Research Letters

# Spatial bias in sampling small rodents in the Atlantic Forest: A landscape and accessibility perspective

Thadeu Sobral-Souza [b],[*],[1], Nicolas Silva Bosco [a],[1], Lana Pavão Candelária [a], Rosane Garcia Collevatti [d],[e], Viviane Maria Guedes Layme [a], Domingos de Jesus Rodrigues [c]

[a] *Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade, Instituto de Biociências, Universidade Federal de Mato Grosso, Av. Fernando Corrêa 2367, E-78060-900 Cuiabá, Mato Grosso, Brazil*

[b] *Laboratório de Macroecologia e Conservação da Biodiversidade, Departamento de Botânica e Ecologia, Instituto de Biociências, Universidade Federal de Mato Grosso, Av. Fernando Corrêa 2367, E-78060-900 Cuiabá, Mato Grosso, Brazil*

[c] *Núcleo de Estudos da Biodiversidade da Amazônia Mato-Grossense (NEBAM), Universidade Federal de Mato Grosso, Avenida Alexandre Ferronato - de 1002 a 1530 - lado par, E-78550728 Sinop, Mato Grosso, Brazil*

[d] *Laboratório de Genética & Biodiversidade, Instituto de Ciências Biológicas, Universidade Federal de Goiás - UFG, Goiânia, Goiás, Brazil*

[e] *Laboratório de Ecologia Espacial (LEEC), Departamento de Biodiversidade, Instituto de Biociências, Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP), Rio Claro, São Paulo, Brazil*

## HIGHLIGHTS

- Less than 1% of the Atlantic Forest spatial surface is well sampled for small rodents.
- Most sampling sites are close to roads, urban centers.
- Well-sampled sites are in larger fragments, and in landscapes with higher percentage of forest cover.
- Knowledge of small mammal occurrence at a large-scale is spatially limited.
- Sampling of small rodents in Atlantic Forest should prioritize small remote fragments, representing current common landscape conditions.

## GRAPHICAL ABSTRACT
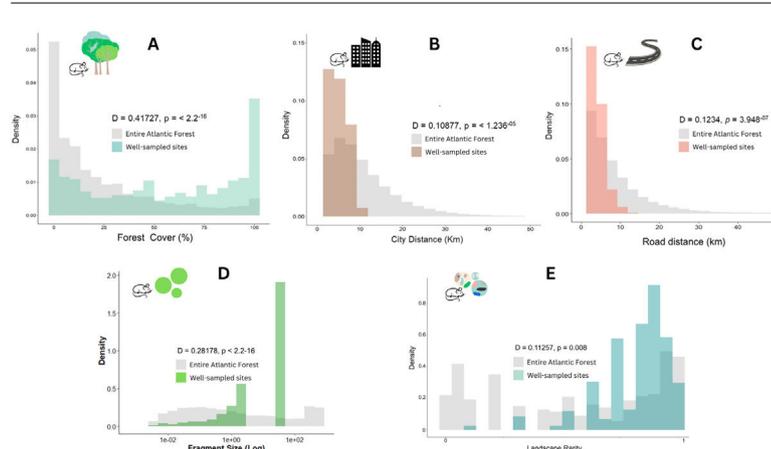


## ARTICLE INFO

## ABSTRACT

Understanding the effects of habitat loss and fragmentation on species spatial distribution is challenging, mainly because knowledge of species occurrences is biased. Survey efforts are unevenly distributed causing spatial sampling biases that are normally neglected. Assessing sampling bias is particularly urgent for threatened ecoregions, such as the Atlantic Forest, a global biodiversity hotspot. Here, we assess spatial sampling biases of small rodents in the Atlantic Forest, using an integrative approach with accessibility and landscape metrics. We built a robust dataset of 11,495 primary records of the Atlantic Forest's small rodent species, based on information from digitally accessible repositories. We expect that well-sampled sites are spatially aggregated and nearer roads, urban centers, on landscapes with larger forest fragments, and with higher percentage of forest cover. We also expect gaps of small rodents sampling in rare landscape conditions. Our results indicated that only less than 1% of the Atlantic Forest (at 1 km² cell-size

resolution) are well sampled. Following our expectations, the well-sampled sites were spatially aggregated biased toward roads, urban centers, larger forest fragments, and landscapes with higher percentage of forest cover. We also found a survey gap on common landscape conditions. Our findings suggest that the spatial distribution of small rodents at landscape level ($1\,km^2$) remains unknown across most of the Atlantic Forest spatial extension. Our findings also point to new priority sites for small mammals sampling on common landscape conditions, in smaller fragments and on remote areas improving spatial distribution knowledge and contributing to conservation policies at landscape level.

## Introduction

Human-induced activities have modified landscapes, directly affecting the spatial distribution of species worldwide (Ceballos and Ehrlich, 2002; Chase et al., 2020). Natural habitat modification may cause defaunation (Dirzo et al., 2014), and species population declines or extinctions (Dirzo and Raven, 2003; Hanski, 2011). Understanding how human-induced habitat changes drive biodiversity loss is still a challenge for science because of the several gaps in biodiversity knowledge that need to be better understood and quantified (Hortal et al., 2015).

Primary biodiversity data on species occurrences are still scarce or spatially biased (Girardello et al., 2018). The Wallacean Shortfall is characterized by uneven distribution of sampling effort over time and space, because some localities are temporally more sampled while many others remain unsampled, generating a spatially clustered pattern in biodiversity knowledge (Hortal et al., 2015; Almeida et al., 2021; Sobral-Souza et al., 2021a). Though, Wallacean Shortfall is dependent on the spatial scale and cell-size resolution of sampling effort assessment (Hortal et al., 2015; Bosco et al., 2022). Bias at local scale normally occurs because sampling tends to be performed close to access routes, whether in terrestrial or aquatic ecosystems (Sousa-Baena et al., 2014; Oliveira et al., 2016), and close to urban centers (Lessa et al., 2019; Almeida et al., 2021). Besides accessibility-related biases, landscape-based biases lead to greater sampling near large and connected forest fragments (Sobral-Souza et al., 2021a), or in protected areas (Almeida et al., 2021). These biases can be explained by the interest of researchers in sampling a greater number of species (Reddy and Dávalos, 2003; Sastre and Lobo, 2009) in a shorter temporal period, especially in intensively modified landscapes.

Global initiatives have emerged to address the problems of limited biodiversity data and their quality, such as making such data available on digitally accessible platforms (e.g., Global Biodiversity Information Facility - GBIF and speciesLink). Digital biodiversity knowledge has been important for understanding patterns of the spatial distribution of biodiversity; however, data gaps and biases can compromise the interpretation of large-scale biodiversity patterns (Hortal et al., 2007), thereby confounding conservation actions (Brooks et al., 2004; Sousa-Baena et al., 2014). Spatial bias can also distort estimated spatial patterns in biodiversity (Hortal et al., 2008; Boakes et al., 2010; Yang et al., 2013) and hinder the use of niche-based modeling techniques (Hortal et al., 2008), which also emerged as a technique to solve the problem of limited biodiversity data (Beck et al., 2014; Rocchini et al., 2011). Shortcomings of biodiversity data are most prominent for high impacted and biodiverse ecoregions, such as the global biodiversity hotspots.

The Atlantic Forest is one of the global biodiversity hotspots (Myers et al., 2000) yet, it still has limited biodiversity knowledge (Collen et al., 2008; Dirzo et al., 2014). The Atlantic Forest is the most studied Brazilian ecoregion due to its history of colonization, natural resources exploitation, its richness of endemic species, as well as the presence of large urban centers with several research institutions (Teixido et al., 2020; Candelária et al., 2021). Even so, biodiversity knowledge of the Atlantic Forest is biased for some regions (Sobral-Souza et al., 2021a). Testing the biases and gaps in biodiversity knowledge of the Atlantic Forest using well resolved

taxonomic groups, with extensive information of species distributions and sampling inventories such as small rodents, can help improving conservation strategies for the taxonomic group and the ecoregion.

