Journal of Biogeography

WILEY

Journal of Biogeography

RESEARCH ARTICLE OPEN ACCESS

Biogeographic Regionalization and Transition Zones

Data Completeness and Disciplinary Bias Shape Marine Biogeographic Regionalisation: Annelids From the SW Atlantic

Karoline Azevedo[1] | Victor Emmanuel Lopes da Silva[2] | Karla Paresque[1] | Paulo Pagliosa[3]

[1]Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Maceió, Brazil | [2]Laboratório de Ecologia, Peixes e Pesca – Instituto de Ciências Biológicas e da Saúde, Universidade Federal de Alagoas, Maceió, Brazil | [3]Laboratório de Biodiversidade e Conservação Marinha, Núcleo de Estudos Do Mar, Universidade Federal de Santa Catarina, Florianópolis, Brazil

Correspondence: Karoline Azevedo (karolineakaf@gmail.com)

Received: 2 January 2026 | **Revised:** 12 February 2026 | **Accepted:** 17 March 2026

Keywords: benthic ecology | biodiversity databases | biogeographic regionalisation | marine annelids | null hypotheses | sampling bias | transition zones

ABSTRACT

Aim: Biogeographic regionalisation reliability depends on how well the data capture true species distributions. We examined how data incompleteness and disciplinary bias (taxonomic vs. ecological) influence the detection of marine biogeographic regions and transition zones.

Location: Southwestern Atlantic.

Taxon: Marine Annelids (Polychaeta).

Methods: Using occurrence records from NONATObase (1859–2020), we compared taxonomic, ecological and combined datasets across five marine ecoregions. We estimated diversity using Hill numbers and standardised comparisons through a sample coverage framework. Pairwise null model tests (EcoTest and BiogTest) were applied to assess regional differentiation in composition and diversity structure.

Results: Sampling was heavily concentrated in Eastern and Southeastern Brazil, while Amazonia and Rio Grande were severely undersampled. Ecological and taxonomic datasets showed minimal spatial overlap and distinct methodological traditions (e.g., preference for grabs vs. manual collection). Coverage-based standardisation revealed that many perceived regional similarities were artefacts of data scarcity. Amazonia and Rio Grande emerged as distinct units only after accounting for completeness. In contrast, the Northeastern and Southeastern regions functioned as a transitional system with high species turnover but a similar diversity structure.

Main Conclusions: Marine regionalisation is highly sensitive to disciplinary bias and sampling completeness. Biogeographic boundaries inferred from raw data often reflect research effort rather than biological structure. Integrating disparate data sources and using completeness-standardised metrics are essential for robust biogeographic inference, particularly for understudied marine invertebrates.

1 | Introduction

Understanding how biological diversity is structured across space has long been a central goal of biogeography. Large-scale syntheses of species distributions have played a key role in identifying biogeographic regions, assessing similarities and differences among areas, as well as exploring the processes that shape faunal boundaries (Mora et al. 2008; Sousa-Baena et al. 2013; Troia and McManamay 2016). These regional frameworks underpin much of contemporary biogeographic reasoning

and are widely used in ecological research, conservation planning and biodiversity management (Abram et al. 2015; O'Hara et al. 2017). Their reliability, however, depends fundamentally on the extent to which available data capture the true spatial distribution of species.

Biodiversity data are often incomplete, unevenly distributed and shaped by the geography of scientific effort (Hughes et al. 2021; Thyrring et al. 2024). Limitations in taxonomic resolution and geographic coverage, commonly described as the Linnean and Wallacean shortfalls, arise from differences in accessibility, research infrastructure, funding priorities and methodological approaches (Bini et al. 2006; Hortal et al. 2015; Oliveira et al. 2016; Menegotto and Rangel 2018). As a result, apparent patterns of species richness and composition may reflect where sampling has occurred rather than underlying biological structure, complicating the interpretation of regional boundaries and faunal turnover (Cardoso et al. 2024).

In marine systems, these challenges are particularly pronounced. Sampling in these areas is constrained by depth, logistics and cost, leading to strong concentration of records in coastal and shallow environments, while vast areas remain poorly surveyed (Webb et al. 2010; Duffy and Chown 2017). Recent basin-scale analyses have shown that such biases have direct consequences for biogeographic inference. In the Western Atlantic, for example, biodiversity records are strongly structured by accessibility and research investment, producing uneven environmental representation and spatial gaps that influence estimates of species richness, endemism and regional differentiation (Cardoso et al. 2024). These findings highlight that some perceived biogeographic patterns may be sensitive to data completeness, raising questions about the stability of regional boundaries derived from heterogeneous datasets (Meineke and Daru 2021).

An additional challenge arises from uneven research attention among taxonomic groups. Invertebrates, despite their ecological dominance, are frequently underrepresented in biodiversity inventories and conservation frameworks, largely due to limited public visibility and funding priorities (Martín-López et al. 2009; Cardoso et al. 2011). This imbalance is particularly relevant for biogeographic regionalisation, as regions and transition zones defined primarily from well-studied groups may not accurately reflect patterns present in invertebrate-dominated assemblages. Consequently, the extent to which recognised marine regions represent general biogeographic units across taxa remains uncertain.

Marine annelids provide a suitable system to explore this problem. As one of the most diverse metazoan phyla, with approximately 12,000 valid species, annelids dominate benthic communities across a wide range of marine habitats and play key roles in sediment dynamics, nutrient cycling and trophic interactions (Pamungkas et al. 2019; Capa and Hutchings 2021). Despite their ecological importance, knowledge of annelid spatial distribution remains uneven, particularly in tropical and Southern Hemisphere regions, reflecting broader patterns of sampling bias in marine biodiversity research (Capa and Hutchings 2021).

