRCTTACCWTGT TACGACTT 3’	Knight and Mindell (1994)
16S rRNA	L2510mod(16sar)	5’ CCGACTGTTTAMCAA AAACA 3’	Palumbi et al. (1991)
	H3056mod(16Sbr)	5’ CTCCGGTCTGAACTC AGATCACGTRGG 3’	Palumbi et al. (1991)
CYTB	703Botp.mod	5’ TCAAAYATCTCAACCT GATGAAAYTTYGG 3’	Pook et al. (2000)
	MVZ16p.mod	5’ GGCAAATAGGAAGTA TCAYTCTGGYTT 3’	Pook et al. (2000)
BDNF	Bdnff	5’ GACCATCCTTTTCCTK ACTATGGTTATTTCAT ACTT 3	Noonan and Chippindale (2006)
	Bdnfr	5’ CTATCTTCCCCTTTTA ATGGTCAGTGTACAA AC 3’	Noonan and Chippindale (2006)

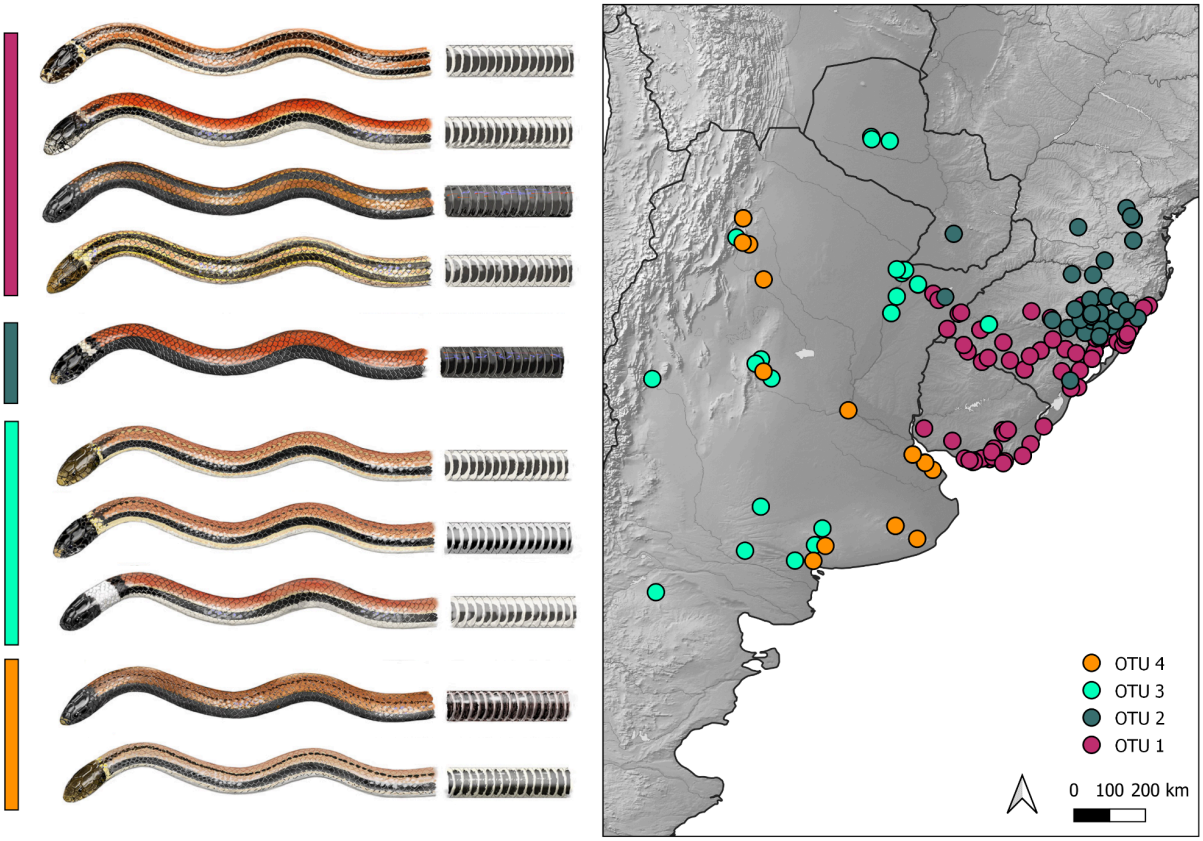
Table 2. Overview of evaluated ancestral character estimation model statistical values. Models: ER = Equal Rates; ARD = All Rates Different; SYM = Symmetric. Statistical analyses: Log(L)= Log-Likelihood; AIC = Akaike Information Criterion; AICw = Weighted Akaike Information Criterion.

Model	Log(L)	AIC	AICw
ER	-16.8168140195602	35.6336280391203	0.110809726246511
ARD	-13.8674130318253	31.7348260636506	0.778380547506978
SYM	-16.8168140195602	35.6336280391203	0.110809726246511

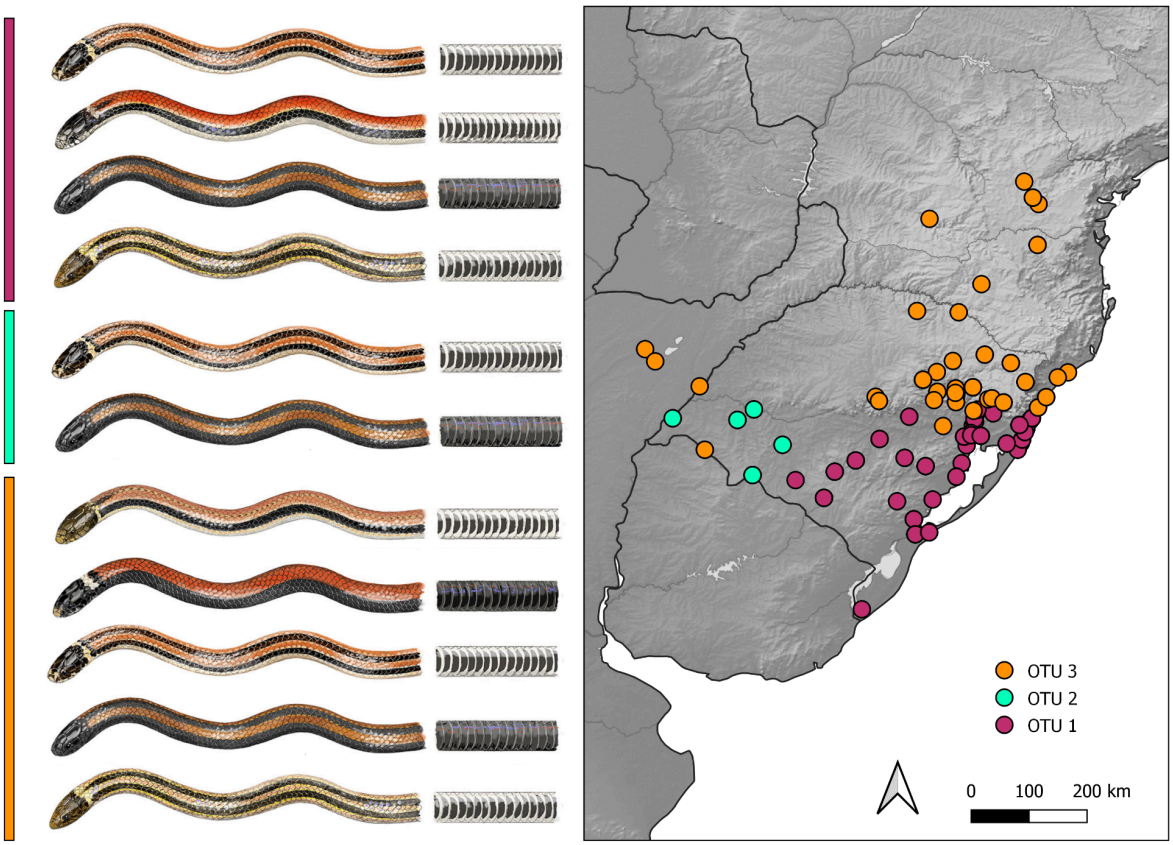
Supplementary Figures



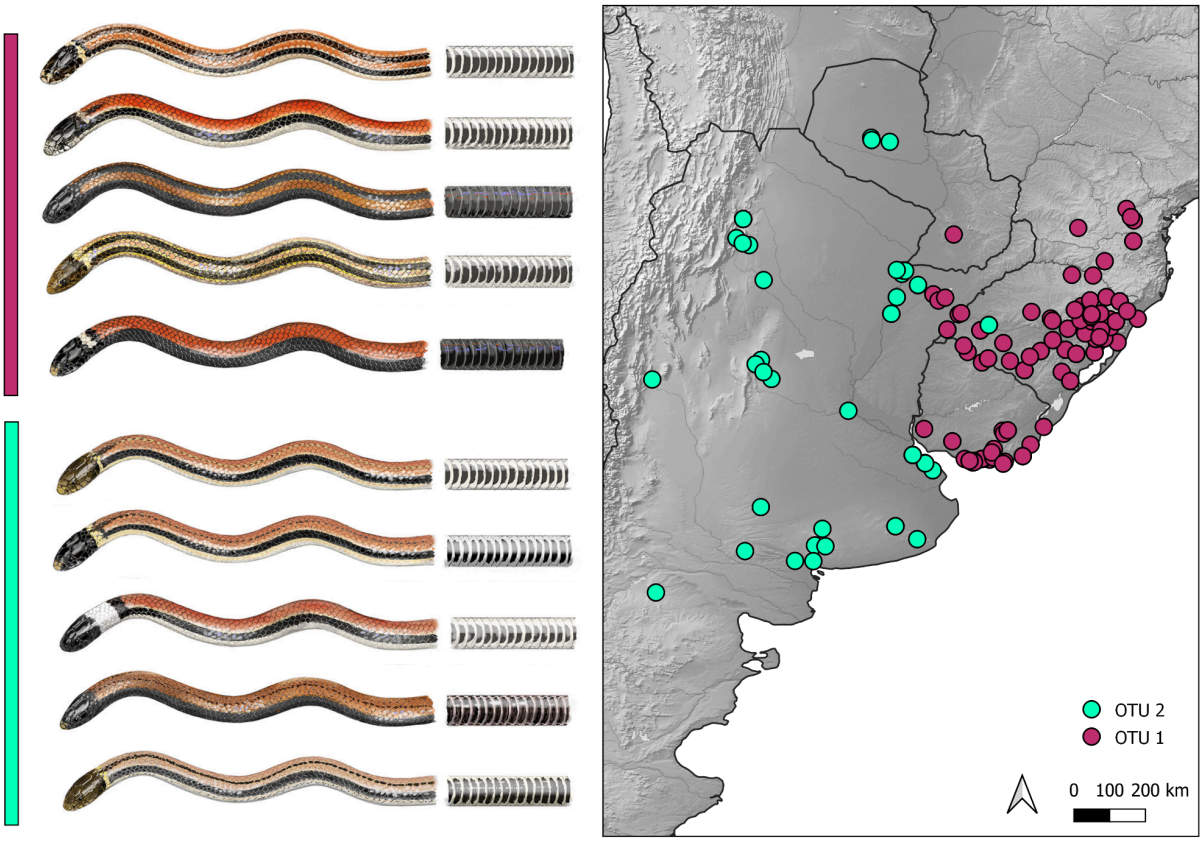
Supplementary Figure 1. Sequential landmarks used for skull shape analysis the *Phalotris bilineatus* species group, as inferred over the generalized Procrustes analysis.



Supplementary Figure 2. Overview of morphological-based OTU hypothesis. Left: morphology of specimens assigned to each OTU; Right: geographic distribution of each OTU.



Supplementary Figure 3. Overview of haplotype-based OTU hypothesis. Left: morphology of specimens assigned to each OTU; Right: geographic distribution of each OTU.



Supplementary Figure 4. Overview of geographical-based OTU hypothesis. Left: morphology of specimens assigned to each OTU; Right: geographic distribution of each OTU.

Supplementary Files

Supplementary File 1. Generated dataset of external morphological characters for species of the *Phalotris bilineatus* species group (.xls).

Supplementary File 2. Alignment of mitochondrial and nuclear DNA gene fragments used for phylogenetic inferences (.xls).