Small rodents are widely distributed (Wilson and Reeder, 2005) and considered a diverse taxon in the Atlantic Forest, with 94 described species (Bovendorp et al., 2017; Galindo-Leal and Câmara, 2003). Ordinarily, small rodent species are generalists and can occupy different habitat types (Bonvicino et al., 2008). They may also respond differently to local habitat changes (Bergallo et al., 2005; Rosa et al., 2018), because they are dependent on habitat quality (Delciellos et al., 2016). Even though the Atlantic Forest is the most surveyed ecoregion for small rodents in Brazil (Candelária et al., 2021), sampling bias and spatial gaps have not yet been assessed.

In fact, problems of biodiversity data quality and quantity occur in all domains of ecology (Hortal et al., 2015). Therefore, identifying gaps and biases in biodiversity data is important for indicating priority areas for sampling that can help better understand large-scale biodiversity patterns. Here, we assess spatial biases and gaps in Atlantic Forest's small rodent species occurrence data. Specifically, we quantify biodiversity knowledge of small rodents to test the hypothesis that accessibility and landscape structure explain sampling biases and gaps. We expect that small rodent surveys are spatially biased towards more accessible regions close to urban centers and roads, a pattern found for other taxa (e.g., Oliveira et al., 2016 – for different taxa groups, Sobral-Souza et al., 2021a, for butterflies). Small rodents can respond relatively fast to landscape changes (e.g., Pardini, 2004; Püttker et al., 2008), therefore they are good models for testing hypotheses related to biodiversity and species loss. In macroecological perspectives, higher forest cover and more connected remnants are commonly more sampled than others (Sobral-Souza et al., 2021a). Thus, we expect to find more surveys in larger fragments (ha) and in landscapes with greater forest cover. Currently, Atlantic Forest's landscapes comprise mainly small fragments far from each other (Vancine et al., 2024), landscape heterogeneity is very common (Carvalho et al., 2021). Thus, we also expect gaps of smalls rodents' sampling in common landscape conditions, i.e., landscapes with small fragments and smaller forest cover.

## Material and methods

### Study area

The Atlantic Forest extends along the Brazilian coast, from the northeast to south, and a small part of Paraguay and Argentina in the southwest (Ribeiro et al., 2009; Tabarelli et al., 2010). The historical and geological context of the Atlantic Forest explains its environmental heterogeneity, which has contributed to its high species richness and rates of endemism (Ribeiro et al., 2009). Silva and Casteleti (2003) proposed biogeographic subdivisions for the Atlantic Forest, namely: Serra do Mar, Pernambuco, Bahia, Northeastern Brejos, Diamantina, Interior Forests, Araucaria Forests and São Francisco. These subregions are characterized by the presence of two or more endemic species with overlapping spatial distribu-
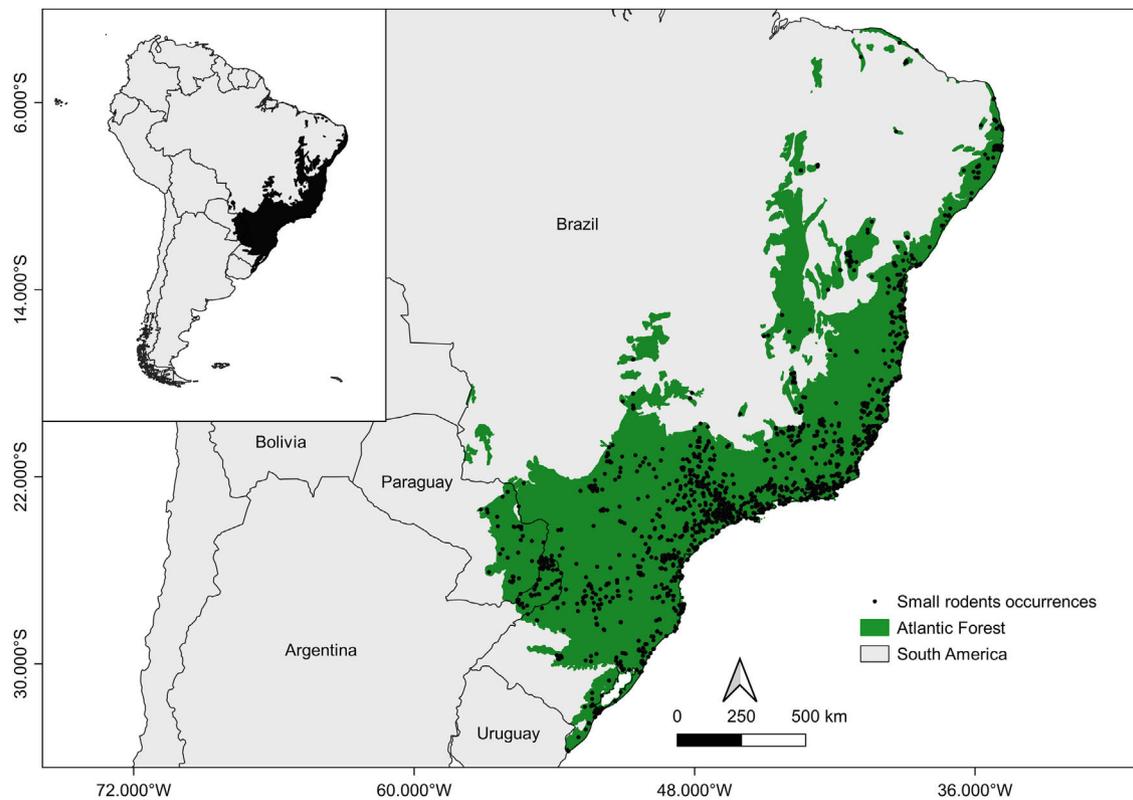
**Fig. 1.** Spatial distribution of small rodents' occurrences compiled for the Atlantic Forest (shapefile from Muylaert et al., 2018b). The thumbnail map depicts the Atlantic Forest (dark area) in South America.

tions or by the specificities of some habitats (Silva and Casteleti, 2003). Currently, 28 % of the Atlantic Forest is composed of small and isolated forest remnants (Rezende et al., 2018). The Serra do Mar is the most conserved biogeographic subregion (about 36.5% of remaining forest area), while the Interior and São Francisco subregions are the most deforested (<8% of forest area remaining; Ribeiro et al., 2009).

*Species dataset*

We built a dataset containing primary records of small rodent species of the families Caviidae, Cricetidae, Ctenomyidae, Echimyidae, Erethizontidae, Muridae and Sciuridae. The primary records were obtained from different digitally accessible databases *Global Biodiversity Information Facility* (GBIF 2021- DOI https://doi.org/10.15468/dl.bnbeqo), speciesLink (2021), Information System on Brazilian Biodiversity (*Sistema de Informação sobre a Biodiversidade Brasileira* - SiBBr 2021) and Biodiversity Portal (*Portal da Biodiversidade* - PortalBio 2021). We also used two published databases, Bovendorp et al. (2017) and Figueiredo et al. (2017), that compiled information on richness and abundance of non-flying small mammals from scientific articles, theses and unpublished data. Additionally, we included the primary records of six invasive small rodents' species from the Neotropical Invasive Mammals database (Rosa et al., 2020). After, we filtered the primary records based on: i) taxonomic accuracy, and removed all the records not identified to species level ; ii) spatial accuracy, and excluded sites outside the Atlantic Forest boundaries (Muylaert et al., 2018a), or lack of georeferenced information; iii) duplicate records, and removed occurrences with identical georeferenced sites, dates, voucher numbers, or included in two or more datasets accessed by us; and iv) primary data sampled after years 2000, and excluded data sampled before 2000, due to the low changes in land use and the stability of the road network after this period. We checked

and corrected synonyms and current taxonomy using the ASM Mammals Diversity Database of the American Society of Mammalogists (ASM, https://www.mammaldiversity.org/), Wilson and Reeder's Mammal Species of the World, 3rd edition (MSW3, http://www.departments.bucknell.edu/biology/resources/msw3/), and Patton et al. (2015). In total, our final dataset contained 11,495 primary records (Fig. 1).