In the Southwestern Atlantic, annelids are comparatively well documented relative to many other marine invertebrates, with approximately 1400 species reported along the Brazilian coast (Lana et al. 2017). This information has been synthesised in NONATObase, a database compiling published occurrence records from the region and distinguishing between taxonomic and ecological sources (Pagliosa et al. 2014). The availability of this database offers a rare opportunity to evaluate how data completeness and disciplinary origin influence perceived patterns of regional differentiation in a diverse marine invertebrate group.

In this study, we use marine annelids from the Southwestern Atlantic as a model to examine how data incompleteness and disciplinary bias affect the detection of biogeographic regions and potential transition zones. Using occurrence records from NONATObase, we compare patterns of species richness and composition derived from taxonomic, ecological and combined datasets across marine ecoregions. Specifically, we aim to quantify spatial asymmetries in sampling completeness, evaluate how standardizstion by sample coverage influences regional contrasts and examine the role of research effort in shaping inferred biogeographic patterns. By addressing these issues, this study evaluates the robustness of marine biogeographic regionalisation under realistic data limitations and contributes to a broader understanding of how incomplete knowledge shapes biogeographic inference in marine systems.

2 | Material and Methods

2.1 | Data Source and Study Area

This study was based on occurrence records of marine annelids along the Brazilian coast, covering the period from 1859 to 2020. Data were extracted from NONATObase, a database that compiles published records of Polychaeta from the Southwestern Atlantic and adjacent Antarctic regions, spanning from 5° N to 80° S (Pagliosa et al. 2014). NONATObase was selected because, in addition to integrating a broad temporal and spatial range of records, it includes metadata classifying each occurrence according to the disciplinary origin of the study, distinguishing between taxonomic and ecological sources. This feature was essential for evaluating how different research traditions influence perceived biogeographic patterns.

Taxonomic records originated from surveys, taxonomic and phylogenetic studies, and museum collections, whereas ecological records were derived from studies focused on community ecology and biology. To explicitly assess disciplinary effects, all analyses were conducted separately for three datasets: ecological data, taxonomic data and a combined dataset including all records.

The study area was structured according to the marine ecoregions proposed by Spalding et al. (2007), which reflect oceanographic and geomorphological features relevant to benthic assemblages and life history processes of sedentary organisms. Five ecoregions along the Brazilian coast were considered: Amazonia, Northeastern Brazil, Eastern Brazil, Southeastern Brazil and Rio Grande. Oceanic island ecoregions were excluded because their sparse, non-systematic records would create

outliers that distort comparisons of diversity and completeness among mainland regions, rather than clarifying how sampling bias affects biogeographic inference.

2.2 | Data Processing and Spatial Standardisation

Species records from all datasets were organised into an incidence matrix (no duplicates per species per grid cell) using 0.5° grid cells (~55 km). This spatial resolution was chosen to reduce spatial bias associated with uneven sampling density while preserving ecologically meaningful regional patterns. Duplicate records of the same species within a grid cell were removed to avoid inflating occurrence frequencies. All species names were taxonomically validated using the World Register of Marine Species. Records associated with invalid or unavailable names, including nomen dubium, nomen inquirendum and temporary identifications, were excluded to ensure taxonomic consistency across datasets.

2.3 | Diversity Estimation and Sampling Completeness

Species diversity was estimated using analytical rarefaction and extrapolation based on Hill numbers, following the framework proposed by Chao et al. (2014). Diversity was quantified for three orders of q, where $q = 0$ corresponds to species richness, $q = 1$ weights species according to their relative frequencies, and $q = 2$ emphasises the most abundant species. These complementary measures allowed us to evaluate how regional diversity patterns vary depending on the contribution of rare versus common species.

To account for differences in sampling effort among ecoregions and datasets, diversity comparisons were standardised using a sample coverage framework (Chao and Jost 2012). Sample coverage estimates the proportion of the total assemblage represented in the observed sample and provides a direct measure of completeness. Incompleteness was calculated as one minus coverage and can be interpreted as the probability of encountering a previously unrecorded species if an additional individual were sampled. Standardising comparisons by coverage rather than by sample size allowed more reliable comparisons among assemblages with uneven sampling intensity. All rarefaction and extrapolation analyses were conducted using the iNEXT package in R.

2.4 | Tests of Regional Differentiation

To assess whether assemblages from different ecoregions differed in composition, richness or abundance structure, we applied pairwise null model tests following Cayuela et al. (2015). The EcoTest evaluates whether two assemblages could be drawn from the same species pool, focusing on compositional differences, whereas the BiogTest evaluates whether assemblages share similar richness and abundance distributions regardless of species identity. Together, these tests allowed us to distinguish between differences driven by species turnover and differences related to overall diversity structure.

Both sample-based and coverage-based rarefaction approaches were applied to assess how standardisation by completeness influences the detection of regional differences. Randomisation procedures were implemented using 200 bootstrap replicates with the R package rareNMtests (Cayuela and Gotelli 2014).

2.5 | Assessment of Methodological Bias

To evaluate whether taxonomic and ecological datasets differed in sampling practices, we examined the distribution of sampling devices used in each dataset. Each record in NONATObase includes information on one or more sampling methods. A contingency table relating dataset type and sampling device was constructed, and independence between these variables was tested using Fisher's exact test with Monte Carlo simulation of one million replicates due to table sparsity. Standardised residuals from a chi-square test with simulated p values were used to identify sampling devices that were over or underrepresented in each dataset.

3 | Results

3.1 | Dataset Composition and Spatial Distribution

A total of 16,738 records, representing 1033 species and 62 families of marine annelids, distributed across the five marine ecoregions recognised for the Brazilian coast by Spalding et al. (2007): Amazonia (AMA), Northeastern Brazil (NEBR), Eastern Brazil (EBR), Southeastern Brazil (SEBR) and Rio Grande (RG). Among all records, 11,904 originated from ecological studies and 4834 from taxonomic sources. The number of species and sampling units differed consistently among ecoregions, with higher values in SEBR and EBR and lower values in AMA and RG (Table 1).