Supplementary File 3. Pairwise uncorrected p-distances for the Cytochrome B (CYTB) mitochondrial DNA gene fragment, as inferred by MEGA (.xls).

Supplementary File 4. Pairwise patristic distances for the Cytochrome B (CYTB) mitochondrial DNA gene fragment, as inferred by Ape package (.xls).

APPENDIX

Appendix I. R-script for PCA analysis.

```
# install.packages("nortest")
# needed for lillie.test
library(nortest)
# install.packages("MASS")
# needed for LDA and MDS
library(MASS)
# install.packages("corrplot")
# needed for corrplot
library(corrplot)
# install.packages("coin")
# needed for independence_test
# library(coin)
# install.packages("wPerm")
# needed for perm.ind.test
# library(wPerm)
# install.packages("caret")
# needed for PLSDA and RF
# library(caret)
# install.packages("ggplot2")
# needed for multivatiated plots
library(ggplot2)
# install.packages("gridExtra")
# needed to plot in grid
library(gridExtra)
# install.packages("randomForest")
# library(randomForest)
# install.packages("rfPermute")
# library(rfPermute)
# install.packages("psych")
# needed for some summary statistics
# library(psych)
# install.packages("vegan")
# needed for Permanova
# library(vegan)
# install.packages("RVAideMemoire")
# needed for pairwise Post-hoc Permanova and PLSDA.VIP
# library(RVAideMemoire)
# install.packages("doSNOW")
# needed to run CV in parallel
# library(doSNOW)

########################################
# READ AND SORT DATA
########################################
```

```
# read .csv file
x<-read.csv("dataset_20231014.csv",header=T)

########################################

# sort variables
x_disc<-x[,grep("disc",colnames(x))]
x_cont<-x[,grep("cont",colnames(x))]

########################################

# create directories for results
dir.create("./00_scatterplot",showWarnings = FALSE)
dir.create("./01_tables",showWarnings = FALSE)
dir.create("./02_histograms",showWarnings = FALSE)
dir.create("./03_QQplots",showWarnings = FALSE)
dir.create("./04_boxplots",showWarnings = FALSE)
dir.create("./05_correlograms",showWarnings = FALSE)
dir.create("./06_PCAs",showWarnings = FALSE)
dir.create("./07_PLSDAs",showWarnings = FALSE)
dir.create("./08_RF",showWarnings = FALSE)
dir.create("./09_LDA",showWarnings = FALSE)
dir.create("./10_Combined",showWarnings = FALSE)

########################################

# recombine variables in different dataframes
x_all<-cbind(x_disc,x_cont)

########################################

# first add NA to missing values
for(i in 1:length(grep("OTUs",colnames(x)))){
   x[x[,grep("OTUs",colnames(x))[i]]=="",grep("OTUs",colnames(x))[i]]<-NA
}
# read OTUs as factor
for(i in 1:length(grep("OTUs",colnames(x)))){
   x[,grep("OTUs",colnames(x))[i]]<-as.factor(x[,grep("OTUs",colnames(x))[i]])
}

########################################

# set hypotheses of OTU clustering
Hindex<-grep("OTUs",colnames(x))
Hlabel<-colnames(x)[grep("OTUs",colnames(x))]
OTUs<-list()
for (k in seq(Hlabel)) {
   OTUs[Hlabel[k]]<-list(levels(x[,Hindex[k]]))
}

########################################

# HISTOGRAMS TO CHECK FOR ERRORS

########################################
```

```
# Plot histograms for ALL variables NO groups
# OBS: Only to check for errors, weird distributions and outliers
dir.create("./02_histograms/01_all_variables-no_groups",showWarnings = FALSE)

# histogram for ALL variables
for(i in seq(x_all[1,])){
   pdf(paste("./02_histograms/01_all_variables-no_groups/histo_",colnames(x_all)[i],".pdf",sep=""),
family="ArialMT", width=9, height=12)
            a<-hist(x_all[,i],freq=F,xlab=NULL,,main=colnames(x_all)[i])
   dev.off()
}

########################################

# Plot histograms for ALL variables grouped by SEX
# OBS: Only to check for errors and weird distributions
dir.create("./02_histograms/02_all_variables-sex",showWarnings = FALSE)

#histogram for ALL variables
for(i in seq(x_all[1,])){
   pdf(paste("./02_histograms/02_all_variables-sex/histo_",colnames(x_all)[i],".pdf",sep=""), family="ArialMT",
width=9, height=12)
   layout(matrix(c(1,2),2,1))

a<-hist(x_all[,i][x$Sex=="F"],freq=F,xlab="F",,main=colnames(x_all)[i],xlim=c(min(na.omit(x_all[,i])),max(na.om
it(x_all[,i]))))

a<-hist(x_all[,i][x$Sex=="M"],freq=F,xlab="M",,main=colnames(x_all)[i],xlim=c(min(na.omit(x_all[,i])),max(na.o
mit(x_all[,i]))))
   dev.off()
}

########################################

# BOXPLOT + HISTOGRAM TO CHECK FOR ERRORS - first attempt

########################################

# Plot Boxplot + histogram for ALL variables - complete dataset
# Only to check for outliers

dir.create("./04_boxplots/01_all_variables-no-groups-check")

for(i in 1:ncol(x_all)){
   #get histogram data
   b1<-hist(x_all[,i],breaks=length(x_all[,i]),plot=FALSE)
   b2<-hist(x_all[,i],plot=FALSE)
   #open pdf file

pdf(paste("./04_boxplots/01_all_variables-no-groups-check/boxplot_",colnames(x_all)[i],".pdf",sep=""),family="Ar
ialMT", width=9, height=12)
            #remove zeros from counts
            q1<-b1$counts
            q1[q1==0]<-NA
            q2<-b2$counts
            q2[q2==0]<-NA
```

```
        #transform counts to boxplot coordinates
        m1<-(0.8-1.2)/(max(na.omit(q1))-min(na.omit(q1)))
        y1<-(m1*q1)+(1.2-(m1*min(na.omit(q1))))
        m2<-(0.8-1.2)/(max(na.omit(q2))-min(na.omit(q2)))
        y2<-(m2*q2)+(1.2-(m2*min(na.omit(q2))))
        #plot boxplot

boxplot(x_all[,i],main=colnames(x_all)[i],outline=F,ylim=c(min(na.omit(x_all[,i])),max(na.omit(x_all[,i]))),lwd=1.5
)
        #plot points from histogram
        points(y1,b1$mids,col="blue",cex=0.8)
        #plot bars from histogram
        rect(y2,b2$breaks[-length(b2$breaks)],rep(1.2,length(y2)),b2$breaks[-1],border="blue",col="blue",
lty=3,lwd=0.3,density = 10)
        #plot mean as a line
        segments(0.8,mean(na.omit(x_all[,i])),1.2,mean(na.omit(x_all[,i])),col="red")
        #plot N for variable
        text(1,max(na.omit(x_all[,i])), labels=paste("N = ",length(na.omit(x_all[,i])),sep=""),pos=3)
        #get range amplitude
        k<-max(na.omit(x_all[,i]))-min(na.omit(x_all[,i]))
        #plot axis for histogram
        segments(0.8,min(na.omit(x_all[,i]))-(0.01*k),1.2,min(na.omit(x_all[,i]))-(0.01*k),col="grey",adj=0)
        #plot max line for histogram
        segments(0.8,min(na.omit(x_all[,i])),0.8,max(na.omit(x_all[,i])),lty=3,col="grey")
        segments(0.8,min(na.omit(x_all[,i]))-(0.02*k),0.8,min(na.omit(x_all[,i])),col="grey")
        #plot min line for histogram
        segments(1.2,min(na.omit(x_all[,i])),1.2,max(na.omit(x_all[,i])),lty=3,col="grey")
        segments(1.2,min(na.omit(x_all[,i]))-(0.02*k),1.2,min(na.omit(x_all[,i])),col="grey")
        #plot labels for the histogram axis
        text(1,min(na.omit(x_all[,i]))-(0.025*k),labels="Count",cex=0.8)
        text(0.8,min(na.omit(x_all[,i]))-(0.03*k),labels=max(na.omit(q1)),cex=0.8)
        text(1.2,min(na.omit(x_all[,i]))-(0.03*k),labels=min(na.omit(q1)),cex=0.8)

   dev.off()
}

########################################

# CLEANING BAD VARIABLES

########################################

# adjust dataset by removing non-informative variables
# OBS: non random missing data (e.g. missing variable for all individuals from an OTU) can have pervasive effects
in multivariate analyses

# remove all variables that are completely missing for an OTU
to_remove<-list()
list_chars<-vector()
# loop for each taxon set
for(k in seq(Hlabel)){
   # loop for each variable
   for (i in 1:ncol(x_all)){
        for (j in seq(OTUs[[k]])) {
             if(length(na.omit(x_all[,i][x[,Hindex[k]]==OTUs[[k]][j]]))<1){
                  list_chars<-c(list_chars,colnames(x_all[i]))
```

```
                                to_remove[[k]]<-unique(list_chars)
                        }
                }
        }
}

# save table with a list of automatically removed variables
sink("./01_tables/Removed_Characters_auto.txt")
   print(to_remove)
sink()

# cleaning datasets
to_remove<-unique(unlist(to_remove))

# x_cln0<-x[,!colnames(x)%in%to_remove]
x_all_cln0<-x_all[,!colnames(x_all)%in%to_remove]

# Check the differences among taxon sets...
# If the missing variables are the same, follow the script. Otherwise, choose the variables to remove.
# Ex:
# to_remove2<-c("cat_LLd","cat_LLe",cat_con_fx_posoc")
to_remove2<-c("disc_Temp1","disc_Temp2","disc_Temp3","disc_SupL","disc_SupR","disc_PreV")

# save table with a list of manually removed variables
write.table(to_remove2,file=paste("./01_tables/Removed_Characters_manual.txt",sep=""),sep="\t",col.names=F)

# cleaning datasets
# x_cln0<-x_cln0[,!colnames(x_cln0)%in%to_remove2]
x_all_cln0<-x_all_cln0[,!colnames(x_all_cln0)%in%to_remove2]

########################################

# BOXPLOT + HISTOGRAM TO CHECK FOR ERRORS - second attempt

########################################

# Plot Boxplot + histograms for ALL variables grouped by OTUs
# Only to check for outliers

# loop for each taxon set
for(k in seq(Hlabel)){
   dir.create(paste("./04_boxplots/02_all_variables-",Hlabel[k],"-check",sep=""),showWarnings = FALSE)
   #loop for each variable
   for(i in 1:ncol(x_all_cln0)){
           #open pdf file

pdf(paste("./04_boxplots/02_all_variables-",Hlabel[k],"-check/boxplot_",colnames(x_all_cln0)[i],".pdf",sep=""),fam
ily="ArialMT", width=length(OTUs[[k]]), height=12)
                        #plot boxplot
                        par(mar = c(7, 4, 4, 2) + 0.1)
                        boxplot(x_all_cln0[,i] ~
x[,Hindex[k]],main=colnames(x_all_cln0)[i],outline=F,ylim=c(min(na.omit(x_all_cln0[,i])),max(na.omit(x_all_cln0
[,i]))),boxwex = 0.8,xaxt = "n", xlab= Hlabel[k], ylab ="")
   #       axis(1, labels = FALSE)
                        xlabels <- levels(x[,Hindex[k]])
```

```
                text(1:length(levels(x[,Hindex[k]])), par("usr")[3], offset=1, srt = 45, adj = 1, labels = xlabels, xpd
= TRUE)
                #loop for each OTUs
                for(j in 1:length(levels(x[,Hindex[k]]))){
                        if(length(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))==0){
                                #plot N for variable
                                text(j,mean(na.omit(x_all_cln0[,i])), labels=paste("N =
",length(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                        }
                        else{
                                #transform counts to boxplot coordinates
                                q<-as.numeric(table(x_all_cln0[x[,Hindex[k]]==xlabels[j],i]))
                                m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                                if(m!=-Inf){
                                        y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                                }else{
                                        if(length(q)>1){
                                                y=q
                                                y[!is.na(y)]<-j+0.4
                                        }else{
                                                y=j+0.4
                                        }
                                }
                                #plot points from histogram
                                if(length(q)<2){

points(y,unique(na.omit(x_all_cln0[x[,Hindex[k]]==xlabels[j],i])),col="blue",cex=0.8)
                                }else{
                                        # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln0[x[,Hindex[k]]==xlabels[j],i]),"dimnames")[[1]]),col="blue",cex=0.8)
                                }
                                #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),j+0.4,mean(na.omit(x_all_cl
n0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),col="red")
                                #plot N for variable
                                text(j,max(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),
labels=paste("N =
",length(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                #get range amplitude
                                rango<-max(na.omit(x_all_cln0[,i]))-min(na.omit(x_all_cln0[,i]))
                                #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))-(0.01*rango),j+0.4,min(na.om
it(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))-(0.01*rango),col="grey",adj=0)
                                #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),j-0.4,max(na.omit(x_all_cln0[
,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))-(0.02*rango),j-0.4,min(na.omi
t(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),col="grey")
                                #plot mean line for histogram
```

```
segments(j,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),j,max(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))-(0.02*rango),j,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),col="grey")
                                        #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),j+0.4,max(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))-(0.02*rango),j+0.4,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]])),col="grey")
                                        #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))-(0.025*rango),labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))-(0.03*rango),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln0[,i][x[,Hindex[k]]==levels(x[,Hindex[k]])[j]]))-(0.03*rango),labels=min(na.omit(q)),cex=0.8)
                                }
                        }
                dev.off()
        }
}

#######################################

# CLEANING BAD TERMINALS

#######################################

# adjust dataset by manually removing outliers, etc...
# OBS: consider removing outliers completelly if there are doubts about collected data or too many missing variables for them

# include in the vector "to_remove01" all IDs for outliers that you want to completelly remove from your analyses.
# EX:
# to_remove3<-c("1381","2","1337","820","1272","410")

to_remove3<-c("157")

x_cln<-x[!x$ID%in%to_remove3,]
x_all_cln1<-x_all_cln0[!x$ID%in%to_remove3,]

# write list of removed terminals
removed_terminals<-cbind(x$ID[which(x$ID%in%to_remove3)],x$Voucher[which(x$ID%in%to_remove3)])

# save table with a list of removed terminals
write.table(removed_terminals,file=paste("./01_tables/Removed_Terminals.txt",sep=""),sep="\t",col.names=F)

#######################################

# CLEANING BAD VARIABLES - second turn
```

```
########################################

# adjust dataset by removing non-informative variables or problematic (doubts) data
# OBS: nonrandom missing data (e.g. missing variable for too many individuals from an OTU) can have pervasive
effects in multivariate analyses

# create table to check unequal sample size among variables within sets of OTUs
summaryTable<-list()

for(k in seq(Hlabel)){
   for(i in seq(ncol(x_all_cln1))){
            summaryTable[[Hlabel[k]]][[colnames(x_all_cln1)[i]]]<-table(x_cln[,Hindex[k]][!is.na(x_all_cln1[,i])])
   }
}

# save list of automatically removed variables
sink("./01_tables/Summary_Table-N_per_variable_per_OTU.txt")
   print(summaryTable)
sink()

# Check the differences among taxon sets...
# Compare Summary table and boxplots for each variable to decide if the variable is "informative"
# Manually choose the variables to remove.
# Ex:
# to_remove4<-c("cat_LLd","cat_LLe",cat_con_fx_posoc")

# to_remove4<-c()
to_remove4<-c()

# other uninformative (your opinion)
to_remove5<-c()
# # OBS: variables without an evident meaningful pattern

# save table with the list of removed variables
write.table(c(to_remove4,to_remove5),file=paste("./01_tables/Removed_Characters_manual-secondTurn.txt",sep="
"),sep="\t",col.names=F)

# cleaning datasets
# x_cln<-x_cln1[,!colnames(x_cln1)%in%c(to_remove4,to_remove5)]
x_all_cln<-x_all_cln1[,!colnames(x_all_cln1)%in%c(to_remove4,to_remove5)]

########################################
# CHECK CORRELATION BETWEEN SIZE AND other VARIABLES (ALLOMETRY vs ISOMETRY)

########################################
# set list of variables to test
allo<-list()

x_cont_cln<-list()

for(k in 1:length(Hlabel)){
   x_cont_cln[[Hlabel[k]]]<-x_all_cln[,grep("cont",colnames(x_all_cln))]
}
```

```
for(k in 1:length(Hlabel)){
   x_cont_cln[[Hlabel[k]]]$cont_TrL<-x_cont_cln[[Hlabel[k]]]$cont_SVL-x_cont_cln[[Hlabel[k]]]$cont_HL
}

# evaluate allometry/isometry by
for(k in 1:length(Hlabel)){
   dir.create(paste("./00_scatterplot/01_all_variables-no-groups/",Hlabel[k],sep=""),showWarnings = FALSE)
   for(i in 1:length(x_cont_cln[[Hlabel[k]]][1,])){
            pdf(paste0("00_scatterplot/",colnames(x_cont_cln[[Hlabel[k]]])[i],".pdf"))

plot(x_cont_cln[[Hlabel[k]]]$cont_TrL,x_cont_cln[[Hlabel[k]]][,i],main=colnames(x_cont_cln[[Hlabel[k]]])[i],xlab
="Trunk length",ylab=colnames(x_cont_cln[[Hlabel[k]]])[i])
                        abline(lm(x_cont_cln[[Hlabel[k]]][,i] ~ x_cont_cln[[Hlabel[k]]]$cont_TrL))

allo[[Hlabel[k]]][colnames(x_cont_cln[[Hlabel[k]]])[i]]<-list(summary(lm(x_cont_cln[[Hlabel[k]]][,i] ~
x_cont_cln[[Hlabel[k]]]$cont_TrL)))
            dev.off()
   }
}

########################################

no_isometric<-c("cont_WNCS","cont_BNCS","cont_InS")

x_cont_cln_iso<-vector()

x_cont_cln_iso<-x_cont_cln[[Hlabel[1]]][,!colnames(x_cont_cln[[Hlabel[1]]])%in%c(no_isometric,"cont_SVL")]

x_cont_corrected<-x_cont_cln_iso/x_cont_cln_iso$cont_TrL

#
x_cont_corrected<-data.frame(rat_HL=x_cont_cln_iso$cont_HL/x_cont_cln_iso$cont_TrL,rat_TL=x_cont_cln_iso
$cont_TL/x_cont_cln_iso$cont_TrL)

########################################

# Plot Boxplot + histograms for ALL variables grouped by OTUs
# Only to check for outliers

x_all_cln_noOut_noZ<-list()

# cleaning dataset from outliers
x_all_cln_noOut_noZ<-list()
for(k in seq(Hlabel)){
   x_all_cln_noOut_noZ[[Hlabel[k]]]<-x_cont_corrected
}


########################################

# Boxplot + histogram for ALL variables - per OTUs
# Final
for(k in 1:length(Hlabel)){
   dir.create(paste("./04_boxplots/03_all_variables-RATIOS-",Hlabel[k],sep=""),showWarnings = FALSE)
   #loop for each variable
   for(i in 1:ncol(x_all_cln_noOut_noZ[[k]])){
```

```
		#open pdf file

pdf(paste("./04_boxplots/03_all_variables-RATIOS-",Hlabel[k],"/boxplot_",colnames(x_all_cln_noOut_noZ[[k]])[i]
,".pdf",sep=""),family="ArialMT", width=9, height=12)
			#plot boxplot
			par(mar = c(7, 4, 4, 2) + 0.1)
			boxplot(x_all_cln_noOut_noZ[[k]][,i] ~
x_cln[,Hindex[k]],main=colnames(x_all_cln_noOut_noZ[[k]])[i],outline=F,ylim=c(min(na.omit(x_all_cln_noOut_n
oZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),boxwex = 0.8,xaxt = "n", xlab= Hlabel[k], ylab ="")
   #		axis(1, labels = FALSE)
			xlabels <- levels(x_cln[,Hindex[k]])
			text(1:length(levels(x_cln[,Hindex[k]])), par("usr")[3], offset=1, srt = 45, adj = 1, labels = xlabels,
xpd = TRUE)
			#loop for each OTUs
			for(j in 1:length(levels(x_cln[,Hindex[k]]))){

if(length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]==levels(x_cln[,Hindex[k]])[j]]))==0){
				#plot N for variable
				text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]==levels(x_cln[,Hindex[k]])[j]])),sep=""),cex=0.8,
pos=3)
			}
			else{
				#transform counts to boxplot coordinates

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]==xlabels[j],i]))
				m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
				if(m!=-Inf){
					y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
				}else{
					if(length(q)>1){
						y=q
						y[!is.na(y)]<-j+0.4
					}else{
						y=j+0.4
					}
				}
				#plot points from histogram
				if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]==xlabels[j],i])),col="blue",cex=0.8)
				}else{