*Survey completeness*

To assess sampling biases and gaps we used survey completeness analysis, that helps identify well-sampled sites at a given spatial resolution and extension (Lobo et al., 2018). We calculated the completeness for cells with 1 km$^2$ spatial resolution within the boundaries of the Atlantic Forest (Muylaert et al., 2018a). We chose this spatial resolution because landscape structure and configuration have been shaped by changes in land-use that occur and affect ecological processes at smaller spatial scales (Benítez-López et al., 2010; Sobral-Souza et al., 2021a).

To account for the effect of the minimum number of records and cell-size resolution on survey completeness (Tessarolo et al., 2021; Bosco et al., 2022), we performed a sensitivity analysis with different thresholds (minimum number of records) for a cell to be considered well-sampled (see Fig. 2). We set the minimum threshold at 20 occurrences for survey completeness calculation, which represented a balance between the risk of high completeness due to low occurrences, and sufficient data representation (Fig. 2).

Then, for cells with more than 20 occurrences, we calculated survey completeness as the ratio between observed and predicted numbers of species. The predicted number of species for each 1 km$^2$ cell was derived from a species accumulation curve generated using the *Clench* function of the KnowBr package (Lobo et al., 2018) implemented in R 4.3.1 (R Development Core Team, 2021). Survey completeness ranges from 0 to 1, with 0 being no survey and
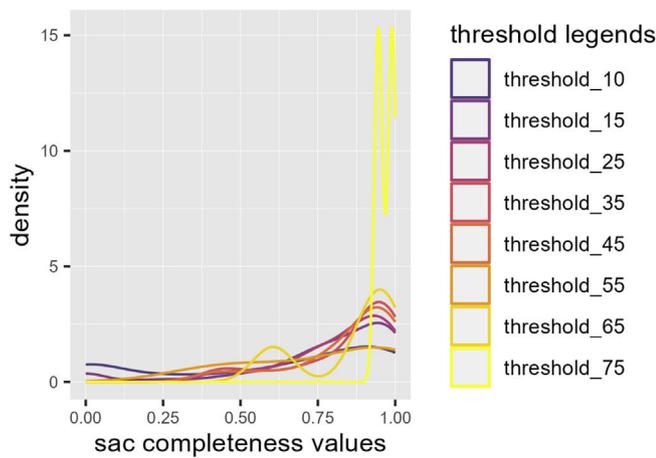
**Fig. 2.** Sensitivity analysis with different thresholds to set a minimum number of records for a cell to be considered well-sampled.

1 being maximum survey completeness. We considered as well-sampled sites those with survey completeness equal to or greater than 0.7 (see Hortal and Lobo, 2005; Lobo et al., 2018; Sobral et al., 2021a; Bosco et al., 2022). We calculated how many individuals can be captured in a $1 \, km^2$ cell of the Atlantic Forest filtering the sampling effort for an area of $1 \, km^2$ and calculating the average number of individuals captured per locality in the Atlantic Forest database. We found that 20 records represent 10% of the number of individuals sampled in our dataset.

*Spatial bias*

We assessed spatial bias of the well-sampled sites based on accessibility and landscape metrics. For accessibility, we used the Euclidean distance of the sampling site to the nearest road (m) and to the nearest urban center (m). To obtain the distance to roads, we used the shapefile of federal and state paved roads of the National Department of Transport Infrastructure (*Departamento Nacional de Infraestrutura de Transportes - DNIT*). For urban centers, we used the shapefiles of urban areas of the Brazilian Foundation for Sustainable Development (*Fundação Brasileira para o Desenvolvimento Sustentável - FBDS)*, the Brazilian Institute of Geography and Statistics (*Instituto de Geografia e Estatística - IBGE*), and the SOS Atlantic Forest Foundation *(Fundação SOS Mata Atlântica).*

We also used fragment size (ha) and percentage of forest cover (%) to assess how landscape configuration and composition can influence spatial biases and gaps in small rodent sampling. For this we used the shapefiles of landcover of the *Fundação Brasileira para o Desenvolvimento Sustentável* (FBDS), *Fundação SOS Mata Atlântica*, *Instituto Nacional de Pesquisas Espaciais* (INPE), and the University of Maryland (Global Forest Change Project, Hansen et al. 2013.) Fragment size was calculated as the sum of the forest area (in hectare) of each cell. To infer forest cover (%) for each cell, the proportion of forest within a square window of 30 m, centered in the focal cell, was calculated (=amount of vegetation cells/total number of cells in the window). Thus, forest cover (%) varied between 0 and 100%. All metrics were calculated at a resolution of 30 m × 30 m and adjusted for $1 \, km^2$. Detailed descriptions of landscape data can be found in Sobral-Souza et al. (2021a,2021b).

We built a landscape rarity map using the methods proposed by Sobral-Souza et al. (2021a). Briefly, we first performed a PCA analysis to synthetize forest percentage cover, fragment size and functional connectivity into a 2D multivariate environment axis. Then, we inferred the frequency of each PCA axis condition in an environmental space defined by 0.25 × 0.25 km grid-cells. We standardized the PCA axis (z score standardization) to vary between 0

and 1, where values near 1 represent rare landscape environmental conditions that occur in a unique or few grid cells or sites. In contrast, values close to zero indicate common landscape conditions that occur in many sites of the Atlantic Forest. To analyze whether well-sampled sites provided a representative subset of the overall landscape variation and encompassed regions of rare landscape conditions (Sobral-Souza et al., 2021b), we compared the frequencies of rarity landscape conditions of all Atlantic Forest and well-sampled cells using Komolgorov-Smirnov test. We also tested accessibility bias towards the well-sampled sites. We compared the frequency of the distance to the nearest road for the entire Atlantic Forest extension with the frequency of distance to nearest road for the well-sampled sites based on Komolgorov-Smirnov test. We used the same procedure to compare the frequency distribution of distance to nearest urban center for the entire Atlantic Forest with that of the well-sampled sites. For the landscape conditions we compared the frequency distribution of fragment size for the entire Atlantic Forest with the frequency distribution of fragment size of the well-sampled sites. We used the same procedure for forest cover (%), comparing the frequency distribution of forest percentage of the entire Atlantic Forest with that of the well-sampled sites.

Finally, we used Komolgorov-Smirnov test to infer the difference between the frequency of the entire Atlantic Forest landscape conditions and the landscape conditions of well-sampled sites, and then for each of Atlantic Forest biogeographic regions proposed by Silva and Casteleti (2003). All analyses were performed in R 4.3.1 (R Development Core Team, 2021).

**Result**

Only 507 cells ($1 \, km^2$) of Atlantic Forest (out of 1,649,932 cells) were considered well-sampled. Thus, only 0.03% of the spatial surface of the Atlantic Forest is well-sampled for small rodents (Fig. 3). Serra do Mar was the best-known biogeographic subregion (Fig. 4) with 177 well-sampled sites (0.15% of the subregion area), followed by Pernambuco and Bahia subregions with between 0.07 and 0.09% of their areas well-sampled. Araucaria, Florestas do Interior and Diamantina subregions had less than 0.05% of their areas as well-sampled. São Francisco subregion had no sites with high completeness (Fig. 4).

We found a sampling bias towards sites closer to paved roads (D = 0.1234, $p = 3.948^{-07}$) and urban centers (D = 0.10877, $p = 1.236^{-05}$). Well-sampled sites were mainly at distances lower than 15 km from cities and paved roads (Fig. 5). Following our predictions, landscape structure also influenced sampling bias (Fig. 5,) with well-sampled sites biased towards landscapes with larger fragments (D = 0.28178, $p < 2.2^{-16}$), and with higher percentages of forest cover (D = 0.41727, p-value < $2.2^{-16}$). Although Atlantic Forest's landscapes had less than 50% of forest cover, most well-sampled sites were at landscapes with >50% (Fig. 5). The well-sampled sites were also biased to rare landscape conditions (D = 0.11257; p = 0.008), showing gaps of sampling of smalls rodents in common landscape conditions (Fig. 5).