Within each ecoregion, between 62% and 92% of species and between 7% and 86% of samples were unique to one dataset type. Overall, the proportion of unique species was similar between ecological and taxonomic datasets (65%). In contrast, the distribution of unique samples differed: 22% of ecological samples were exclusive, whereas 58% of taxonomic samples were unique (Figure 1).

3.2 | Diversity Patterns and Sampling Completeness

Sampling completeness varied markedly among ecoregions and between datasets (Figure 2). The ecological dataset showed comparatively higher and more homogeneous sample coverage across ecoregions, suggesting relatively well-sampled assemblages. In contrast, the taxonomic dataset displayed pronounced regional heterogeneity in coverage, with Amazonia and Rio Grande consistently showing substantially lower completeness, indicating strong undersampling. When ecological and taxonomic records were combined, coverage increased overall and became more balanced among ecoregions, although lower completeness in Amazonia persisted. Across all datasets, differences in coverage were consistent across

TABLE 1 | Number of species occurrences and samples (grid cells) for three different datasets from all marine ecoregions along the Brazilian coast.

	Species			Samples		
Marine ecoregions	**Ecological**	**Taxonomic**	**All data**	**Ecological**	**Taxonomic**	**All data**
Amazonia	43 (32)	29 (18)	61 (50)	10 (5)	7 (2)	12 (7)
North eastern	192 (144)	202 (154)	346 (298)	21 (5)	37 (21)	42 (26)
Eastern	204 (174)	203 (173)	377 (347)	21 (5)	105 (89)	110 (94)
South eastern	479 (338)	369 (228)	707 (566)	70 (5)	92 (27)	97 (32)
Rio Grande	51 (33)	112 (94)	145 (127)	30 (13)	43 (26)	56 (39)
Total	**626 (407)**	**626 (407)**	**1033 (814)**	**152 (33)**	**284 (165)**	**317 (198)**

Note: Numbers on parenthesis represent unique species and samples.

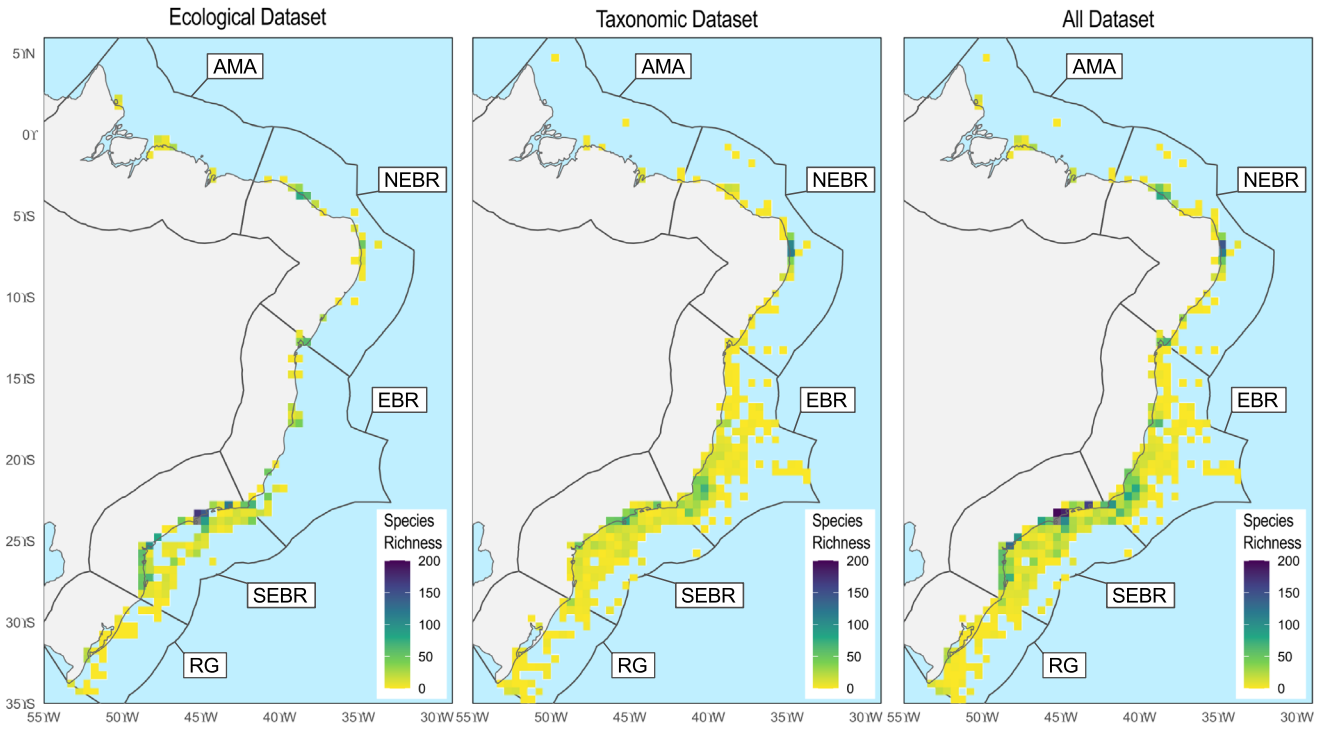

FIGURE 1 | Marine annelid species richness across Brazilian marine ecoregions. Maps show species richness derived from (left) ecological, (center) taxonomic and (right) combined datasets. AMA = Amazonia; EBR = Eastern Brazil; NEBR = Northeastern Brazil; RG = Rio Grande; SEBR = Southeastern Brazil.