					# points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]==xlabels[j],i]),"dimnames")[[1]]),col="
blue",cex=0.8)
				}
				#plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]==levels(x_cln[,Hindex[k]])[j]])),j+0.
4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]==levels(x_cln[,Hindex[k]])[j]])),col="red")
				#plot N for variable

text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]==levels(x_cln[,Hindex[k]])[j]])),
labels=paste("N =
```

```
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),sep=""),cex=0.8,
pos=3)
                                        #get range amplitude

rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))-min(na.omit(x_all_cln_noOut_noZ[[k]][,i]))
                                        #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.01
*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.01*
rango),col="grey",adj=0)
                                        #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j-0.4,
max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02
*rango),j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="gr
ey")
                                        #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j,max(na.
omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02*ran
go),j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                        #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j+0.4
,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02
*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="g
rey")
                                        #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.025*rango),
labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.03*rang
o),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.03*ran
go),labels=min(na.omit(q)),cex=0.8)
                                }
                        }
                dev.off()
        }
}

########################################

x_mult<-cbind(x_cont_corrected[,-ncol(x_cont_corrected)],x_all_cln[,grep("disc",colnames(x_all_cln))])

colnames(x_mult)<-gsub("cont_","rat_",colnames(x_mult))

########################################
```

```
# REMOVING OUTLIERS

########################################

x_all_cln<-x_mult

# removing outliers based on data values - removing a specific value only when it is out of 2*(interquartile range)
for that particular variable
out_check<-list()

# loop for each taxon set
for(k in seq(Hlabel)){
   #loop for each variable
   for(i in 1:ncol(x_all_cln)){
            outlier_values <- boxplot(x_all_cln[,i] ~ x_cln[,Hindex[k]],range=2,plot=F)$stats
            outliers_ind<-vector()
            for (j in seq(OTUs[[k]])) {

outliers_ind<-c(outliers_ind,x_cln$ID[x_cln[,Hindex[k]]==OTUs[[k]][j]][x_all_cln[x_cln[,Hindex[k]]==OTUs[[k]]
[j],i]<outlier_values[1,j]|x_all_cln[x_cln[,Hindex[k]]==OTUs[[k]][j],i]>outlier_values[5,j]])
            }
            out_check[[Hlabel[k]]][[colnames(x_all_cln)[i]]]<-unique(na.omit(outliers_ind))
            # names(out_check[k])[i]<-colnames(x_all_cln)[i]
   }
}

# write list of removed outliers
sink("./01_tables/Removed_Outliers.txt")
   print(out_check)
sink()

# cleaning dataset from outliers
x_all_cln_noOut<-list()
for(k in seq(Hlabel)){
   x_all_cln_noOut[[Hlabel[k]]]<-x_all_cln
   for(i in seq(out_check[[k]])){

x_all_cln_noOut[[k]][which(x_cln$ID%in%out_check[[k]][[i]]),grep(names(out_check[[k]][i]),colnames(x_all_cln_
noOut[[k]]))]<-NA
   }
}

########################################

# REMOVING ZERO-VARIANCE

########################################

# removing zero-variance variables after correcting for outliers
zero_var<-list()
x_all_cln_noOut_noZ<-list()
for(k in seq(Hlabel)){
   zero_var[[Hlabel[k]]]<-which(apply(x_all_cln_noOut[[k]], 2, function(x) var(x, na.rm = TRUE)) == 0)
```

```
x_all_cln_noOut_noZ[[Hlabel[k]]]<-x_all_cln_noOut[[k]][,!seq(x_all_cln_noOut[[k]][1,])%in%zero_var[[Hlabel[k]]]]
}

# write list of zero-variance variables
sink("./01_tables/Removed_Zero-Variance.txt")
   print(zero_var)
sink()

########################################

# PLOT QQPLOTS

########################################

# Plot Quantiles-Quantiles normal plot for ALL variables
# Only to check distribution.
# OBS: Discrete and categorical variables are not Normal distributed by definition (they can follow Poisson, binomial, etc). Only use this to check for remaining outliers and possible problems

dir.create("./03_QQplots/01_all_variables-no-groups",showWarnings = FALSE)

#qqnorm for ALL variables
for(k in seq(Hlabel)){
   dir.create(paste("./03_QQplots/01_all_variables-no-groups/",Hlabel[k],sep=""),showWarnings = FALSE)
   for (i in 1:length(x_all_cln_noOut_noZ[[k]][1,])){

pdf(paste("./03_QQplots/01_all_variables-no-groups/",Hlabel[k],"/qqnorm_",colnames(x_all_cln_noOut_noZ[[k]])[i],".pdf",sep=""), family="ArialMT", width=9, height=12)
            qqnorm(x_all_cln_noOut_noZ[[k]][,i], main = colnames(x_all_cln_noOut_noZ[[k]][i]))
            qqline(x_all_cln_noOut_noZ[[k]][,i], col=2)
            if (length(x_all_cln_noOut_noZ[[k]][,i])<=50){

text(-1,max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))-(diff(range(na.omit(x_all_cln_noOut_noZ[[k]][,i])))*0.1),paste("Shapiro-Wilk - p.value = ",round(shapiro.test(x_all_cln_noOut_noZ[[k]][,i])$p.value,4)))
            } else {

text(-0.9,max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))-(diff(range(na.omit(x_all_cln_noOut_noZ[[k]][,i])))*0.1),paste("Lilliefors - p.value = ",round(lillie.test(na.omit(x_all_cln_noOut_noZ[[k]][,i]))$p.value,4)))
            }
   dev.off()
   }
}

########################################

# TEST FOR SEXUAL DIMORPHISM

########################################

# Statistical test for sexual dimorphism (SD) for ALL variables
# The t-test will be used if the variable is continuous or a ratio, normally distributed with equal variances, otherwise the Wilcoxon test will be applied.
```

```
# OBS: This test will not evaluate SD within OTUs to avoid small sample sizes, but if your sampling is comprehensive enough or your OTUs are phylogenetically distant consider changing the script, or running sex as covariate.

#new matrix
dimorphism<-list()

for(k in seq(Hlabel)){
   dimorphism[[Hlabel[k]]]<-matrix(nrow=ncol(x_all_cln_noOut_noZ[[k]]),ncol=2)
   #row names
   rownames(dimorphism[[k]])<-colnames(x_all_cln_noOut_noZ[[k]])
   colnames(dimorphism[[k]])<-c("Test","p.value")
   #loop to fill table
   for (i in 1:ncol(x_all_cln_noOut_noZ[[k]])){

if(length(grep("cont",colnames(x_all_cln_noOut_noZ[[k]])[i]))==1|length(grep("rat_",colnames(x_all_cln_noOut_noZ[[k]])[i]))==1){
                        if
(fligner.test(x_all_cln_noOut_noZ[[k]][,i]~x_cln$Sex,data=x_all_cln_noOut_noZ[[k]])$p.value>=0.05) {
                                dimorphism[[k]][i,1]<-"T.test"

dimorphism[[k]][i,2]<-round(t.test(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i],x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i])$p.value,4)
                        } else {
                                dimorphism[[k]][i,1]<-"Wilcoxon.test"

dimorphism[[k]][i,2]<-round(wilcox.test(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i],x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i])$p.value,4)
                        }
                } else {
                        dimorphism[[k]][i,1]<-"Wilcoxon.test"

dimorphism[[k]][i,2]<-round(wilcox.test(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i],x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i])$p.value,4)
                }
   }
   rownames(dimorphism[[k]])<-gsub("disc_","",rownames(dimorphism[[k]]))
   rownames(dimorphism[[k]])<-gsub("cont_","",rownames(dimorphism[[k]]))
   rownames(dimorphism[[k]])<-gsub("rat_","",rownames(dimorphism[[k]]))
   rownames(dimorphism[[k]])<-gsub("cat_","",rownames(dimorphism[[k]]))
   write.table(dimorphism[[k]],file=paste("./01_tables/Dimorphism-no_groups-",Hlabel[k],".txt",sep=""),sep="\t")
}

# dimorphism<-read.table("./01_tables/dimorphism-no_groups.txt")

########################################

# BOXPLOT + HISTOGRAM - FINAL

########################################

# Boxplot + histogram for ALL variables - per OTUs
# Final
for(k in 1:length(Hlabel)){
   dir.create(paste("./04_boxplots/03_all_variables-",Hlabel[k],sep=""),showWarnings = FALSE)
   #loop for each variable
```

```
   for(i in 1:ncol(x_all_cln_noOut_noZ[[k]])){
          #open pdf file

pdf(paste("./04_boxplots/03_all_variables-",Hlabel[k],"/boxplot_",colnames(x_all_cln_noOut_noZ[[k]])[i],".pdf",sep=""),family="ArialMT", width=9, height=12)
                    #plot boxplot
                    par(mar = c(7, 4, 4, 2) + 0.1)
                    boxplot(x_all_cln_noOut_noZ[[k]][,i] ~
x_cln[,Hindex[k]],main=colnames(x_all_cln_noOut_noZ[[k]])[i],outline=F,ylim=c(min(na.omit(x_all_cln_noOut_noZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),boxwex = 0.8,xaxt = "n", xlab= Hlabel[k], ylab ="")
   #        axis(1, labels = FALSE)
                    xlabels <- levels(x_cln[,Hindex[k]])
                    text(1:length(levels(x_cln[,Hindex[k]])), par("usr")[3], offset=1, srt = 45, adj = 1, labels = xlabels, xpd = TRUE)
                    #loop for each OTUs
                    for(j in 1:length(levels(x_cln[,Hindex[k]]))){

if(length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))==0){
                              #plot N for variable
                              text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i])), labels=paste("N = ",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),sep=""),cex=0.8, pos=3)
                    }
                    else{
                              #transform counts to boxplot coordinates

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]]==xlabels[j],i]))
                              m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                              if(m!=-Inf){
                                        y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                              }else{
                                        if(length(q)>1){
                                                  y=q
                                                  y[!is.na(y)]<-j+0.4
                                        }else{
                                                  y=j+0.4
                                        }
                              }
                              #plot points from histogram
                              if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]]==xlabels[j],i])),col="blue",cex=0.8)
                              }else{
                                        # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]]==xlabels[j],i]),"dimnames")[[1]]),col="blue",cex=0.8)
                              }
                              #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j+0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="red")
                              #plot N for variable

text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),
labels=paste("N =
```

```
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),sep=""),cex=0.8,
pos=3)
                                        #get range amplitude

rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))-min(na.omit(x_all_cln_noOut_noZ[[k]][,i]))
                                        #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.01
*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.01*
rango),col="grey",adj=0)
                                        #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j-0.4,
max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02
*rango),j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="gr
ey")
                                        #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j,max(na.
omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02*ran
go),j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                        #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j+0.4
,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02
*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="g
rey")
                                        #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.025*rango),
labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.03*rang
o),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.03*ran
go),labels=min(na.omit(q)),cex=0.8)
                                }
                        }
                dev.off()
        }
}

########################################

# Boxplot + histogram for ALL variables - per Sex (if dimorphic)- per OTUs
# Need to run/load test for sexual dimorphism first
for(k in 1:length(Hlabel)){
   dir.create(paste("./04_boxplots/04_all_variables-",Hlabel[k],"-sex",sep=""),showWarnings = FALSE)
   #loop for each variable
```

```
   for(i in 1:ncol(x_all_cln_noOut_noZ[[k]])){
          if(as.numeric(dimorphism[[k]][i,2])>=0.05){
                #open pdf file

pdf(paste("./04_boxplots/04_all_variables-",Hlabel[k],"-sex/boxplot_",colnames(x_all_cln_noOut_noZ[[k]])[i],".pdf
",sep=""),family="ArialMT", width=3+length(OTUs[[k]]), height=14)
                #plot boxplot
                boxplot(x_all_cln_noOut_noZ[[k]][,i] ~
x_cln[,Hindex[k]],main=colnames(x_all_cln_noOut_noZ[[k]])[i],outline=F,ylim=c(min(na.omit(x_all_cln_noOut_n
oZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),boxwex = 0.8,xaxt = "n", xlab= Hlabel[k], ylab ="")
                xlabels <- levels(x_cln[,Hindex[k]])
                text(1:length(levels(x_cln[,Hindex[k]])), par("usr")[3], offset=1, srt = 45, adj = 1, labels
= xlabels, xpd = TRUE)
                #loop for each OTUs
                for(j in 1:length(levels(x_cln[,Hindex[k]]))){

if(length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))==0){
                        #plot N for variable
                        text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),sep=""),cex=0.8,
pos=3)
                }
                else{

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]]==xlabels[j],i]))
                        #transform counts to boxplot coordinates
                        m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                        if(m!=-Inf){
                                y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                        }else{
                                if(length(q)>1){
                                        y=q
                                        y[!is.na(y)]<-j+0.4
                                }else{
                                        y=j+0.4
                                }
                        }
                        #plot points from histogram
                        if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]]==xlabels[j],i])),col="blue",cex=0.8)
                        }else{
                                # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln[,Hindex[k]]==xlabels[j],i]),"dimnames")[[1]]),col="
blue",cex=0.8)
                        }
                        #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j+0.
4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="red")
                        #plot N for variable

text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),
labels=paste("N =
```

```
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),sep=""),cex=0.8,
pos=3)
                                                #get range amplitude

rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))-min(na.omit(x_all_cln_noOut_noZ[[k]][,i]))
                                                #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.01
*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.01*
rango),col="grey",adj=0)
                                                #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j-0.4,
max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02
*rango),j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="gr
ey")
                                                #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j,max(na.
omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02*ran
go),j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                                #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),j+0.4
,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.02
*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]])),col="g
rey")
                                                #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.025*rango),
labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.03*rang
o),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln[,Hindex[k]]==levels(x_cln[,Hindex[k]])[j]]))-(0.03*ran
go),labels=min(na.omit(q)),cex=0.8)
                                                }
                                        }
                        dev.off()
                } else {
                        #open pdf file

pdf(paste("./04_boxplots/04_all_variables-",Hlabel[k],"-sex/boxplot_",colnames(x_all_cln_noOut_noZ[[k]])[i],".pdf
",sep=""),family="ArialMT", width=10+length(OTUs[[k]]), height=14)
                                        layout(matrix(c(1,2),1,2))
                                        #plot boxplot for females
                                        boxplot(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i] ~
x_cln[,Hindex[k]][x_cln$Sex=="F"],main=colnames(x_all_cln_noOut_noZ[[k]])[i],outline=F,ylim=c(min(na.omit(x
```

```
_all_cln_noOut_noZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),boxwex = 0.8,xaxt = "n", xlab=
paste(Hlabel[k]," Females",sep=""), ylab ="")
                                xlabels <- levels(x_cln[,Hindex[k]])
                                text(1:length(levels(x_cln[,Hindex[k]])), par("usr")[3], offset=1, srt = 45, adj = 1, labels
= xlabels, xpd = TRUE)
                                #loop for each OTUs
                                for(j in 1:length(levels(x_cln[,Hindex[k]]))){

if(length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==levels(x_cln
[,Hindex[k]])[j]]))==0){
                                                #plot N for variable
                                                text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i])),
labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==levels(x_cln[
,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                        }else{

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==levels(x_
cln[,Hindex[k]])[j]]))
                                                # q[q==0]<-NA
                                                #transform counts to boxplot coordinates
                                                m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                                                if(m!=-Inf){
                                                        y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                                                }else{
                                                        if(length(q)>1){
                                                                y=q
                                                                y[!is.na(y)]<-j+0.4
                                                        }else{
                                                                y=j+0.4
                                                        }
                                                }
                                                #plot points from histogram
                                                if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==level