**Discussion**

Our findings suggest a spatial bias in the sampling sites of small rodents in the Atlantic Forest. Well-sampled sites were closer to urban centers and roads and involved larger fragments and landscapes with greater forest cover. Additionally, our results also highlight bias towards rare landscape conditions of Atlantic Forest. In turn, our findings also indicate a sampling gap for small fragments and landscapes with low forest cover, sites further from urban areas and roads, and in common landscape conditions, pre-
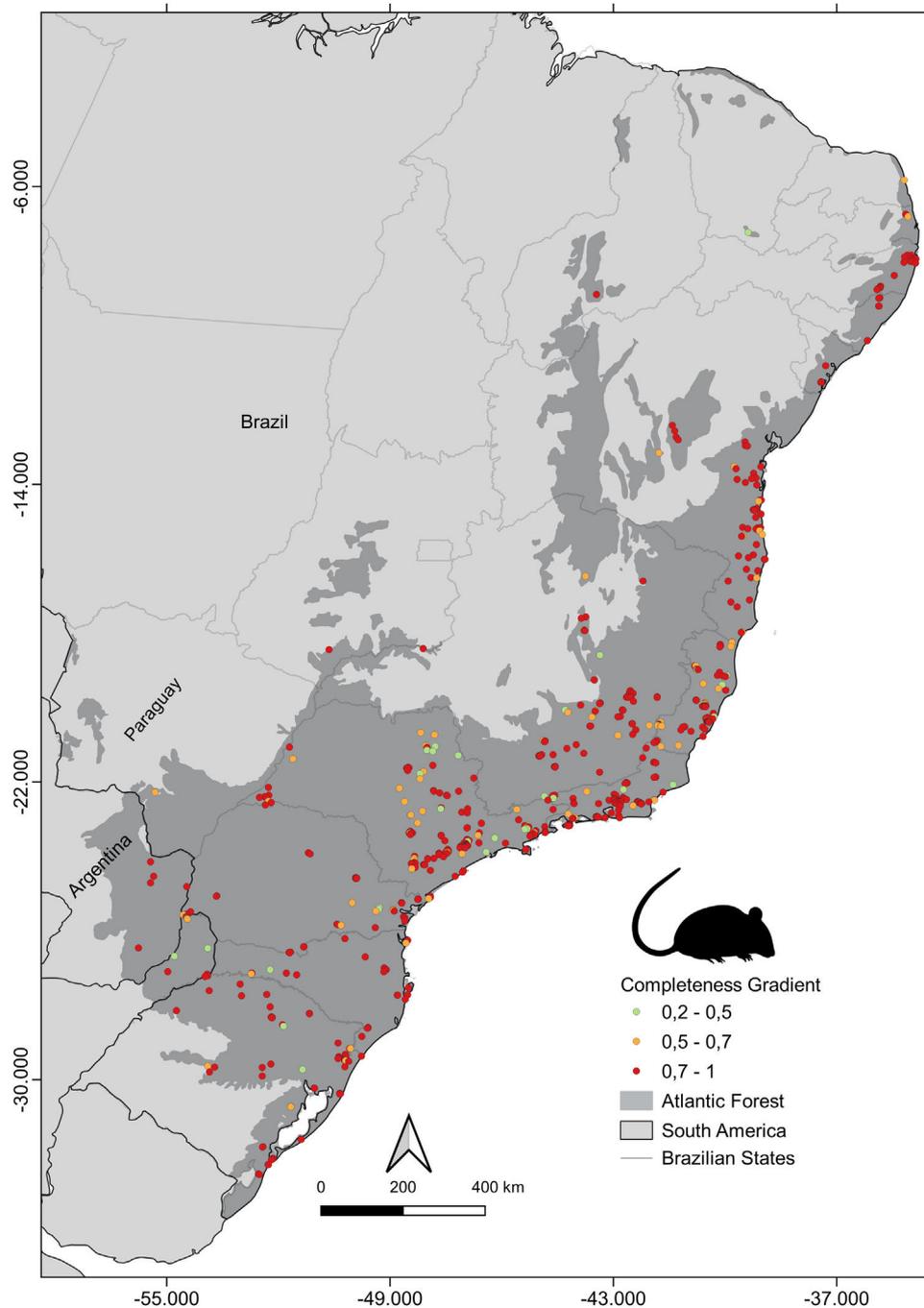
**Fig. 3.** Well-sampled sites (completeness => 0.7) and poorly sampled sites (completeness < 0.7) of small rodent inventories in the Atlantic Forest (shapefile from Muylaert et al., 2018b). Red dots indicate well-sampled sites (n = 507). Green and yellow dots indicate sites where less than 69% of the biodiversity is known (n = 41 and 91, respectively).

cluding broad scale inferences of the relationships between habitat fragmentation and forest loss and small rodents' biodiversity. In addition, based on digitally accessible data, the spatial aggregation of sampling efforts generated a 99% of spatial gap in the knowledge of small rodents in the Atlantic Forest, with well-sampled sites covering only 0.03% of the Atlantic Forest's extension. Our results demonstrate that broad-scale knowledge of small rodents is spatially limited, which impairs inferences about the factors that drive and change their spatial distribution, a key factor for design sound conservation actions.

Although accessible digital platforms aim to provide more data on biodiversity to increase biodiversity knowledge (Devictor and Bensaude-Vincent, 2016), and to help fill knowledge gaps, the data

currently available is still limited (Hortal et al., 2007). The limitations in biodiversity information are mainly due to the different methods used for data sampling, since the data are derived from different researchers and with different purposes (Wüest et al., 2020).

Our findings also suggest a spatial bias of well-sampled sites, which were associated with more accessible regions. Proximity to access routes and urban centers are well-known causes of bias in biodiversity knowledge for different taxonomic groups (Almeida et al., 2021; Correia et al., 2019; Oliveira et al., 2016). The historic exploitation of natural resources of the Atlantic Forest caused forest loss, remaining now only 28% of the original forest (Rezende et al., 2018). Changes in the Atlantic Forest landscape resulted in frag-
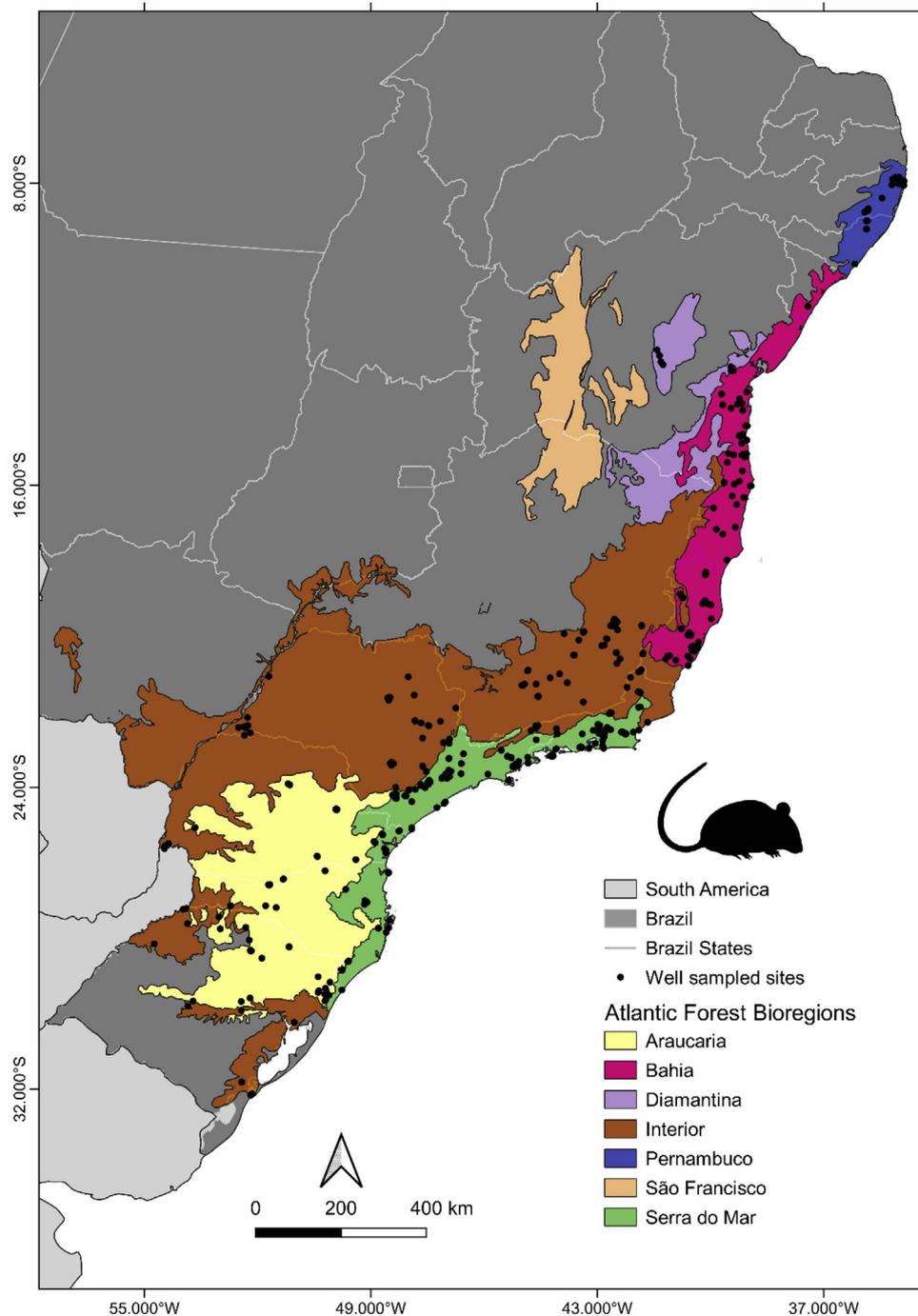
**Fig. 4.** Small rodents well-sampled sites in the Brazilian Atlantic Forest biogeographic subregions proposed by Silva and Casteleti (2003).