Hill numbers, indicating that patterns of undersampling affected richness ($q = 0$), common species ($q = 1$) and dominant species ($q = 2$) in a similar manner.

3.3 | Regional Differentiation and Transition Patterns

Coverage-based rarefaction revealed clear patterns of regional differentiation across marine ecoregions, elucidating contrasts that were often obscured under sample-based standardisation. Across datasets, regional structure emerged most prominently when sampling completeness was accounted for, enabling a sharper distinction between robust biogeographic units and regions exhibiting transitional characteristics. Sample-based rarefaction tends to yield higher rates of non-rejection, because the curves tend to converge at very small sample sizes, masking real differences (Figure 3). The results of all 30 paired comparison tests are fully reported in the Data S1.

In the ecological dataset, Amazonia consistently emerged as a distinct biogeographic unit whenever comparisons were evaluable. Coverage-based analyses rejected both ecological and biogeographic null hypotheses across all diversity orders, indicating pronounced differentiation in species composition, richness and abundance structure (Figure 3). These contrasts were less consistently detected under sample-based standardisation, particularly for Shannon ($q = 1$) and Simpson ($q = 2$) diversity, underscoring the role of incomplete standardisation in inflating apparent similarity. Comparisons between Amazonia

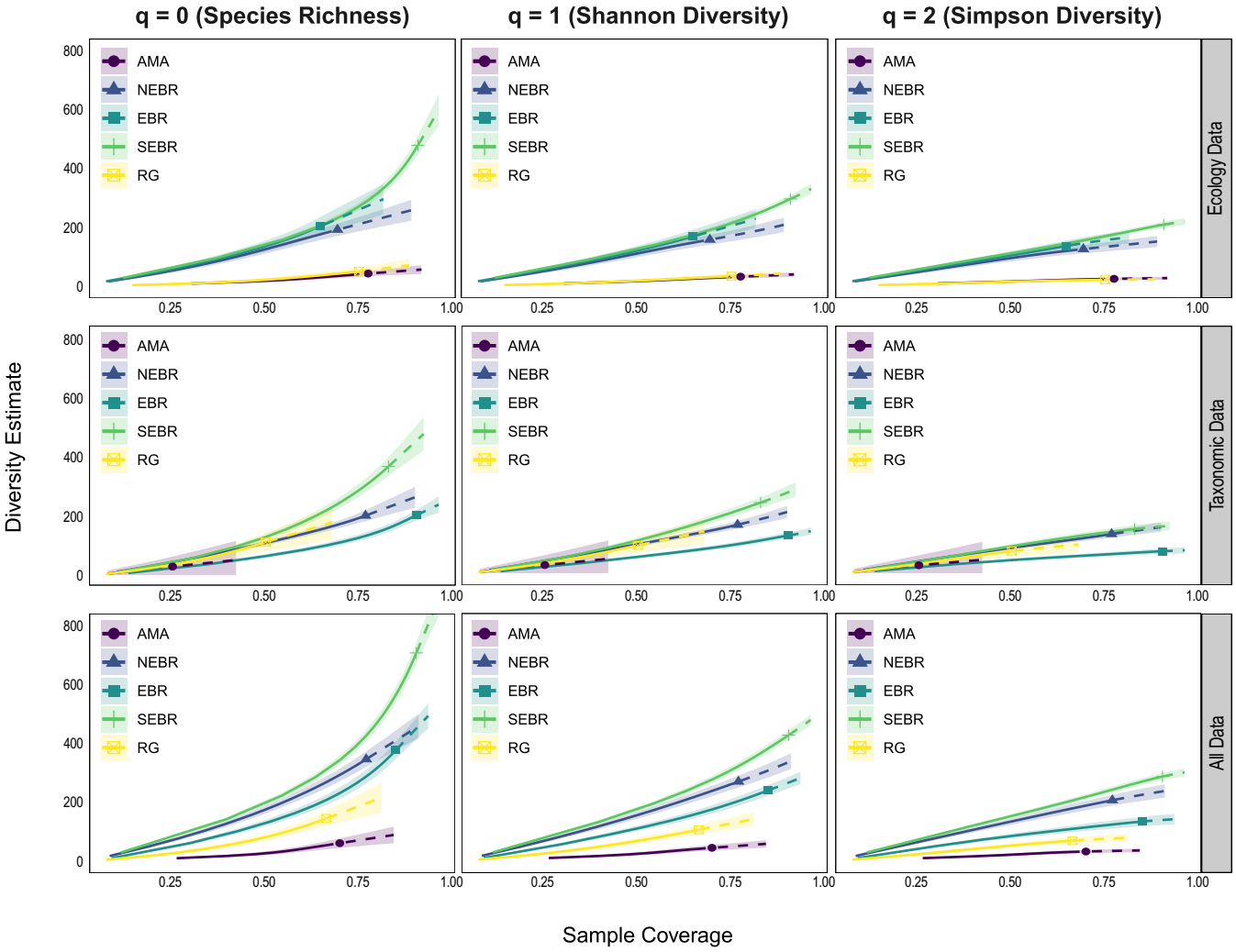

FIGURE 2 | Coverage-based rarefaction curves for five marine ecoregions along the Brazilian coast among three different datasets based on the purposes of study, using Hill numbers (0, 1, 2). AMA = Amazonia; EBR = Eastern Brazil; NEBR = Northeastern Brazil; RG = Rio Grande; SEBR = Southeastern Brazil.

and Rio Grande could not be evaluated due to insufficient overlap in sampling units. Beyond Amazonia, Rio Grande frequently showed differentiation from other ecoregions, particularly under coverage-based analyses, revealing divergence in both composition and diversity structure relative to Eastern, Northeastern and Southeastern Brazil. In contrast, comparisons among Eastern, Northeastern and Southeastern regions predominantly indicated similar richness and abundance profiles, suggesting faunal continuity across adjacent regions as part of a broader transitional system rather than sharply delimited biogeographic units. No comparison in the ecological dataset under coverage-based tests showed rejection restricted to a single Hill number (Figure 3).