s(x_cln[,Hindex[k]])[j]])),col="blue",cex=0.8)
                                                }else{
                                                        # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==
levels(x_cln[,Hindex[k]])[j]]),"dimnames")[[1]]),col="blue",cex=0.8)
                                                }
                                                #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==
levels(x_cln[,Hindex[k]])[j]])),j+0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]]
[x_cln$Sex=="F"]==levels(x_cln[,Hindex[k]])[j]])),col="red")
                                                #plot N for variable

text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==levels(x_c
ln[,Hindex[k]])[j]])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==levels(x_cln[
,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                                #get range amplitude
```

```
rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i]))-min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln
$Sex=="F",i]))
                                        #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==le
vels(x_cln[,Hindex[k]])[j]]))-(0.01*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,
Hindex[k]][x_cln$Sex=="F"]==levels(x_cln[,Hindex[k]])[j]]))-(0.01*rango),col="grey",adj=0)
                                        #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==le
vels(x_cln[,Hindex[k]])[j]])),j-0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_
cln$Sex=="F"]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==le
vels(x_cln[,Hindex[k]])[j]]))-(0.02*rango),j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,
Hindex[k]][x_cln$Sex=="F"]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                        #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==level
s(x_cln[,Hindex[k]])[j]])),j,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Se
x=="F"]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==level
s(x_cln[,Hindex[k]])[j]]))-(0.02*rango),j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[
k]][x_cln$Sex=="F"]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                        #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==l
evels(x_cln[,Hindex[k]])[j]])),j+0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][
x_cln$Sex=="F"]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==l
evels(x_cln[,Hindex[k]])[j]]))-(0.02*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[
,Hindex[k]][x_cln$Sex=="F"]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                        #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==levels(x_cl
n[,Hindex[k]])[j]]))-(0.025*rango),labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==levels(
x_cln[,Hindex[k]])[j]]))-(0.03*rango),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="F",i][x_cln[,Hindex[k]][x_cln$Sex=="F"]==levels(
x_cln[,Hindex[k]])[j]]))-(0.03*rango),labels=min(na.omit(q)),cex=0.8)
                                        }
                                }
                                #plot boxplot for males
                                boxplot(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i] ~
x_cln[,Hindex[k]][x_cln$Sex=="M"],main=colnames(x_all_cln_noOut_noZ[[k]])[i],outline=F,ylim=c(min(na.omit(
x_all_cln_noOut_noZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),boxwex = 0.8,xaxt = "n", xlab=
paste(Hlabel[k]," Males",sep=""), ylab ="")
                                xlabels <- levels(x_cln[,Hindex[k]])
                                text(1:length(levels(x_cln[,Hindex[k]])), par("usr")[3], offset=1, srt = 45, adj = 1, labels
= xlabels, xpd = TRUE)
                                #loop for each OTUs
```

```
                        for(j in 1:length(levels(x_cln[,Hindex[k]]))){

if(length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==levels(x_c
ln[,Hindex[k]])[j]]))==0){
                                        #plot N for variable
                                        text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i])),
labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==levels(x_cl
n[,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                    }else{

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==levels(
x_cln[,Hindex[k]])[j]]))
                                        #transform counts to boxplot coordinates
                                        m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                                        if(m!=-Inf){
                                                y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                                        }else{
                                                if(length(q)>1){
                                                        y=q
                                                        y[!is.na(y)]<-j+0.4
                                                }else{
                                                        y=j+0.4
                                                }
                                        }
                                        #plot points from histogram
                                        if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==lev
els(x_cln[,Hindex[k]])[j]])),col="blue",cex=0.8)
                                        }else{
                                                # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]
==levels(x_cln[,Hindex[k]])[j]]),"dimnames")[[1]]),col="blue",cex=0.8)
                                        }
                                        #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]=
=levels(x_cln[,Hindex[k]])[j]])),j+0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[
k]][x_cln$Sex=="M"]==levels(x_cln[,Hindex[k]])[j]])),col="red")
                                        #plot N for variable

text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==levels(x_
cln[,Hindex[k]])[j]])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==levels(x_cl
n[,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                        #get range amplitude

rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i]))-min(na.omit(x_all_cln_noOut_noZ[[k]][x_cl
n$Sex=="M",i]))
                                        #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==
levels(x_cln[,Hindex[k]])[j]]))-(0.01*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cl
n[,Hindex[k]][x_cln$Sex=="M"]==levels(x_cln[,Hindex[k]])[j]]))-(0.01*rango),col="grey",adj=0)
```

```
                                                    #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==
levels(x_cln[,Hindex[k]])[j]])),j-0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][
x_cln$Sex=="M"]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==
levels(x_cln[,Hindex[k]])[j]]))-(0.02*rango),j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln
[,Hindex[k]][x_cln$Sex=="M"]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                                    #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==lev
els(x_cln[,Hindex[k]])[j]])),j,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$
Sex=="M"]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==lev
els(x_cln[,Hindex[k]])[j]]))-(0.02*rango),j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hind
ex[k]][x_cln$Sex=="M"]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                                    #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]=
=levels(x_cln[,Hindex[k]])[j]])),j+0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k
]][x_cln$Sex=="M"]==levels(x_cln[,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]=
=levels(x_cln[,Hindex[k]])[j]]))-(0.02*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_c
ln[,Hindex[k]][x_cln$Sex=="M"]==levels(x_cln[,Hindex[k]])[j]])),col="grey")
                                                    #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==levels(x_
cln[,Hindex[k]])[j]]))-(0.025*rango),labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==level
s(x_cln[,Hindex[k]])[j]]))-(0.03*rango),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln$Sex=="M",i][x_cln[,Hindex[k]][x_cln$Sex=="M"]==level
s(x_cln[,Hindex[k]])[j]]))-(0.03*rango),labels=min(na.omit(q)),cex=0.8)
                                                }
                                        }
                                dev.off()
                        }
                }
        }

########################################

# REMOVING ZERO-VARIANCE - again

########################################

# removing zero-variance variables after correcting for outliers
zero_var<-list()
x_all_cln_noOut_noZ_again<-list()
for(k in seq(Hlabel)){
   zero_var[[Hlabel[k]]]<-which(apply(x_all_cln_noOut_noZ[[k]], 2, function(x) var(x, na.rm = TRUE)) == 0)
```

```
x_all_cln_noOut_noZ_again[[Hlabel[k]]]<-x_all_cln_noOut_noZ[[k]][,!seq(x_all_cln_noOut_noZ[[k]][1,])%in%ze
ro_var[[Hlabel[k]]]]
}

# write list of zero-variance variables
sink("./01_tables/Removed_Zero-Variance-AGAIN.txt")
   print(zero_var)
sink()

########################################

# CLEANING DATA FOR each set of OTUs

# dataset

x_cln_red<-list()

x_multi_cln_noOut_noZ_clnOOM<-list()

for (k in seq(Hlabel)) {
   x_cln_red[[Hlabel[k]]]<-x_cln[!is.na(x_cln[,Hindex[k]]),]
   x_multi_cln_noOut_noZ_clnOOM[[Hlabel[k]]]<-x_all_cln_noOut_noZ[[Hlabel[k]]][!is.na(x_cln[,Hindex[k]]),]
}


########################################

# CLEANING DATA FOR MULTIVARIATE ANALYSES

# dataset cannot have missing entries for most of the multivariate analyses.

########################################
# OVERALL OTUs MEDIAN (OOM)- NO OUTLIERS
########################################

# prepare dataset for multivariate for MULTI (RATES, transformed CAT and DISC) variables - NO OUTLIERS
# fill missing entries with OVERALL OTUs MEDIAN (OOM)
# x_multi_cln_noOut_noZ_clnOOM<-x_all_cln_noOut_noZ

for (k in seq(Hlabel)) {
   for (i in 1:length(x_multi_cln_noOut_noZ_clnOOM[[k]][1,])) {
           for (j in seq(OTUs[[k]])){
                   m <-
median(na.omit(x_multi_cln_noOut_noZ_clnOOM[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==OTUs[[k]][j]]))
                   if(is.na(m)){
                           w <- median(na.omit(x_multi_cln_noOut_noZ_clnOOM[[k]][,i]))

x_multi_cln_noOut_noZ_clnOOM[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==OTUs[[k]][j]][is.na(x_multi_cln_n
oOut_noZ_clnOOM[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==OTUs[[k]][j]])] <- w
                   }

x_multi_cln_noOut_noZ_clnOOM[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==OTUs[[k]][j]][is.na(x_multi_cln_n
oOut_noZ_clnOOM[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==OTUs[[k]][j]])] <- m
           }
   }
```

```
}

# x_multi_cln_noOut_noZ_cln<-x_multi_cln_noOut_noZ_clnOOM

########################################

# SET COLOR AND SHAPES FOR PLOTS - MULTIVARIATE ANALYSES

########################################

# create fixed color pallete for multivariated plots

gg_color_hue <- function(n) {
 hues = seq(15, 375, length = n + 1)
 hcl(h = hues, l = 65, c = 100)[1:n]
}

all_OTUs<-vector()

for(k in seq(Hindex)){
   all_OTUs<-c(all_OTUs,levels(x[,Hindex[k]]))
}

all_OTUs<-unique(all_OTUs)

myColors <- gg_color_hue(length(all_OTUs))

names(myColors) <- all_OTUs

colScale <- scale_colour_manual(name = "OTUS",values = myColors)
fillScale <- scale_fill_manual(name = "OTUS",values = myColors)


########################################

# PCA

########################################

x_multi_cln_noOut_noZ_noAllo_cln_noCor<-x_multi_cln_noOut_noZ_clnOOM

# correct zero var for sex
zero_var_F<-list()
x_multi_cln_noOut_noZ_noAllo_cln_noCor_F<-list()
for(k in seq(Hlabel)){

zero_var_F[[Hlabel[k]]]<-which(apply(x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]][x_cln_red[[Hlabel[k]]]$Se
x=="F",], 2, function(x) var(x, na.rm = TRUE))==0)

x_multi_cln_noOut_noZ_noAllo_cln_noCor_F[[Hlabel[k]]]<-x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]][x_c
ln_red[[Hlabel[k]]]$Sex=="F",][,!seq(x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]][x_cln_red[[Hlabel[k]]]$Sex
=="F",][1,])%in%zero_var_F[[Hlabel[k]]]]
}


zero_var_M<-list()
```

```
x_multi_cln_noOut_noZ_noAllo_cln_noCor_M<-list()
for(k in seq(Hlabel)){

zero_var_M[[Hlabel[k]]]<-which(apply(x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",], 2, function(x) var(x, na.rm = TRUE))==0)

x_multi_cln_noOut_noZ_noAllo_cln_noCor_M[[Hlabel[k]]]<-x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",][,!seq(x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",][1,])%in%zero_var_M[[Hlabel[k]]]]
}

zero_var<-list()
for(k in seq(Hlabel)){
   zero_var[[Hlabel[k]]]<-which(apply(x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]], 2, function(x) var(x, na.rm = TRUE))==0)

x_multi_cln_noOut_noZ_noAllo_cln_noCor[[Hlabel[k]]]<-x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]][,!seq(x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]][1,])%in%zero_var[[Hlabel[k]]]]
}


# plot PCA for multi variables - polygon instead of ellipse - PC1 & PC2

g1<-list()
g2<-list()
g3<-list()

for(k in seq(Hlabel)){
   dir.create(paste("./06_PCAs/01_PCA_multi_variables-",Hlabel[k],sep=""),showWarnings = FALSE)

   # plot pca - ALL variables
   pca_multi <- prcomp(x_multi_cln_noOut_noZ_noAllo_cln_noCor[[k]], scale.=T)
   dfscores <- as.data.frame(pca_multi$x)
   prop_pca = pca_multi$sdev^2/sum(pca_multi$sdev^2)
   forChull<-cbind(dfscores[,1:2],OTUS=x_cln_red[[Hlabel[k]]][,Hindex[k]],Labels=x_cln_red[[Hlabel[k]]]$ID)
   forChull$OTUS<-factor(forChull$OTUS,all_OTUs[all_OTUs%in%forChull$OTUS])
   afterChull<-vector()
   for(i in 1:length(OTUs[[k]])){

afterChull<-rbind(afterChull,forChull[forChull[,3]==OTUs[[k]][i],][chull(forChull[forChull[,3]==OTUs[[k]][i],1:2]),])
   }
   g <- ggplot(data=forChull, aes(x=`PC1`,y=`PC2`))
   g <- g + geom_polygon(data=afterChull, aes(x=`PC1`,y=`PC2`, fill = `OTUS` ), alpha = 0.3)
   g <- g + geom_point(data=forChull, aes(x=`PC1`,y=`PC2`, color = OTUS, fill=OTUS), size=1.5)
   # g <- g + geom_text(data=forChull, aes(x=`Comp 1`,y=`Comp 2`, color = OTUS, label=Labels), hjust = -0.5, nudge_x = 0.0) # uncomment this line if you want to check the lables associated to each point (IDs)
   # g <- g + scale_shape_manual(values = myShapes)
   g <- g + scale_colour_manual(values = myColors)
   g <- g + scale_color_discrete(name = '')
   g <- g + colScale + fillScale
   g <- g + theme(legend.direction = 'horizontal', legend.position = 'top')
   g <- g + guides(shape="none")
   g <- g + labs(x = paste("PC1 (", round(prop_pca[1]*100,1), "%)", sep=""), y = paste("PC2 (", round(prop_pca[2]*100,1), "%)", sep=""))
   g1[[k]] <-g
```

```
   # pdf(paste("./06_PCAs/01_PCA_multi_variables-",Hlabel[k],"/PCA-plot2.pdf",sep=""), family="ArialMT",
width=8, height=11)
   #     print(g1)
   # dev.off()

   # plot pca for females - ALL variables
   pca_multi_F <- prcomp(x_multi_cln_noOut_noZ_noAllo_cln_noCor_F[[k]], scale.=T)
   dfscores_F <- as.data.frame(pca_multi_F$x)
   prop_pca_F = pca_multi_F$sdev^2/sum(pca_multi_F$sdev^2)

forChull<-cbind(dfscores_F[,1:2],OTUS=x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"],La
bels=x_cln_red[[Hlabel[k]]]$ID[x_cln_red[[Hlabel[k]]]$Sex=="F"])
   forChull$OTUS<-factor(forChull$OTUS,all_OTUs[all_OTUs%in%forChull$OTUS])
   afterChull<-vector()
   for(i in 1:length(OTUs[[k]])){

afterChull<-rbind(afterChull,forChull[forChull[,3]==OTUs[[k]][i],][chull(forChull[forChull[,3]==OTUs[[k]][i],1:2]
),])
   }
   g <- ggplot(data=forChull, aes(x=`PC1`,y=`PC2`))
   g <- g + geom_polygon(data=afterChull, aes(x=`PC1`,y=`PC2`, fill = `OTUS` ), alpha = 0.3)
   g <- g + geom_point(data=forChull, aes(x=`PC1`,y=`PC2`, color = OTUS, fill=OTUS), size=1.5)
   # g <- g + geom_text(data=forChull, aes(x=`Comp 1`,y=`Comp 2`, color = OTUS, label=Labels), hjust = -0.5,
nudge_x = 0.0) # uncomment this line if you want to check the lables associated to each point (IDs)
   # g <- g + scale_shape_manual(values = myShapes)
   g <- g + scale_colour_manual(values = myColors)
   g <- g + scale_color_discrete(name = '')
   g <- g + colScale + fillScale
   g <- g + theme(legend.direction = 'horizontal', legend.position = 'top')
   g <- g + guides(shape="none")
   g <- g + labs(x = paste("PC1 (", round(prop_pca_F[1]*100,1), "%)", sep=""), y = paste("PC2 (",
round(prop_pca_F[2]*100,1), "%)", sep=""))
   g2[[k]] <-g
   # pdf(paste("./06_PCAs/01_PCA_multi_variables-",Hlabel[k],"/PCA-plot-F2.pdf",sep=""), family="ArialMT",
width=8, height=11)
   #     print(g2)
   # dev.off()

   # plot pca for males - ALL variables
   pca_multi_M <- prcomp(x_multi_cln_noOut_noZ_noAllo_cln_noCor_M[[k]], scale.=T)
   dfscores_M <- as.data.frame(pca_multi_M$x)
   prop_pca_M = pca_multi_M$sdev^2/sum(pca_multi_M$sdev^2)

forChull<-cbind(dfscores_M[,1:2],OTUS=x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"],
Labels=x_cln_red[[Hlabel[k]]]$ID[x_cln_red[[Hlabel[k]]]$Sex=="M"])
   forChull$OTUS<-factor(forChull$OTUS,all_OTUs[all_OTUs%in%forChull$OTUS])
   afterChull<-vector()
   for(i in 1:length(OTUs[[k]])){

afterChull<-rbind(afterChull,forChull[forChull[,3]==OTUs[[k]][i],][chull(forChull[forChull[,3]==OTUs[[k]][i],1:2]
),])
   }
   g <- ggplot(data=forChull, aes(x=`PC1`,y=`PC2`))
   g <- g + geom_polygon(data=afterChull, aes(x=`PC1`,y=`PC2`, fill = `OTUS` ), alpha = 0.3)
   g <- g + geom_point(data=forChull, aes(x=`PC1`,y=`PC2`, color = OTUS, fill=OTUS), size=1.5)
```

```
   # g <- g + geom_text(data=forChull, aes(x=`Comp 1`,y=`Comp 2`, color = OTUS, label=Labels), hjust = -0.5,
nudge_x = 0.0) # uncomment this line if you want to check the lables associated to each point (IDs)
   # g <- g + scale_shape_manual(values = myShapes)
   g <- g + scale_colour_manual(values = myColors)
   g <- g + scale_color_discrete(name = '')
   g <- g + colScale + fillScale
   g <- g + theme(legend.direction = 'horizontal', legend.position = 'top')
   g <- g + guides(shape="none")
   g <- g + labs(x = paste("PC1 (", round(prop_pca_M[1]*100,1), "%)", sep=""), y = paste("PC2 (",
round(prop_pca_M[2]*100,1), "%)", sep=""))
   g3[[k]] <-g
   # pdf(paste("./06_PCAs/01_PCA_multi_variables-",Hlabel[k],"/PCA-plot-M2.pdf",sep=""), family="ArialMT",
width=8, height=11)
   #     print(g3)
   # dev.off()

   # plot all combined
   pdf(paste("./06_PCAs/01_PCA_multi_variables-",Hlabel[k],"/PCA-plot-combined-PC1y2-nNEW.pdf",sep=""),
family="ArialMT", width=8, height=11)
         grid.arrange(g1[[k]] , g2[[k]] , g3[[k]] , nrow = 3, ncol = 1)
   dev.off()

}

########################################

# BOXPLOT + HISTOGRAM - FINAL

########################################

x_all_cln_noOut_noZ<-x_multi_cln_noOut_noZ_noAllo_cln_noCor

# Boxplot + histogram for ALL variables - per OTUs
# Final
for(k in 1:length(Hlabel)){
   dir.create(paste("./04_boxplots/03_all_variables-FINAL-",Hlabel[k],sep=""),showWarnings = FALSE)
   #loop for each variable
   for(i in 1:ncol(x_all_cln_noOut_noZ[[k]])){
         #open pdf file

pdf(paste("./04_boxplots/03_all_variables-FINAL-",Hlabel[k],"/boxplot_",colnames(x_all_cln_noOut_noZ[[k]])[i],"
.pdf",sep=""),family="ArialMT", width=9, height=12)
                  #plot boxplot
                  par(mar = c(7, 4, 4, 2) + 0.1)
                  boxplot(x_all_cln_noOut_noZ[[k]][,i] ~
x_cln_red[[Hlabel[k]]][,Hindex[k]],main=colnames(x_all_cln_noOut_noZ[[k]])[i],outline=F,ylim=c(min(na.omit(x_
all_cln_noOut_noZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),boxwex = 0.8,xaxt = "n", xlab=