mented landscapes with high density of access routes and cities (Lapola et al., 2014; Oliveira et al., 2016), including protected areas that are easily accessible by roads. In the southeast r Atlantic Forest, the protected areas are spatially connected, comprising the largest and most conserved forest remnants, which are easily accessible. This spatial pattern may lead to a broad-scale spatial biases due to both accessibility and landscape structure. In addition, historically, funding of biological research in the Atlantic Forest has favored projects with long-term sampling and on local scale (e.g., Long-Term Ecological Projects, PELDs in Portuguese). These projects were able to clarify and describe diverse ecological and biodiversity patterns on a local scale, while failing, on the other hand, to clarify patterns on a broader scale, like as responses of biodiversity to

landscape modification and fragmentation on a macroecological perspectives. Consequently, we found a sampling gap at distances greater than 15 km from urban centers and access routes, which should be priority sites for future sampling efforts. These spatial biases in survey sampling were also found for other taxonomic groups and ecoregions. For instance, in Brazilian Amazon, accessibility and proximity to research facilities explained sampling probability of an area for different taxa (Carvalho et al., 2023).

Sampling close to roads and cities may lead to biodiversity indices underestimation, mainly because these sites may not represent regional biodiversity (Benítez-López et al., 2010; Laurance, 2009) and may be affected by biological invasions (Hobbs et al., 2009). Access routes facilitate habitat degradation (Freitas et al.,
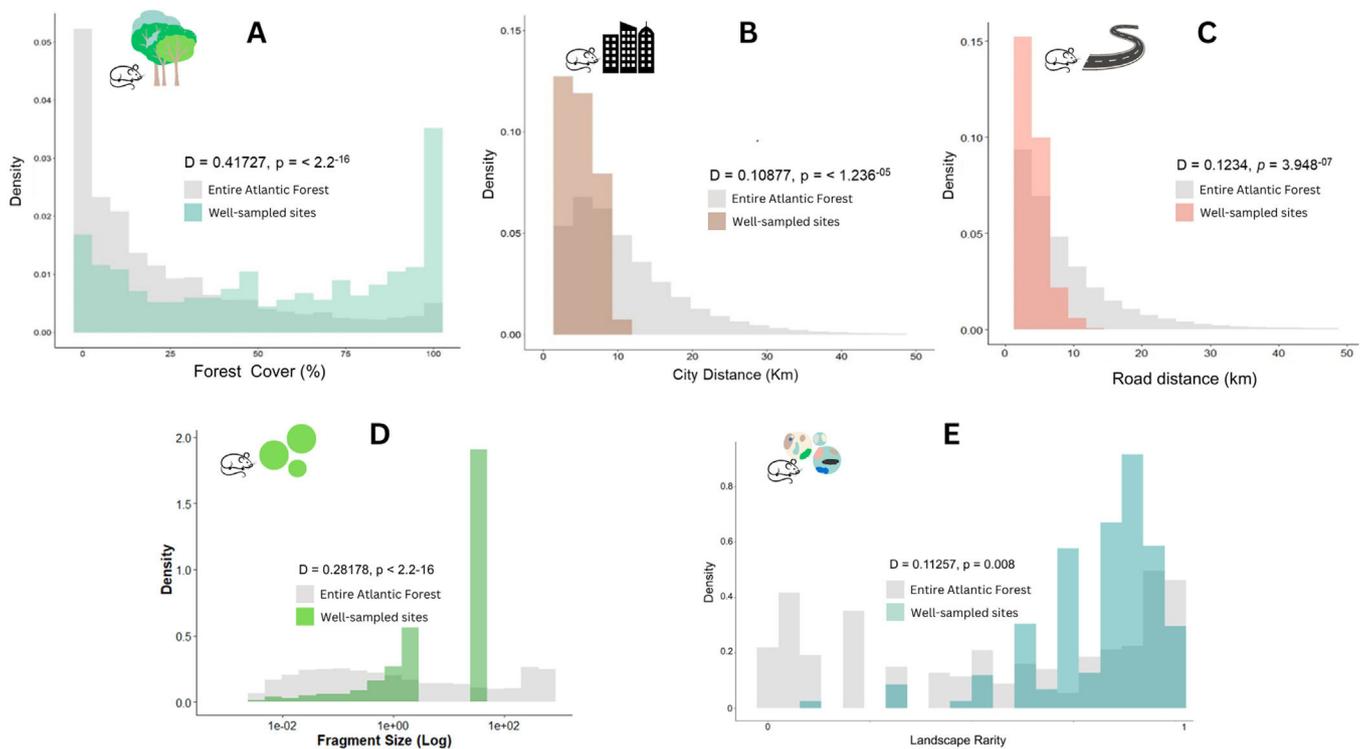
**Fig. 5.** Spatial bias of sites well-sampled for small rodents related to the accessibility and landscape. The well-sampled sites are biased towards locations closer to larger fragments, higher percentage of habitat cover, closed to roads and cities, and rare landscape conditions.

2010), and change landscape configuration, which in turn alter small rodent species composition, due to edge effect (Rosa et al., 2018). On the other hand, sampling sites close to large cities and roads offer better infrastructure, more qualified human resources and reduced costs for researchers with little financial support (Lessa et al., 2019; Meyer et al., 2015).

We also found bias for larger fragments and landscapes with greater forest cover. These biases may be explained by the limited amount of time and funding for sampling, leading the researchers to choose more accessible larger fragments where the probability to record a greater number of species is higher (Reddy and Dávalos, 2003; Sastre and Lobo, 2009; Sobral-Souza et al., 2021a). Additionally, in the Atlantic Forest, larger fragments are temporally more stable and experience less effects from human-induced landscape modifications (Hansen et al., 2020). Human colonization and habitat fragmentation in Atlantic Forest resulted in a spatial configuration with the main forest remnants on sites close to high human density areas (Ribeiro et al., 2009). Thus, the largest forest fragments are spatially connected, which explains the spatial aggregation pattern of well-sampled sites. The main forest remnants of the Atlantic Forest are also in protected areas, which can foster long term habitat stability and provide sites for long-term temporal samplings with little human efforts. The same sampling biases towards larger forest fragments, closer to cities and roads was found for fruit-feeding butterfly in Atlantic Forest, (Sobral-Souza et al., 2021a) and for herptile, in global perspective (Nori et al., 2023), reassuring the gap knowledge for small and disconnected forest fragments.