Patterns in the taxonomic dataset were more strongly constrained by limited sampling, particularly in Amazonia, where most pairwise comparisons could not be evaluated under either rarefaction approach (Figure 3). Where coverage-based tests were feasible, results revealed a mixture of differentiation and similarity. Eastern and Southeastern regions exhibited patterns broadly consistent with those observed in the ecological dataset, whereas contrasts involving Amazonia and Southeastern Brazil highlight the limitations imposed by sparse taxonomic coverage, reducing the power to consistently reject null hypotheses. Several taxonomic comparisons displayed discordance between ecological and biogeographic null hypotheses, with some contrasts indicating compositional turnover without corresponding shifts in richness or abundance structure, and others showing the opposite pattern.

When ecological and taxonomic data were combined, regional patterns became more coherent and internally consistent (Figure 3). Coverage-based analyses consistently identified Amazonia as distinct from all other ecoregions across all diversity orders, indicating that differences in richness and abundance structure persist once sampling completeness is accounted for. Contrasts involving Rio Grande also became more pronounced. Northeastern and Southeastern Brazil continued to reinforce their characterisation as a transitional region marked by species turnover but broadly similar diversity structure. Across datasets, coverage-based results showed greater consistency across Hill numbers, reflecting the stabilising effect of accounting for sampling completeness on inferences of regional differentiation.

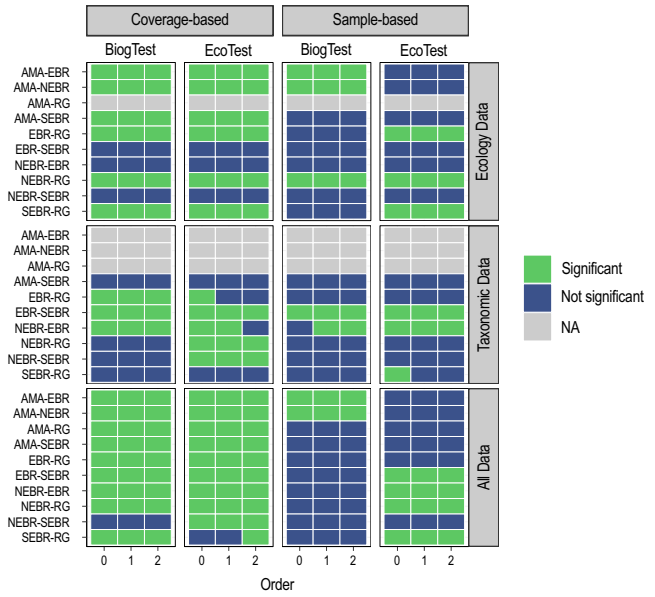

FIGURE 3 | Results of pairwise ecological (EcoTest) and biogeographic (BiogTest) null hypothesis tests among marine ecoregions across diversity orders ($q = 0$, 1, 2), comparing coverage-based and sample-based rarefaction approaches. Panels are organised by dataset type (taxonomic, ecological and combined) and standardisation method. Colours indicate test outcomes: Red denotes significant differences (rejection of the null hypothesis), green indicates non-significant results (failure to reject the null hypothesis), and grey represents comparisons that could not be evaluated due to insufficient data. Comparisons are shown for all regional pairs included in the analysis. AMA = Amazonia; EBR = Eastern Brazil; NEBR = Northeastern Brazil; RG = Rio Grande; SEBR = Southeastern Brazil.

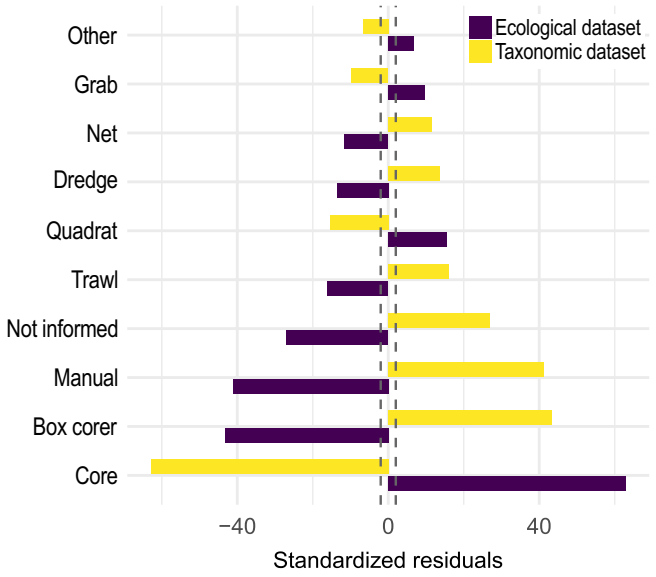

FIGURE 4 | Differences in sampling gear use between taxonomic and ecological datasets. Bars represent standardised residuals from the chi-square test; positive and negative values indicate over- and underrepresentation relative to independence expectations. Dashed lines mark approximate significance thresholds (±2).

3.4 | Influence of Sampling Methods

Sampling device usage differed markedly between ecological and taxonomic datasets (Fisher's exact test, $\chi^2 = 7.700$, $p < 0.001$). Ecological studies were positively associated with the use of box corers, grabs and quadrats, while taxonomic studies were more frequently associated with core sampling, manual collection and dredging (Figure 4). These differences were statistically significant and reflected distinct methodological traditions, with direct implications for spatial coverage and taxonomic resolution.