Hlabel[k], ylab ="")
   #        axis(1, labels = FALSE)
                  xlabels <- levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])
                  text(1:length(levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])), par("usr")[3], offset=1, srt = 45, adj =
1, labels = xlabels, xpd = TRUE)
                  #loop for each OTUs
                  for(j in 1:length(levels(x_cln_red[[Hlabel[k]]][,Hindex[k]]))){
```

```
if(length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][
,Hindex[k]])[j]]))==0){
                                        #plot N for variable
                                        text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][,
Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                }
                                else{
                                        #transform counts to boxplot coordinates

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]][,Hindex[k]]==xlabels[j],i]))
                                        m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                                        if(m!=-Inf){
                                                y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                                        }else{
                                                if(length(q)>1){
                                                        y=q
                                                        y[!is.na(y)]<-j+0.4
                                                }else{
                                                        y=j+0.4
                                                }
                                        }
                                        #plot points from histogram
                                        if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]][,Hindex[k]]==xlabels[j],i])),col="blue",c
ex=0.8)
                                        }else{
                                                # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]][,Hindex[k]]==xlabels[j],i]),"dimna
mes")[[1]]),col="blue",cex=0.8)
                                        }
                                        #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[
[Hlabel[k]]][,Hindex[k]])[j]])),j+0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k
]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="red")
                                        #plot N for variable

text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]
][,Hindex[k]])[j]])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][,
Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                        #get range amplitude

rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))-min(na.omit(x_all_cln_noOut_noZ[[k]][,i]))
                                        #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]]))-(0.01*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]]
[,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.01*rango),col="grey",adj=0)
                                        #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
```

```
Hlabel[k]]][,Hindex[k]])[j]])),j-0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]=
=levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][
,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlab
el[k]]][,Hindex[k]])[j]])),j,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x
_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlab
el[k]]][,Hindex[k]])[j]]))-(0.02*rango),j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[
k]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]])),j+0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]=
=levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]]
[,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]
][,Hindex[k]])[j]]))-(0.025*rango),labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabel
[k]]][,Hindex[k]])[j]]))-(0.03*rango),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabe
l[k]]][,Hindex[k]])[j]]))-(0.03*rango),labels=min(na.omit(q)),cex=0.8)
                                }
                        }
                dev.off()
        }
}

########################################

# Boxplot + histogram for ALL variables - per Sex (if dimorphic)- per OTUs
# Need to run/load test for sexual dimorphism first
for(k in 1:length(Hlabel)){
   dir.create(paste("./04_boxplots/04_all_variables-FINAL-",Hlabel[k],"-sex",sep=""),showWarnings = FALSE)
   #loop for each variable
   for(i in 1:ncol(x_all_cln_noOut_noZ[[k]])){
           if(as.numeric(dimorphism[[k]][i,2])>=0.05){
                   #open pdf file

pdf(paste("./04_boxplots/04_all_variables-FINAL-",Hlabel[k],"-sex/boxplot_",colnames(x_all_cln_noOut_noZ[[k]]
)[i],".pdf",sep=""),family="ArialMT", width=3+length(OTUs[[k]]), height=14)
                                #plot boxplot
                                boxplot(x_all_cln_noOut_noZ[[k]][,i] ~
x_cln_red[[Hlabel[k]]][,Hindex[k]],main=colnames(x_all_cln_noOut_noZ[[k]])[i],outline=F,ylim=c(min(na.omit(x_
```

```
all_cln_noOut_noZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),boxwex = 0.8,xaxt = "n", xlab=
Hlabel[k], ylab ="")
                        xlabels <- levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])
                        text(1:length(levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])), par("usr")[3], offset=1, srt =
45, adj = 1, labels = xlabels, xpd = TRUE)
                        #loop for each OTUs
                        for(j in 1:length(levels(x_cln_red[[Hlabel[k]]][,Hindex[k]]))){

if(length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]==levels(x_cln_red[[Hlabel[k]]][
,Hindex[k]])[j]]))==0){
                                #plot N for variable
                                text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]==levels(x_cln_red[[Hlabel[k]]][,
Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                }
                                else{

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]][,Hindex[k]==xlabels[j],i]))
                                #transform counts to boxplot coordinates
                                m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                                if(m!=-Inf){
                                        y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                                }else{
                                        if(length(q)>1){
                                                y=q
                                                y[!is.na(y)]<-j+0.4
                                        }else{
                                                y=j+0.4
                                        }
                                }
                                #plot points from histogram
                                if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]][,Hindex[k]==xlabels[j],i])),col="blue",c
ex=0.8)
                                }else{
                                        # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]][,Hindex[k]==xlabels[j],i]),"dimna
mes")[[1]]),col="blue",cex=0.8)
                                }
                                #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]==levels(x_cln_red[
[Hlabel[k]]][,Hindex[k]])[j]])),j+0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]
]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="red")
                                #plot N for variable

text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]==levels(x_cln_red[[Hlabel[k]]
][,Hindex[k]])[j]])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]==levels(x_cln_red[[Hlabel[k]]][,
Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                #get range amplitude

rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))-min(na.omit(x_all_cln_noOut_noZ[[k]][,i]))
                                #plot axis for histogram
```

```
segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]])))-(0.01*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]]
[,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])))-(0.01*rango),col="grey",adj=0)
                                        #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]])),j-0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]=
=levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][
,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlab
el[k]]][,Hindex[k]])[j]])),j,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x
_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlab
el[k]]][,Hindex[k]])[j]]))-(0.02*rango),j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[
k]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]])),j+0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]=
=levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[
Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]]
[,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabel[k]]
][,Hindex[k]])[j]]))-(0.025*rango),labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabel
[k]]][,Hindex[k]])[j]]))-(0.03*rango),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][,i][x_cln_red[[Hlabel[k]]][,Hindex[k]]==levels(x_cln_red[[Hlabe
l[k]]][,Hindex[k]])[j]]))-(0.03*rango),labels=min(na.omit(q)),cex=0.8)
                                        }
                                }
                        dev.off()
                } else {
                        #open pdf file

pdf(paste("./04_boxplots/04_all_variables-FINAL-",Hlabel[k],"-sex/boxplot_",colnames(x_all_cln_noOut_noZ[[k]]
)[i],".pdf",sep=""),family="ArialMT", width=10+length(OTUs[[k]]), height=14)
                                layout(matrix(c(1,2),1,2))
                                #plot boxplot for females
                                boxplot(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i] ~
x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"],main=colnames(x_all_cln_noOut_noZ[[k]])[
i],outline=F,ylim=c(min(na.omit(x_all_cln_noOut_noZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),box
wex = 0.8,xaxt = "n", xlab= paste(Hlabel[k]," Females",sep=""), ylab ="")
                                xlabels <- levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])
```

```
                        text(1:length(levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])), par("usr")[3], offset=1, srt =
45, adj = 1, labels = xlabels, xpd = TRUE)
                        #loop for each OTUs
                        for(j in 1:length(levels(x_cln_red[[Hlabel[k]]][,Hindex[k]]))){

if(length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k
]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))==0){
                                #plot N for variable

text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]
][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                        }else{

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hinde
x[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))
                                # q[q==0]<-NA
                                #transform counts to boxplot coordinates
                                m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                                if(m!=-Inf){
                                        y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                                }else{
                                        if(length(q)>1){
                                                y=q
                                                y[!is.na(y)]<-j+0.4
                                        }else{
                                                y=j+0.4
                                        }
                                }
                                #plot points from histogram
                                if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hi
ndex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="blue",cex=0.8)
                                }else{
                                        # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]
][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]),"dimnames")[[1]
]),col="blue",cex=0.8)
                                }
                                #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]
][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),j+0.4,mean(na.o
mit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[
Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="red")
                                #plot N for variable

text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[
k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]
][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                #get range amplitude
```

```
rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i]))-min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i]))
                                        #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.01*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.01*rango),col="grey",adj=0)
                                        #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),j-0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),j,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),j+0.4,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.025*rango),labels="Count",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.03*rango),labels=max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="F",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="F"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.03*rango),labels=min(na.omit(q)),cex=0.8)
```

```
                        }
                }
                #plot boxplot for males
                boxplot(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i] ~
x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"],main=colnames(x_all_cln_noOut_noZ[[k]])
[i],outline=F,ylim=c(min(na.omit(x_all_cln_noOut_noZ[[k]][,i])),max(na.omit(x_all_cln_noOut_noZ[[k]][,i]))),box
wex = 0.8,xaxt = "n", xlab= paste(Hlabel[k]," Males",sep=""), ylab ="")
                xlabels <- levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])
                text(1:length(levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])), par("usr")[3], offset=1, srt =
45, adj = 1, labels = xlabels, xpd = TRUE)
                #loop for each OTUs
                for(j in 1:length(levels(x_cln_red[[Hlabel[k]]][,Hindex[k]]))){

if(length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[
k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))==0){
                        #plot N for variable

text(j,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k
]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                        }else{

q<-as.numeric(table(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hinde
x[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))
                        #transform counts to boxplot coordinates
                        m<-((j-0.4)-(j+0.4))/(max(na.omit(q))-min(na.omit(q)))
                        if(m!=-Inf){
                                y<-(m*q)+((j+0.4)-(m*min(na.omit(q))))
                        }else{
                                if(length(q)>1){
                                        y=q
                                        y[!is.na(y)]<-j+0.4
                                }else{
                                        y=j+0.4
                                }
                        }
                        #plot points from histogram
                        if(length(q)<2){

points(y,unique(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hi
ndex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="blue",cex=0.8)
                        }else{
                                # points(y,b$mids,col="blue")

points(y,as.numeric(attr(table(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]
]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]),"dimnames")[[
1]]),col="blue",cex=0.8)
                        }
                        #plot mean as a line

segments(j-0.4,mean(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]
]][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),j+0.4,mean(na
.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_re
d[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="red")
                        #plot N for variable
```

```
text(j,max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex
[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])), labels=paste("N =
",length(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k
]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),sep=""),cex=0.8,pos=3)
                                        #get range amplitude

rango<-max(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i]))-min(na.omit(x_all_cln_noO
ut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i]))
                                        #plot axis for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]]
[,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.01*rango),j+
0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k]
][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.01*rango),col="grey",ad
j=0)
                                        #plot max line for histogram

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]]
[,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),j-0.4,max(na.o
mit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[
[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]]
[,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j-
0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k]
][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot mean line for histogram

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hi
ndex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),j,max(na.omit(x_all
_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[[Hlabel[
k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hi
ndex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j,min(
na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_
red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot min line for histogram

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]
][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),j+0.4,max(na.o
mit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[k]][x_cln_red[
[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),lty=3,col="grey")

segments(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]
][,Hindex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.02*rango),j
+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[
k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]])),col="grey")
                                        #plot labels for the histogram axis

text(j,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hindex[
k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.025*rango),labels="C
ount",cex=0.8)

text(j-0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hin
```

```
dex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.03*rango),labels=
max(na.omit(q)),cex=0.8)

text(j+0.4,min(na.omit(x_all_cln_noOut_noZ[[k]][x_cln_red[[Hlabel[k]]]$Sex=="M",i][x_cln_red[[Hlabel[k]]][,Hi
ndex[k]][x_cln_red[[Hlabel[k]]]$Sex=="M"]==levels(x_cln_red[[Hlabel[k]]][,Hindex[k]])[j]]))-(0.03*rango),labels
=min(na.omit(q)),cex=0.8)
                                        }
                                    }
                        dev.off()
                }
    }
}

########################################
```