The Serra do Mar range is the most conserved biogeographic region of the Atlantic Forest (Ribeiro et al., 2009) with 35% of the ecoregion's well-sampled sites for small rodents (Fig. 4). It contains the largest protected forest remnants of Atlantic Forest, however, embedded in large urban centers, such as São Paulo city. Although the Floresta do Interior biogeographic region is six times larger than the Serra do Mar, our results revealed that knowledge of

its biodiversity is smaller, with 0.02% of its surface well-sampled for small rodents. Sampling sites in the Floresta do Interior are spatially closer to the Serra do Mar region, probably due to the presence of large urban centers and research institutes in the former region (Almeida et al., 2021; Candelária et al., 2021; Lessa et al., 2019). The biogeographic regions at the northeast Atlantic Forest are practically unknown, especially São Francisco, which has no well-sampled sites. The northeast biogeographic regions were classified based on the endemism level, which usually attracts taxonomists to discover new species (Silva-Soares et al., 2021). Even so, areas with high endemism are poorly sampled, such as Diamantina, Bahia and Pernambuco. This can be explained by the low number of universities and protected areas in these regions. The northeast is also an ecotone between the Caatinga and the Cerrado ecoregions, which can hinder the sampling of Atlantic Forest's taxa. The sampling gap at the biogeographic regions of the northeast Atlantic Forest may also explain the sampling gap at ecotone landscape conditions. São Francisco, Diamantina and Pernambuco, which showed no well-sampled sites occur in a semi-arid climate dominated by seasonally dry forests (Pennington et al., 2000). Thus, our findings point to the critical need to prioritize these regions in future inventories.

Biased biodiversity data also hamper the design of species distribution maps, because they will show the sampling bias and not the true species distribution (Rocchini et al., 2011; Ladle and Hortal, 2013). Landscape configuration also affects species dispersal (Fahrig, 2005); thus, landscape sampling bias may impair inferences about the effects of landscape modifications on local and regional biodiversity (Sobral-Souza et al., 2021b; Santos et al., 2020). The Wallacean Shortfall is dependent on the spatial scale on which it is assessed. At low spatial resolutions for broad spatial scales, biodiversity knowledge may be considered complete as all grid cells would have enough occurrences to reach a high level of completeness (Lobo et al., 2018). On the other hand, at high spatial resolutions, the spatial gaps in biodiversity knowledge can be

found, such as those we found here for small rodents of the Atlantic Forest.

In conclusion, knowledge of biodiversity of small rodents in the Atlantic Forest is currently insufficient to understand how landscape modifications affect the spatial distributions of species at a large scale. The bias highlighted here showed biodiversity knowledge gaps for small fragments, and landscapes with small forest cover and far from roads and cities. Most Atlantic Forest remnants are small fragments, with small forest cover being common landscape conditions, as they are composed of secondary forests or forest restoration (Ribeiro et al., 2009). Therefore, to contribute to increasing survey completeness across the range of landscape configuration and conditions, samplings of small rodents should prioritize inaccessible sites, away from cities and roads, and landscapes with small fragments and with small forest cover, the most common landscape conditions in Atlantic Forest.

## Funding

## Author contributions

L.P.C; V.M.G.L T.S-S and D.J.R conceived the research.
N.S.B, T.S-S and R.G.C performed the analysis.
TS-S; L.P.C; N.S.B and R.G.C wrote the manuscript draft.
All authors contributed to the final version of the manuscript.

## Competing interests

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. There is no conflict of interest.

## Acknowledgements

## Appendix A.  Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.pecon.2024.07.004.

## References

Almeida, T.C.de, Tessarolo, G., Nabout, J.C., Teresa, F.B., 2021. Non-stationary drivers on fish sampling efforts in Brazilian freshwaters. Divers. Distrib. 27, 1224–1234, http://dx.doi.org/10.1111/ddi.13269.

Beck, J., Böller, M., Erhardt, A., Schwanghart, W., 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. Ecol. Inform. 19, 10–15, http://dx.doi.org/10.1016/j.ecoinf.2013.11.002.

Benítez-López, A., Alkemade, R., Verweij, P.A., 2010. The impacts of roads and other infrastructure on mammal and bird populations: a meta-analysis. Biol. Conserv. 143, 1307–1316, http://dx.doi.org/10.1016/j.biocon.2010.02.009.

Bergallo, H.G., Luz, J.L., Raíces, D.S., Hatano, F.H., Martins-Hatano, F., 2005. Habitat use by Oryzomys subflavus (Rodentia) in an open shrubland formation in Restinga de Jurubatiba National Park, RJ, Brazil. Braz. J. Biol. 65, 583–588, http://dx.doi.org/10.1590/s1519-69842005000400004.

Boakes, E.H., McGowan, P.J.K., Fuller, R.A., Chang-Qing, D., Clark, N.E., O'Connor, K., Mace, G.M., 2010. Distorted views of biodiversity: Spatial and temporal bias in species occurrence data. PLoS Biol. 8, http://dx.doi.org/10.1371/journal.pbio.1000385.

Bonvicino, C.R., Oliveira, J.Ade, D'Andrea, O.S., 2008. Guia dos roedores do Brasil, com chaves para gêneros baseadas em caracteres externos. Centro Pan-Americano de Febre Aftosa-OPAS/OMS, Rio de Janeiro.

Bosco, N.S., Prasniewski, V.M., Santos, J.P., Silveira, N.S., Culot, L., Ribeiro, M.C., Tessarolo, G., Sobral-Souza, T., 2022. Scale affects the understanding of biases on the spatial knowledge of Atlantic Forest primates. PECON 20, 338–345, http://dx.doi.org/10.1016/j.pecon.2022.08.002.

Bovendorp, R.S., Villar, N., de Abreu-Junior, E.F., Bello, C., Regolin, A.L., Percequillo, A.R., Galetti, M., 2017. Atlantic small-mammal: a dataset of communities of rodents and marsupials of the Atlantic forests of South America. Ecology 98, 2226, http://dx.doi.org/10.1002/ecy.1893.

Brooks, T.M., Da Fonseca, G.A.B., Rodrigues, A.S.L., 2004. Protected areas and species. Conserv. Biol. 18, 616–618, http://dx.doi.org/10.1111/j.1523-1739.2004.01836.x.

Candelária, L.P., Zucchetto, M., Gonçalves, S.R.A., GDSF, Braga, Izzo, T.J., Teixido, A.L., 2021. Biogeographic and fragmentation-related research biases on antbirds and non-flying small mammals in Brazil. J. Trop. Ecol. 37, 175–184, http://dx.doi.org/10.1017/S0266467421000250.

Carvalho, C.S., Martello, F., Galetti, M., Pinto, F., Francisco, M.R., Silveira, L.F., Galetti, P.M., 2021. Environmental heterogeneity and sampling relevance areas in an Atlantic forest endemism region. Persp. Ecol. Conser. 19, 3, http://dx.doi.org/10.1016/j.pecon.2021.05.001.

Carvalho, R.L., et al., 2023. Pervasive gaps in Amazonian ecological research. Curr. Biol. 33, 3544.

Ceballos, G., Ehrlich, P.R., 2002. Mammal population losses and the extinction crisis. Science 296, 904–907, http://dx.doi.org/10.1126/science.1069349.

Chase, J.M., Blowes, S.A., Knight, T.M., Gerstner, K., May, F., 2020. Ecosystem decay exacerbates biodiversity loss with habitat loss. Nature 584, 238–243, http://dx.doi.org/10.1038/s41586-020-2531-2.

Collen, B., Ram, M., Zamin, T., McRae, L., 2008. The tropical biodiversity data gap: addressing disparity in global monitoring. Trop. Conserv. Sci. 1, 75–88, http://dx.doi.org/10.1177/194008290800100202.

Correia, R.A., Ruete, A., Stropp, J., Malhado, A.C.M., dos Santos, J.W., Lessa, T., Alves, J.A., Ladle, R.J., 2019. Using ignorance scores to explore biodiversity recording effort for multiple taxa in the Caatinga. Ecol. Indic. 106, 105539, http://dx.doi.org/10.1016/j.ecolind.2019.105539.