4 | Discussion

Our results demonstrate that the reliability of marine biogeographic regionalisation for marine annelids is strongly contingent on data completeness and disciplinary bias, with direct implications for the recognition of biogeographic regions and transition zones. Although ecoregional frameworks such as those proposed by Spalding et al. (2007) are widely used as proxies for biogeographic structure, our findings indicate that uneven sampling and methodological heterogeneity can either obscure or exaggerate regional differentiation, particularly for understudied taxa. Consequently, biogeographic patterns inferred without explicit consideration of data limitations may reflect artefacts of data structure rather than underlying biological processes.

At a basin scale, the Southwestern Atlantic provides a clear illustration of how historical and institutional factors shape the spatial distribution of marine biodiversity data. Records of marine annelids were heavily concentrated in Eastern and Southeastern Brazil, regions that host most research institutions and trained specialists, whereas Amazonia and Rio Grande remained comparatively underrepresented (Lana et al. 2017; Longo and Amado Filho 2014). This uneven geographic footprint directly affects the stability of regional comparisons, as well-sampled ecoregions consistently yielded evaluable and coherent null model outcomes, while poorly sampled regions were more prone to non-evaluable results. These spatial asymmetries are compounded by disciplinary differences in how data are produced, sustained by the level of research capacity, where accessibility and research investment strongly structure available data and influence estimates of species richness and regional differentiation (Cardoso et al. 2024).

As documented by the limited overlap among sampling units across datasets, most grid cells were exclusive to a single dataset type, with taxonomic studies contributing disproportionately to taxonomic resolution and ecological datasets emphasising standardised replication and abundance structure (Halme et al. 2015). The strong separation observed between taxonomic and ecological datasets highlights how differences in research goals, sampling strategies and historical trajectories shape the composition of available data (Díaz-Díaz et al. 2017; Lana et al. 2017). Although combining datasets increased overall spatial coverage, it also implies that regional patterns derived from a single disciplinary perspective might be necessarily incomplete, echoing concerns raised in discussions of the Linnean and Wallacean shortfalls in biodiversity knowledge (Cardoso et al. 2011; Hortal et al. 2015).

Ecological studies prioritise comparability and environmental inference through standardised sampling designs, whereas

taxonomic studies rely on specimen-centred strategies aimed at detecting rarity and novelty. Consequently, the larger spatial coverage area and the higher proportion of single grid cells in the taxonomic dataset reveal a fragmented sampling mosaic, shaped by historical expeditions and institutional trajectories associated with a small number of specialists and research hubs (Lana et al. 2017; Nascimento et al. 2024). This structural mismatch explains why coverage-based rarefaction frequently failed to stabilise null model tests in the taxonomic dataset, particularly for Amazonia-related comparisons, where NA outcomes dominated across Hill numbers.

When these biases are not accounted for, apparent similarities among regions may be misleading. This effect is particularly pronounced under sample-based rarefaction, which tends to yield higher rates of non-rejection because rarefaction curves converge at very small sample sizes, effectively masking real differences among assemblages (Cayuela et al. 2015). By contrast, coverage-based standardisation mitigates this artefact by equalising sampling completeness rather than sample size. By standardising comparisons using sample coverage, we show that several similarities among ecoregions arose primarily from uneven sampling rather than true biological homogeneity. This effect was particularly evident in comparisons involving Amazonia and Rio Grande, which in several cases shifted from non-significant to significant once coverage was considered, while in others became evaluable only under coverage-based standardisation, indicating that incomplete data masked underlying regional differentiation (Miloslavich et al. 2011; Costello et al. 2015). Coverage-based approaches therefore proved essential for distinguishing genuinely similar assemblages from those that merely appear similar due to data scarcity, reinforcing recent basin-scale assessments that advocate completeness-standardised metrics in biogeographic studies (Cardoso et al. 2024).

Importantly, not all patterns were sensitive to sampling variation. The persistent differentiation between Eastern Brazil and Southeastern Brazil across datasets, diversity orders, and null models suggests a robust biogeographic boundary that is resilient to data limitations (Lana et al. 2017; Miloslavich et al. 2011). Such consistency indicates that some marine ecoregions represent stable and biologically meaningful units, even under imperfect sampling conditions. These findings support the idea that well-sampled regions are more likely to yield reliable signals of regionalisation, whereas poorly sampled areas are prone to misclassification or omission in regional schemes (Cardoso et al. 2011; Costello et al. 2015).

Between these extremes, patterns of partial similarity observed among adjacent tropical ecoregions, particularly between Northeastern Brazil and Southeastern Brazil, point to the presence of transition zones rather than sharply delimited biogeographic units. These areas exhibited compositional differences coupled with gradual faunal turnover or species overlap (Spalding et al. 2007; Miloslavich et al. 2011). Transition zones are increasingly recognised as integral components of biogeographic systems; however, our results show that their detection is highly sensitive to data completeness and analytical choices (Cardoso et al. 2011; Hortal et al. 2015), with the risk that sampling biases may either obscure them or artificially exaggerate continuity between regions. In this context, the persistent non-significance observed between Rio Grande and Southeastern ecoregions across different datasets and null models presented here may support the interpretation that these regions cannot be statistically distinguished with the available data, potentially reflecting a single coherent unit at the upper slope (200–1000 m), as suggested by independent deep-sea analyses (Gaurisas and Bernardino 2023).

Finally, methodological heterogeneity between ecological and taxonomic datasets further reinforces these conclusions. Distinct associations between dataset type and sampling devices reflect long-standing disciplinary traditions that influence both spatial coverage and taxon detectability (Lana et al. 2017; Díaz-Díaz et al. 2017). Despite its clear potential to bias biogeographic inference, such methodological variation has rarely been incorporated explicitly into marine regionalisation studies. Recognising and accounting for these differences is therefore critical for improving the robustness and interpretability of regional delineations (Cardoso et al. 2011, 2024).