Appendix II. TNT-script for maximum parsimony phylogenetic analysis.

```
macro= ;            /*  enable scripting and comments   */

/*#####################################*/
/* TE Analysis using MP in TNT */
/* 2023.11.30*/
/*#####################################*/

/*#############################################################################################*/
/* Equal Weights TE */
/*#############################################################################################*/

/* piwe off */
piwe -;

/* delete trees in memory */
keep 0;

/* load matrix */
run parsimony.tnt;

/* load matrix */
/* set Heterodon as outgroup */

/* open log */
log TE_EW-LOG.txt

/* run New Tech search */
xmult = consense 5;

/* final bbreak */
bb;

/* close log file */
log /;

/* export all trees */
```

```
export - TE_EW-xmult_cons5-ALL.tre;

/* save all trees */
tsave *TE_EW-xmult_cons5-ALL.tnt;
save /;
tsave /;

/* consensus saved as last tree */
ne *;

/* keep last tree only */
tchoose /;

/* export NE tree */
export - TE_EW-xmult_cons5-NE.tre;

/* save NE tree */
tsave *TE_EW-xmult_cons5-NE.tnt;
save /;
tsave /;

/* turn on tree tags */
ttags =;

/* run bootstrap */
resample boot frequency from 0 rep 100 nogc;

/* keep Boot tree only */
tchoose /;

/* export boot tree */
export - TE_EW-xmult-cons5-NE-BOOT.tre;

/* clear ttags */
ttags -;

/* turn on tree tags */
ttags =;

/* run jacknife */
resample jak sym from 0 rep 100 freq nogc;

/* keep Jack tree only */
tchoose /;

/* export Jack tree */
export - TE_EW-xmult-cons5-NE-JAK.tre;

/* close log file */
log /;

proc/;
```