Delciellos, A.C., Vieira, M.V., Grelle, C.E.V., Cobra, P., Cerqueira, R., 2016. Habitat quality versus spatial variables as determinants of small mammal assemblages in Atlantic Forest fragments. J. Mammal. 97, 253–265, http://dx.doi.org/10.1093/jmammal/gyv175.

Devictor, V., Bensaude-Vincent, B., 2016. From ecological records to big data: the invention of global biodiversity. Hist. Philos. Life Sci. 38, 1–23, http://dx.doi.org/10.1007/s40656-016-0113-2.

Dirzo, R., Raven, P.H., 2003. Global state of biodiversity and loss. Annu. Rev. Environ. Resour. 28, 137–167, http://dx.doi.org/10.1146/annurev.energy.28.050302.105532.

Dirzo, R., Young, H.S., Galetti, M., Ceballos, G., Isaac, N.J.B., Collen, B., 2014. Defaunation in the Anthropocene. Science 345, 401–406, http://dx.doi.org/10.1126/science.1251817.

Fahrig, L., 2005. When is a landscape perspective important? In: Wiens, J., Moss, M. (Eds.), Issues and Perspectives in Landscape Ecology (Cambridge Studies in Landscape Ecology). Cambridge University Press, Cambridge, pp. 3–10, http://dx.doi.org/10.1017/CBO9780511614415.002.

Figueiredo, M.S.L., Barros, C.S., Delciellos, A.C., Guerra, E.B., Cordeiro-Estrela, P., Kajin, M., Alvarez, M.R., Asfora, P.H., Astúa, D., Bergallo, H.G., Cerqueira, R., Geise, L., Gentile, R., Grelle, C.E.V., Iack-Ximenes, G.E., Oliveira, L.C., Weksler, M., Vieira, M.V., 2017. Abundance of small mammals in the Atlantic Forest (ASMAF): a data set for analyzing tropical community patterns. Ecology 98, http://dx.doi.org/10.1002/ecy.2005, 2981–2981.

Freitas, S.R., Hawbaker, T.J., Metzger, J.P., 2010. Effects of roads, topography, and land use on forest cover dynamics in the Brazilian Atlantic Forest. For. Ecol. Manage. 259, 410–417, http://dx.doi.org/10.1016/j.foreco.2009.10.036.

Galindo-Leal, C., Câmara, I.G., 2003. The Atlantic Forest of South America: biodiversity status, threats, and outlook. In: Center for Applied Biodiversity Science at Conservation International State of the Hotspots series. Island Press, Washington.

Girardello, M., Martellos, S., Pardo, A., Bertolino, S., 2018. Gaps in biodiversity occurrence information may hamper the achievement of international biodiversity targets: insights from a cross-taxon analysis. Environ. Conserv. 45, 370–377, http://dx.doi.org/10.1017/S0376892918000115.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. Sci 342, 850–853.

Hansen, M.C., Wang, L., Song, X.P., Tyukavina, A., Turubanova, S., Potapov, P.V., Stehman, S.V., 2020. The fate of tropical forest fragments. Sci. Adv. 6, 11, http://dx.doi.org/10.1126/sciadv.aax8574.

Hanski, I., 2011. Habitat loss, the dynamics of biodiversity, and a perspective on conservation. Ambio 40, 248–255, http://dx.doi.org/10.1007/s13280-011-0147-3.

Hobbs, R.J., Higgs, E., Harris, J.A., 2009. Novel ecosystems: implications for conservation and restoration. Trends Ecol. Evol. 24, 599–605, http://dx.doi.org/10.1016/j.tree.2009.05.012.

Hortal, J., Lobo, J.M., 2005. An ED-based protocol for the optimal sampling of biodiversity. Biodiversity and Conservation 14, 2913–2947, http://dx.doi.org/10.1007/s10531-004-0224-z.

Hortal, J., Lobo, J.M., Jiménez-Valverde, A., 2007. Limitations of biodiversity databases: case study on seed-plant diversity in Tenerife, Canary Islands. Conserv. Biol. 21, 853–863, http://dx.doi.org/10.1111/j.1523-1739.2007.00686.x.

Hortal, J., Jiménez-Valverde, A., Gómez, J.F., Lobo, J.M., Baselga, A., 2008. Historical bias in biodiversity inventories affects the observed environmental niche of the species. Oikos 117, 847–858, http://dx.doi.org/10.1111/j.0030-1299.2008.16434.x.

Hortal, J., De Bello, F., Diniz-Filho, J.A.F., Lewinsohn, T.M., Lobo, J.M., Ladle, R.J., 2015. Seven shortfalls that beset large-scale knowledge of biodiversity. Annu. Rev. Ecol. Evol. Syst. 46, 523–549, http://dx.doi.org/10.1146/annurev-ecolsys-112414-054400.

Ladle, R., Hortal, J., 2013. Mapping species distributions: living with uncertainty. Front. Biogeogr. 5, 4–6, http://dx.doi.org/10.21425/F5FBG12942.

Lapola, D.M., Martinelli, L.A., Peres, C.A., Ometto, J.P.H.B., Ferreira, M.E., Nobre, C.A., Aguiar, A.P.D., Bustamante, M.M.C., Cardoso, M.F., Costa, M.H., Joly, C.A., Leite, C.C., Moutinho, P., Sampaio, G., Strassburg, B.B.N., Vieira, I.C.G., 2014. Pervasive transition of the Brazilian land-use system. Nat. Clim. Chang. 4, 27–35, http://dx.doi.org/10.1038/nclimate2056.

Laurance, W., 2009. Conserving the hottest of the hotspots. Biol. Conserv. 142, http://dx.doi.org/10.1016/j.biocon.2008.10.011.

Lessa, T., Dos Santos, J.W., Correia, R.A., Ladle, R.J., Malhado, A.C.M., 2019. Known unknowns: filling the gaps in scientific knowledge production in the Caatinga. PLoS One 14, 1–12, http://dx.doi.org/10.1371/journal.pone.0219359.

Lobo, J.M., Hortal, J., Yela, J.L., Millán, A., Sánchez-Fernández, D., García-Roselló, E., González-Dacosta, J., Heine, J., González-Vilas, L., Guisande, C., 2018. KnowBR: an application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. Ecol. Indic. 91, 241–248, http://dx.doi.org/10.1016/j.ecolind.2018.03.077.

Meyer, C., Kreft, H., Guralnick, R., Jetz, W., 2015. Global priorities for an effective information basis of biodiversity distributions. Nat. Commun. 6, http://dx.doi.org/10.1038/ncomms9221.

Muylaert, R.L., Vancine, M.H., Bernardo, R., Oshima, J.E.F., Sobral-Souza, T., Tonetti, V.R., Niebuhr, B.B., . . . Ribeiro, M.C., 2018a. Uma nota sobre os limites territoriais da Mata Atlântica. Oecologia Aust. 22, 302–311, http://dx.doi.org/10.4257/oeco.2018.2203.09.

Muylaert, R.L., Vancine, M.H., Bernardo, R., Oshima, J.E.F., Sobral- Souza, T., Tonetti, V.R., . . . Ribeiro, M.C., 2018b. Uma nota sobre os limites territoriais da Mata Atlântica. Oecol. Aust 22 (3), 302–311.

Myers, N., Mittermeler, R.A., Mittermeler, C.G., Da Fonseca, G.A.B., Kent, J., 2000. Biodiversity hotspots for conservation priorities. Nature 403, 853–858, http://dx.doi.org/10.1038/35002501.

Nori, J., Cordier, M., Osorio-Olvera, L., Hortal, J., 2023. Global knowledge gaps of herptile responses to land transformation. Front. Ecol. Environ. 21 (9), 411–417, http://dx.doi.org/10.1002/fee.2625.