Our study demonstrates that marine biogeographic regionalisation cannot be reliably interpreted without explicit consideration of data completeness and disciplinary bias. Although focused on marine annelids, the patterns documented here likely reflect broader processes affecting understudied marine invertebrate taxa, characterised by high diversity and persistent spatial asymmetries in data availability. By concentrating on continental coastal ecoregions, where sampling is more continuous and shaped by sustained research programs rather than episodic expeditions typical of oceanic islands, we show that uneven research effort and methodological heterogeneity strongly influence the recognition of biogeographic regions and transition zones in the Southwestern Atlantic. Integrating coverage-based approaches and complementary data sources therefore provides a more realistic assessment of regional structure while underscoring the need for sustained investment in biodiversity inventories and taxonomic expertise, particularly in underrepresented regions, to strengthen the empirical foundations of marine biogeographic frameworks.

Author Contributions

K.A. conceived the ideas and designed the methodology, together with V.E.L.S. and P.P. K.A. collected and analysed the data. K.A. and V.E.L.S. led the writing of the manuscript. V.E.L.S., K.P. and P.P. reviewed and improved the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

Acknowledgements

No fieldwork was conducted for this study. All data analysed were derived from publicly available, open-access repositories. As such, no field permits or collection licences were required. This work was funded by 'Coordenação de Aperfeiçoamento de Pessoal de Nível Superior'—Brazil (CAPES)—Finance Code 001. K.A. was funded by the National Institute of Science and Technology (INCT) in Ecology, Evolution and Biodiversity Conservation, funded by CNPq (grant 465610/2014-5/409197/2024-6). V.E.L.S. received a postdoc Fellowship from CAPES. The Article Processing Charge for the publication of

this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614).

Funding

This work was supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (001).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

No fieldwork was conducted for this study. All data analyzed were derived from publicly available, open-access repositories. As such, no field permits or collection licenses were required. The data that supports the findings of this study are available in the Supporting Information of this article.

References

Abram, N. K., E. Meijaard, J. A. Wells, et al. 2015. "Mapping Perceptions of Species' Threats and Population Trends to Inform Conservation Efforts: The Bornean Orangutan Case Study." *Diversity and Distributions* 21, no. 5: 487–499. https://doi.org/10.1111/ddi.12286.

Bini, L. M., J. A. F. Diniz-Filho, T. F. L. V. B. Rangel, R. P. Bastos, and M. P. Pinto. 2006. "Challenging Wallacean and Linnean Shortfalls: Knowledge Gradients and Conservation Planning in a Biodiversity Hotspot." *Diversity and Distributions* 12, no. 5: 475–482. https://doi.org/10.1111/j.1366-9516.2006.00286.x.

Capa, M., and P. Hutchings. 2021. "Annelid Diversity: Historical Overview and Future Perspectives." *Diversity* 13, no. 3: 129. https://doi.org/10.3390/d13030129.

Cardoso, M. N. M., F. Azevedo, A. Dias, et al. 2024. "Causes and Effects of Sampling Bias on Marine Western Atlantic Biodiversity Knowledge." *Diversity and Distributions* 30, no. 6: e13839. https://doi.org/10.1111/ddi.13839.

Cardoso, P., P. A. V. Borges, K. A. Triantis, M. A. Ferrández, and J. L. Martín. 2011. "Adapting the IUCN Red List Criteria for Invertebrates." *Biological Conservation* 144: 2432–2440. https://doi.org/10.1016/j.biocon.2011.06.020.

Cayuela, L., and N. J. Gotelli. 2014. "rareNMtests: Ecological and Biogeographical Null Model Tests for Comparing Rarefaction Curves." R package version, 1.

Cayuela, L., N. J. Gotelli, and R. K. Colwell. 2015. "Ecological and Biogeographic Null Hypotheses for Comparing Rarefaction Curves." *Ecological Monographs* 85, no. 3: 437–455. https://doi.org/10.1890/14-1261.1.

Chao, A., N. J. Gotelli, T. C. Hsieh, et al. 2014. "Rarefaction and Extrapolation With Hill Numbers: A Framework for Sampling and Estimation in Species Diversity Studies." *Ecological Monographs* 84, no. 1: 45–67. https://doi.org/10.1890/13-0133.1.

Chao, A., and L. Jost. 2012. "Coverage-Based Rarefaction and Extrapolation: Standardizing Samples by Completeness Rather Than Size." *Ecology* 93, no. 12: 2533–2547. https://doi.org/10.1890/11-1952.1.

Costello, M. J., M. Lane, S. Wilson, and B. Houlding. 2015. "Factors Influencing When Species Are Named and Estimating Global Species Richness." *Global Ecology and Conservation* 4: 243–254. https://doi.org/10.1016/j.gecco.2015.07.001.

Díaz-Díaz, O., D. Bone, C. T. Rodríguez, and V. H. Delgado-Blas. 2017. *Poliquetos de Sudamérica. Volumen Especial del Boletín del Instituto Oceanográfico de Venezuela*, 149. Buletín del Instituto Oceanográfico de Venezuela.

Duffy, G. A., and S. L. Chown. 2017. "Explicitly Integrating a Third Dimension in Marine Species Distribution Modelling." *Marine Ecology Progress Series* 564: 1–8. https://doi.org/10.3354/meps12011.

Gaurisas, D. Y., and A. F. Bernardino. 2023. "Benthic Biogeographic Patterns on the Deep Brazilian Margin." *PeerJ* 11: e14585. https://doi.org/10.7717/peerj.14585.

Halme, P., S. Kuusela, and A. Juslén. 2015. "Why Taxonomists and Ecologists Are Not, but Should Be, Carpooling?" *Biodiversity and Conservation* 24, no. 7: 1831–1836. https://doi.org/10.1007/s10531-015-0899-3.