Appendix III. RAxML-script for maximum likelihood phylogenetic analyses.

```
sudo apt-get install flex bison libgmp3-dev

git clone --recursive https://github.com/amkozlov/raxml-ng
cd raxml-ng
mkdir build && cd build
cmake ..
make

raxmlHPC-HYBRID-AVX -T 4 -f a -n result -s infile.txt -m GTRGAMMA -p 12345 -N 100 -x 12345
```

Appendix IV. R-script for ancestral character estimation and phylogeny plotting.

```
#####################################################
# Compare tree topologies
#####################################################
library(ape)
library(phytools)
#################################
#read trees#
p.mll<-read.tree("SupplementaryFilePHL.tre")
p.mpp<-read.tree("SupplementaryFilePHP.tre")

#extracting ML#
Ac01<-match("P_lemniscatus_Quarai_UFRGS6088",p.mll$tip.label)
Ph01<-match("P_lemniscatus_Bom_Jesus_UFRGS5351",p.mll$tip.label)

node<-getMRCA(p.mll,c(Ac01,Ph01))

pt2<-extract.clade(p.mll,node)
p.ml<-pt2

#extracting MP#
Ac02<-match("P_lemniscatus_Quarai_UFRGS6088",p.mpp$tip.label)
Ph02<-match("P_reticulatus_Bom_Jesus_UFRGS5194",p.mpp$tip.label)

node<-getMRCA(p.mpp,c(Ac02,Ph02))

pt1<-extract.clade(p.mpp,node)
p.mp<-pt1


# check if tip labels match
dif<-c(setdiff(p.ml$tip.label, p.mp$tip.label),setdiff(p.mp$tip.label, p.ml$tip.label))
dif

#################################
#comparing RAXML vs TNT topologies
comp.data<-data.frame(p.ml$tip.label,p.ml$tip.label)

# create cophylo object rotating nodes to match vertical position
obj<-cophylo(p.ml,p.mp,rotate=T)
```

```
###########


boot_values<-p.ml$node.label
boot_values[as.numeric(p.ml$node.label)<70]<-""

colfunc <- colorRampPalette(c("white", "black"))

morethan70<-data.frame(support=70:100,color=colfunc(31))

pie_color<-boot_values


for(i in seq(morethan70[,1])){
  pie_color[as.numeric(p.ml$node.label)==morethan70[i,1]]<-morethan70[i,2]
}


pie_color[boot_values==""]<-"#FFFFFF"

###########


boot_valuez<-p.mp$node.label
boot_valuez[as.numeric(p.mp$node.label)<70]<-""

colfunc <- colorRampPalette(c("white", "black"))

morethan70<-data.frame(support=70:100,color=colfunc(31))

pie_color<-boot_valuez


for(i in seq(morethan70[,1])){
  pie_color[as.numeric(p.mp$node.label)==morethan70[i,1]]<-morethan70[i,2]
}


pie_color[boot_valuez==""]<-"#FFFFFF"

# plot tree comparison
pdf("Phalotris_Fig_XX_Comparison_ML_vs_MP.pdf", width=14, height=16)
  plot(obj,no.margin=T, fsize=1.7, cex=0.7, label.offset=0.001, edge.width=2,
            link.type="curved",link.lty="dashed",
            link.col=make.transparent("red",0.95),link.lwd=1,
            pts=FALSE,part=0.47)
   nodelabels.cophylo(pch=21, bg=pie_color, adj=c(0.5,0.5), cex=1.8, frame = "n", which="left")
  nodelabels.cophylo(pch=21, bg=pie_color, adj=c(0.5,0.5), cex=1.8, frame = "n", which="right")
dev.off()

#####################################################

#
# # values plot tree comparison
# pdf("Suppl_Fig_XX_Comparison_ML_vs_MP.pdf", width=15, height=18)
# plot(obj,no.margin=T, fsize=1.7, cex=0.3, label.offset=0.001)
```

```
# nodelabels.cophylo(p.ml$node.label, cex=0.9, adj= c(-0.2,1.3), bg = g, frame = "n", which="left")
# nodelabels.cophylo(p.mp$node.label, cex=0.9, adj= c(-0.2,1.3), bg = g, frame = "n", which="right")
# dev.off()

############################

# ACE Discrete and Continuous

############################


setwd("~/MSC/ACE")
##########

library(ape)
library(phytools)
library(geiger)
library(diversitree)
############
p.ace<-read.tree("elatre.tre")
snake.tree<-force.ultrametric(p.ace)

snake.tree<-p.ace

total.diversity<-31

#Read ACE csv file #

X<-read.csv("SIZE.csv",row.names=1)

# #
#
# Y<-read.csv("DSN.csv",row.names=1)
# #

stripes<-as.matrix(X)[,1]

#
#
# numbers<-as.matrix(Y)[,1]

# Now we can estimate ancestral states.
# We will also compute variances & 95% confidence intervals for each node: #

fit<-fastAnc(snake.tree,stripes,vars=TRUE,CI=TRUE)
fit

#

fit<-fastAnc(snake.tree,numbers,vars=TRUE,CI=TRUE)
fit


obj<-contMap(snake.tree,stripes,plot=FALSE)
plot(obj,type="phylogram",legend=0.7*max(nodeHeights(snake.tree)),
        fsize=c(0.7,0.9))
```

```
#
# obj<-contMap(snake.tree,numbers,plot=FALSE)
# plot(obj,type="fan",legend=0.7*max(nodeHeights(snake.tree)),
#           fsize=c(0.7,0.9))
# plot #

obj<-contMap(snake.tree,stripes)

obj<-contMap(snake.tree,numbers)

######

APS<-read.csv("APS.csv",row.names=1)

stripes<-as.matrix(APS)[,1]

states <-factor(c("Absent","Present"))

states <-as.factor(states)

stripes.sorted<-stripes[match(snake.tree$tip.label,names(stripes))]

################################
# model fitting
# ################################
fitER<-fitDiscrete(snake.tree,as.factor(stripes),model="ER",ncores=4)
fitSYM<-fitDiscrete(snake.tree,as.factor(stripes),model="SYM",ncores=4)
fitARD<-fitDiscrete(snake.tree,as.factor(stripes),model="ARD",ncores=4)
#
# # compare models through AICc
aicc<-setNames(c(fitER$opt$aicc,fitSYM$opt$aicc,fitARD$opt$aicc),c("ER","SYM","ARD"))
#
# # make a table with AIC values
aic.table<-data.frame(AICc=aicc,AICcW=as.vector(aic.w(aicc)))
#
# # save AIC table
write.table(aic.table, file="table_AIC_CharEvol.txt", quote=FALSE, sep="\t")
#
# # best model (if equal values -> get first)
CharFitBest<-rownames(aic.table)[aic.table[,2]==max(aic.table[,2])][1]
#

ansER<-ace(stripes,snake.tree,model=CharFitBest,type="discrete")

cols<-setNames(c("red","blue"),levels(states))
#

#

pdf("Elapomorphini_ACE.pdf",family="Helvetica", width=6, height=6)

plot.phylo(snake.tree,cex=0.7,no.margin=TRUE,label.offset=0.5)
nodelabels(node=1:snake.tree$Nnode+Ntip(snake.tree),pie=ansER$lik.anc,piecol=cols,cex=0.6)
tiplabels(pie=to.matrix(stripes.sorted[snake.tree$tip.label],levels(states)),piecol=cols,cex=0.5)
legend("topleft",names(cols),pch=22,pt.bg=cols,bty="n",cex=0.8,pt.cex=1.2)
```

```
dev.off()
```