Oliveira, U., Paglia, A.P., Brescovit, A.D., de Carvalho, C.J.B., Silva, D.P., Rezende, D.T., Leite, F.S.F., Batista, J.A.N., Barbosa, J.P.P.P., Stehmann, J.R., Ascher, J.S., de Vasconcelos, M.F., De Marco, P., Löwenberg-Neto, P., Dias, P.G., Ferro, V.G., Santos, A.J., 2016. The strong influence of collection bias on biodiversity knowledge shortfalls of Brazilian terrestrial biodiversity. Divers. Distrib. 22, 1232–1244, http://dx.doi.org/10.1111/ddi.12489.

Pardini, R., 2004. Effects of forest fragmentation on small mammals in an Atlantic Forest landscape. Biodiv. Conser. 13, 2567–2586, http://dx.doi.org/10.1023/B:BIOC.0000048452.18878.2d.

Patton, J.L., Pardiñas, U.F.J., D'Elía, G., 2015. Mammals of South America; Volume 2: Rodents. University of Chicago Press.

Pennington, T.R., Prado, D.E., Pendry, C.A., 2000. Neotropical seasonally dry forests and Quaternary vegetation changes. J. Biogeogr. 27, 261–273, http://dx.doi.org/10.1046/j.1365-2699.2000.00397.x.

Püttker, T., Pardini, R., Meyer-Lucht, Y., et al., 2008. Responses of five small mammal species to micro-scale variations in vegetation structure in secondary Atlantic Forest remnants, Brazil. BMC Ecol. 8, 9, http://dx.doi.org/10.1186/1472-6785-8-9.

R Development Core Team, https://www.Rproject.org, 2021.

Reddy, S., Dávalos, L.M., 2003. Geographical sampling bias and its implications for conservation priorities in Africa. J. Biogeogr. 30, 1719–1727, http://dx.doi.org/10.1046/j.1365-2699.2003.00946.x.

Rezende, C.L., Scarano, F.R., Assad, E.D., Joly, C.A., Metzger, J.P., Strassburg, B.B.N., Tabarelli, M., Fonseca, G.A., Mittermeier, R.A., 2018. From hotspot to hopespot: an opportunity for the Brazilian Atlantic Forest. Perspect. Ecol. Conserv. 16, 208–214, http://dx.doi.org/10.1016/j.pecon.2018.10.002.

Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J., Hirota, M.M., 2009. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. Biol. Conserv. 142, 1141–1153, http://dx.doi.org/10.1016/j.biocon.2009.02.021.

Rocchini, D., Hortal, J., Lengyel, S., Lobo, J.M., Jiménez-Valverde, A., Ricotta, C., Bacaro, G., Chiarucci, A., 2011. Accounting for uncertainty when mapping species distributions: the need for maps of ignorance. Prog. Phys. Geogr. 35, 211–226, http://dx.doi.org/10.1177/0309133311399491.

Rosa, C.A.da, Secco, H., Carvalho, N., Alves, C., Maia, A.C., Bager, A., Catarina, S., 2018. Edge effects on small mammals: differences between arboreal and ground-dwelling species living near roads in Brazilian fragmented landscapes. Austral Ecol. 43, 117–126, http://dx.doi.org/10.1111/aec.12549.

Rosa, C.A.da, Ribeiro, B.R., Bejarano, V., et al., 2020. NEOTROPICAL ALIEN MAMMALS: a data set of occurrence and abundance of alien mammals in the neotropics. Ecology 101, http://dx.doi.org/10.1002/ecy.3115.

Santos, J.P., Sobral-Souza, T., Brown, K.S., Vancine, M.H., Ribeiro, M.C., Freitas, A.V.L., 2020. Effects of landscape modification on species richness patterns of fruit-feeding butterflies in Brazilian Atlantic Forest. Divers Distrib. 26, 196–208, http://dx.doi.org/10.1111/ddi.13007.

Sastre, P., Lobo, J.M., 2009. Taxonomist survey biases and the unveiling of biodiversity patterns. Biol. Conserv. 142, 462–467, http://dx.doi.org/10.1016/j.biocon.2008.11.002.

Silva, J.M.C., Casteleti, C.H.M., 2003. Status of the biodiversity of the Atlantic Forest of Brazil. In: Galindo-Leal, C., Câmara, I.G. (Eds.), The Atlantic Forest of South America: Biodiversity Status; Threats; and Outlook. CABS and Island Press, pp. 43–59.

Silva-Soares, T., Ferreira, R.B., Ornellas, I.A., Zocca, C.Z., Caramaschi, U., Cruz, C.A.G., 2021. A New Species of Ischnocnema (Anura: Brachycephalidae) from the Mountainous Region of Atlantic Forest, Southeastern Brazil, With a New Phylogeny and Diagnose for Ischnocnema parva Series. Zootaxa., http://dx.doi.org/10.11646/ZOOTAXA.5082.3.1.

Sobral-Souza, T., Stropp, J., Santos, J.P., Prasniewski, V.M., Szinwelski, N., Vilela, B., Freitas, A.V.L., Ribeiro, M.C., Hortal, J., 2021a. Knowledge gaps hamper understanding the relationship between fragmentation and biodiversity loss: the case of Atlantic Forest fruit-feeding butterflies. PeerJ, 1–23, http://dx.doi.org/10.7717/peerj.11673.

Sobral-Souza, T., Santos, J.P., Maldaner, M.E., Lima-Ribeiro, M.S., Ribeiro, M.C., 2021b. EcoLand: a multiscale niche modelling framework to improve predictions on biodiversity and conservation. Perspect. Ecol. Conserv. 19, 362–368, http://dx.doi.org/10.1016/j.pecon.2021.03.008.

Sousa-Baena, M.S., Garcia, L.C., Peterson, A.T., 2014. Knowledge behind conservation status decisions: data basis for "data deficient" Brazilian plant species. Biol. Conserv. 173, 80–89, http://dx.doi.org/10.1016/j.biocon.2013.06.034.

Tabarelli, M., Aguiar, A.V., Ribeiro, M.C., Metzger, J.P., Peres, C.A., 2010. Prospects for biodiversity conservation in the Atlantic Forest: lessons from aging human-modified landscapes. Biol. Conserv. 143, 2328–2340, http://dx.doi.org/10.1016/j.biocon.2010.02.005.

Teixido, A.L., Gonçalves, S.R.A., Fernández-Arellano, G.J., Dáttilo, W., Izzo, T.J., Layme, V.M.G., Moreira, L.F.B., Quintanilla, L.G., 2020. Major biases and knowledge gaps on fragmentation research in Brazil: implications for conservation. Biol. Conserv. 251, http://dx.doi.org/10.1016/j.biocon.2020.108749.

Tessarolo, G., Ladle, R.J., Lobo, J.M., Rangel, T.F., Hortal, J., 2021. Using maps of biogeographical ignorance to reveal the uncertainty in distributional data hidden in species distribution models. Ecography 44, 1743–1755, http://dx.doi.org/10.1111/ecog.05793.

Vancine, M.H., Muylaert, R.L., Niebuhr, B.B., Oshima, J.E.F., Tonetti, V., Bernardo, R., De Angelo, C., Rosa, M.R., Grohmann, Ribeiro M.C., 2024. The Atlantic Forest of South America: spatiotemporal dynamics of the vegetation and implications for conservation. Biol. Conserv. 291, 110499, http://dx.doi.org/10.1016/j.biocon.2024.110499.

Wilson, D.E., Reeder, D.M., 2005. Mammal Species of the World: A Taxonomic and Geographic Reference. Johns Hopkins University Press, Baltimore and London.

Wüest, R.O., Zimmermann, N.E., Zurell, D., Alexander, J.M., Fritz, S.A., Hof, C., Kreft, H., Normand, S., Cabral, J.S., Szekely, E., Thuiller, W., Wikelski, M., Karger, D.N., 2020. Macroecology in the age of Big data – where to go from here? J. Biogeogr. 47, 1–12, http://dx.doi.org/10.1111/jbi.13633.

Yang, W., Ma, K., Kreft, H., 2013. Geographical sampling bias in a large distributional database and its effects on species richness-environment models. J. Biogeogr. 40, 1415–1426, http://dx.doi.org/10.1111/jbi.12108.