Hortal, J., F. de Bello, J. A. F. Diniz-Filho, T. M. Lewinsohn, J. M. Lobo, and R. J. Ladle. 2015. "Seven Shortfalls That Beset Large-Scale Knowledge of Biodiversity." *Annual Review of Ecology, Evolution, and Systematics* 46, no. 1: 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400.

Hughes, A. C., M. C. Orr, K. Ma, et al. 2021. "Sampling Biases Shape Our View of the Natural World." *Ecography* 44, no. 9: 1259–1269. https://doi.org/10.1111/ecog.05926.

Lana, P. C., P. R. Pagliosa, P. C. Paiva, et al. 2017. "Polychaetes in Brazil: People and Places, Past, Present and Future." In *Poliquetos de Sudamérica*, 24–50. Institute of Venezuela.

Longo, L. d. L., and G. M. Amado Filho. 2014. "Knowledge of Brazilian Benthic Marine Fauna Throughout Time." *História, Ciências, Saúde-Manguinhos* 21, no. 3: 995–1010. https://doi.org/10.1590/s0104-59702014000300011.

Martín-López, B., C. Montes, L. Ramírez, and J. Benayas. 2009. "What Drives Policy Decision-Making Related to Species Conservation?" *Biological Conservation* 142, no. 7: 1370–1380. https://doi.org/10.1016/j.biocon.2009.01.030.

Meineke, E. K., and B. H. Daru. 2021. "Bias Assessments to Expand Research Harnessing Biological Collections." *Trends in Ecology & Evolution* 36, no. 12: 1071–1082. https://doi.org/10.1016/j.tree.2021.08.003.

Menegotto, A., and T. F. Rangel. 2018. "Mapping Knowledge Gaps in Marine Diversity Reveals a Latitudinal Gradient of Missing Species Richness." *Nature Communications* 9, no. 1: 1–6. https://doi.org/10.1038/s41467-018-07217-7.

Miloslavich, P., E. Klein, J. M. Díaz, et al. 2011. "Marine Biodiversity in the Atlantic and Pacific Coasts of South America: Knowledge and Gaps." *PLoS One* 6, no. 1: e14631. https://doi.org/10.1371/journal.pone.0014631.

Mora, C., D. P. Tittensor, and R. A. Myers. 2008. "The Completeness of Taxonomic Inventories for Describing the Global Diversity and Distribution of Marine Fishes." *Proceedings of the Royal Society B* 275, no. 1631: 149–155. https://doi.org/10.1098/rspb.2007.1315.

Nascimento, R. L., S. Mendes, M. V. C. Vital, and P. C. Paiva. 2024. "Polychaete Research in Brazil: A Bibliometric Analysis." *Ocean and Coastal Research* 72, no. suppl 1: 1–14. https://doi.org/10.1590/2675-2824072.23105.

O'Hara, C. C., J. C. Afflerbach, C. Scarborough, K. Kaschner, and B. S. Halpern. 2017. "Aligning Marine Species Range Data to Better Serve Science and Conservation." *PLoS One* 12, no. 5: e0175739. https://doi.org/10.1371/journal.pone.0175739.

Oliveira, U., A. P. Paglia, A. D. Brescovit, et al. 2016. "The Strong Influence of Collection Bias on Biodiversity Knowledge Shortfalls of Brazilian Terrestrial Biodiversity." *Diversity and Distributions* 22, no. 12: 1232–1244. https://doi.org/10.1111/ddi.12489.

Pagliosa, P. R., G. Doria, D. Misturini, et al. 2014. "NONATObase: A Database for Polychaeta (Annelida) From the Southwestern Atlantic

Ocean." *Database* 2014: bau002. https://doi.org/10.1093/database/bau002.

Pamungkas, J., C. J. Glasby, G. B. Read, S. P. Wilson, and M. J. Costello. 2019. "Progress and Perspectives in the Discovery of Polychaete Worms (Annelida) of the World." *Helgoland Marine Research* 73, no. 4: 1–10. https://doi.org/10.1186/s10152-019-0524-z.

Sousa-Baena, M. S., L. C. Garcia, and A. T. Peterson. 2013. "Completeness of Digital Accessible Knowledge of the Plants of Brazil and Priorities for Survey and Inventory." *Diversity and Distributions* 20, no. 4: 369–381. https://doi.org/10.1111/ddi.12136.

Spalding, M. D., H. E. Fox, G. R. Allen, et al. 2007. "Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas." *Bioscience* 57, no. 7: 573–583. https://doi.org/10.1641/B570707.

Thyrring, J., L. S. Peck, M. K. Sejr, J. M. Węsławski, C. D. Harley, and A. Menegotto. 2024. "Shallow Coverage in Shallow Waters: The Incompleteness of Intertidal Species Inventories in Biodiversity Database Records." *Ecography* 2024, no. 12: e07006. https://doi.org/10.1111/ecog.07006.

Troia, M. J., and R. A. McManamay. 2016. "Filling in the GAPS: Evaluating Completeness and Coverage of Open-Access Biodiversity Databases in the United States." *Ecology and Evolution* 6, no. 14: 4654–4669. https://doi.org/10.1002/ece3.2225.

Webb, T. J., E. van den Berghe, and R. O'Dor. 2010. "Biodiversity's Big Wet Secret: The Global Distribution of Marine Biological Records Reveals Chronic Under-Exploration of the Deep Pelagic Ocean." *PLoS One* 5, no. 8: e10223. https://doi.org/10.1371/journal.pone.0010223.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Full results of pairwise ecological (EcoTest) and biogeographic (BiogTest) null model analyses among marine ecoregions, including all datasets and rarefaction approaches. For each comparison, the table reports sampling descriptors (species, samples, records) and test outcomes across diversity orders ($q = 0, 1, 2$). Non-evaluable cases are indicated.