Appendix V. R-script for Procrustes analysis of skull morphology.

```
#packages
require(ape)
require(geomorph)
require(MASS)
require(Morpho)
require(phytools)
require(Rvcg)
require(stats)
require(vegan)
require(ggplot2)

#Setwork directory
setwd("OMAR2")

# load .tps file
tps<-readland.tps("DORSAL15.tps",specID = "ID", readcurves = FALSE)
dim(tps)
tps
## GPA (Generalized Procrustes Analysis
gpa.object<-gpagen(tps)
gpa.object
shape<-gpa.object$coords
size<-gpa.object$Csize

#Normality test
shapiro.test(size)#W = 0.92736, p-value = 0.2491
#Histogram plot
hist(size)

#Shape procedures evalueting
plotAllSpecimens(shape)

# Find outliers
plotOutliers(shape, inspect.outliers = TRUE)

# Correlation (shape space) (tangent space)
regdist(shape)

# Meanshape
ref<-mshape(shape)

# Landmark links
links<-define.links(ref,ptsize=3)
dev.off()

links.apos<-matrix(c(1,2,2,3,3,4,4,1,3,5,5,10,7,7,6,6,4,9,14,9,8,14,18,8,19,19,18,15,20,20,22,22,23,23,25,25,26,26,3
4,26,27,27,33,33,34,33,32),nrow=50,ncol=2,byrow=T)

# Plot specimens to mean
```

```
GP1<-gridPar(pt.bg="gray",link.col="gray",link.lty=1) # cor dos landmarks e links
plotRefToTarget(ref,shape[,,8],links=links.apos,method="TPS")
plotRefToTarget(ref,shape[,,8],links=links.apos,method="vector")
plotRefToTarget(ref,shape[,,8],links=links.apos,method="points",gridPars=GP1) # alvo = em preto, refer?ncia = em
cinza

plotAllSpecimens(shape,mean=TRUE,links=links.apos)

#########################################################################################################
#########################################################################################
#Classfier data frame
class<-data.frame("species"=c("Phalotris_ilustrator","Phalotris_ilustrator","Phalotris_lemniscatus","Phalotris_lemni
scatus","Phalotris_lemniscatus",

"Phalotris_reticulatus","Phalotris_reticulatus","Phalotris_reticulatus","Phalotris_spegazzinii","Phalotris_spegazzinii
",

"Phalotris_spegazzinii","Phalotris_suspectus","Phalotris_suspectus","Phalotris_trilineatus","Phalotris_trilineatus"),si
ze)
species<-class$species

# MANOVA COMPARING SPECIES SHAPE
MANOVA<-procD.lm(shape~size+species, iter=999)
summary(MANOVA)

#           Df         SS         MS        Rsq       F           Z Pr(>F)
#species 5 0.015829 0.0031658 0.59046 2.5951 3.2644  0.002 **
#Residuals  9 0.010979 0.0012199 0.40954
#Total    14 0.026808

#           Df         SS         MS        Rsq       F           Z Pr(>F)
#size      1 0.0012831 0.0012831 0.04786 1.0551 0.2700  0.402
#species 5 0.0157958 0.0031592 0.58922 2.5978 3.1664  0.002 **
#Residuals  8 0.0097289 0.0012161 0.36291
#Total    14 0.0268078



# PCA
PCA_omar<-gm.prcomp(shape)

gf<-data.frame(PC1=PCA_omar$x[,1],PC2=PCA_omar$x[,2], species, size)


PlotPCA<- ggplot(gf, aes(x = PC1, y = PC2, color = species, fill=species)) +
  geom_point(size = 3 * 2, shape = 21) +
  labs(title = "") +
  ylab("PC2 (23.98%)") +
  xlab("PC1 (30.47%)")

# Graph data
plot_data <- ggplot_build(PlotPCA)$data[[1]]

# Colorscale
cores_do_grafico <- plot_data$colour
hullcol<-unique(cores_do_grafico)
```

```
#########################################################################################################
########

#Shapes for PCA

PC.sc<-PCA_omar$x[,1]
PC.sc2<-PCA_omar$x[,2]
sc.preds1<-shape.predictor(shape,x=PC.sc,Intercept = FALSE,pred1 = min(PC.sc), pred2 = max(PC.sc))
sc.preds2<-shape.predictor(shape,x=PC.sc2,Intercept = FALSE,pred1 = min(PC.sc2), pred2 = max(PC.sc2))



##Colored PCA skull
pdf("PCA_dorsal skullview2.pdf", width = 8, height=5)
xlab<-paste("Principal Component 1 (30.47%)") # Axis 1 Label
ylab<-paste("Principal Component 2  (23.98%)") # Axis 2 Label
mat<-matrix(c(4,5,0,1,1,2,1,1,3),3) # Plot split
layout(mat, widths=c(1,1,1), heights=c(1,1,1))
par(mar=c(4, 5, 1, 1), cex.lab=2, cex.axis=1.5)
plot(PCA_omar$x[,1],PCA_omar$x[,2],pch=21,cex=3,bg=cores_do_grafico,asp=T,xlab=xlab,ylab=ylab)
ordihull(PCA_omar$x,species,draw="polygon", col=hullcol , alpha=60)
plotRefToTarget(ref,sc.preds1$pred1,mesh=refmesh,links=links.apos,method="TPS", size=3,mag=1)
plotRefToTarget(ref,sc.preds1$pred2,mesh=refmesh,links=links.apos,method="TPS", size=3,mag=1)
plotRefToTarget(ref,sc.preds2$pred2,mesh=refmesh,links=links.apos,method="TPS", size=3,mag=1)
plotRefToTarget(ref,sc.preds2$pred1,mesh=refmesh,links=links.apos,method="TPS", size=3,mag=1)
par(xpd=NA)  # Allows legend outside plot
legend("topright", inset = c(0, 1), legend = levels(as.factor(species)), fill = unique(cores_do_grafico), title =
"Species", bg = "white")
dev.off()
# odd plot


GP1<-gridPar(pt.bg="darkolivegreen3",link.col="darkolivegreen3",link.lty=1,link.lwd=3) # Landmark color e
linksGP1<-gridPar(pt.bg="gray",link.col="gray",link.lty=1) # Landmark color


pred_alt_p<-shape.predictor(plot_alo_ALT_p$GM$fitted, x=Databy_Spp_p$BOTALT$Length, Intercept=TRUE,
predmin=min(Databy_Spp_p$BOTALT$Length),predmax=max(Databy_Spp_p$BOTALT$Length))



shape.predictor(PCA_omar$shapes)


# Calculating mean shape by species
shape.2d<-two.d.array(shape)
shape.2d.means<-rowsum(shape.2d,species)/as.vector(table(species))
shape.means<-arrayspecs(shape.2d.means,dim(shape)[1],dim(shape)[2])
shape.means

# Nexus tree
tree<-read.nexus("nexus.tree.sig.nex")
plot(tree)
tree<-compute.brlen(tree,1) # branch length = 1

# Plot (phylomorphoespace)
```

```
pms<-plotGMPhyloMorphoSpace(tree,shape.means,plot.param=list(t.bg="green",txt.col="black",n.bg="black",n.ce
x=1,lwd=2,l.col="blue"))

dev.off()
```

CONCLUSÕES GERAIS

Através do presente estudo, é possível concluir que as linhas de evidência em morfologia externa (morfometria, folidose), morfologia interna (osteologia), e molecular (transcriptoma de glândulas de veneno, sequências de DNA) sugerem uma relação de sinonímia entre *P. lemniscatus*, *P. trilineatus*, e *P. reticulatus*, considerando o primeiro táxon como sinônimo sênior. O status taxonômico de outras espécies consideradas válidas (*P. bilineatus*, *P. illustrator*, *P. suspectus*, *P. spegazzinii*, *P. normanscotti*) permanece amparado a partir de caracteres morfológicos demonstrados como variáveis e polimórficos neste estudo. Uma proposta de sinonímia destas espécies transcende o escopo do trabalho atual, e deve ser precedida por uma análise criteriosa de mais dados morfológicos e moleculares.