# PLOS ONE

# The application of historic sample-specific variables in evaluating the biodiversity patterns of the South African azooxanthellate scleractinians (Cnidaria: Anthozoa). --Manuscript Draft--



#### examples.

This statement is required for submission and will appear in the published article if the submission is accepted. Please make sure it is accurate.

### Unfunded studies

Enter: The author(s) received no specific funding for this work.

#### Funded studies

Enter a statement with the following details:

- Initials of the authors who received each award
- Grant numbers awarded to each author
- The full name of each funder
- URL of each funder website
- Did the sponsors or funders play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript?
- NO Include this sentence at the end of your statement: The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.
- YES Specify the role(s) played.

#### \* typeset

#### Competing Interests

Use the instructions below to enter a competing interest statement for this submission. On behalf of all authors, disclose any [competing interests](http://journals.plos.org/plosone/s/competing-interests) that could be perceived to bias this work—acknowledging all financial support and any other relevant financial or nonfinancial competing interests.

This statement is required for submission and will appear in the published article if the submission is accepted. Please make sure it is accurate and that any funding sources listed in your Funding Information later in the submission form are also declared in your Financial Disclosure statement.

## The authors have declared that no competing interests exist.



#### Format for specific study types

## Human Subject Research (involving human participants and/or tissue)

- Give the name of the institutional review board or ethics committee that approved the study
- Include the approval number and/or a statement indicating approval of this research
- Indicate the form of consent obtained (written/oral) or the reason that consent was not obtained (e.g. the data were analyzed anonymously)

#### Animal Research (involving vertebrate

### animals, embryos or tissues)

- Provide the name of the Institutional Animal Care and Use Committee (IACUC) or other relevant ethics board that reviewed the study protocol, and indicate whether they approved this research or granted a formal waiver of ethical approval
- Include an approval number if one was obtained
- If the study involved non-human primates, add additional details about animal welfare and steps taken to ameliorate suffering
- If anesthesia, euthanasia, or any kind of animal sacrifice is part of the study, include briefly which substances and/or methods were applied

## Field Research

Include the following details if this study involves the collection of plant, animal, or other materials from a natural setting: • Field permit number

• Name of the institution or relevant body that granted permission

## Data Availability

Authors are required to make all data underlying the findings described fully available, without restriction, and from the time of publication. PLOS allows rare exceptions to address legal and ethical concerns. See the [PLOS Data Policy](http://journals.plos.org/plosone/s/data-availability) and [FAQ](http://journals.plos.org/plosone/s/data-availability#loc-faqs-for-data-policy) for detailed information.

Yes - all data are fully available without restriction





 $\triangleq$ 



 by analysing the environmental correlates of museum samples. The associated coordinate data were georeferenced and depth obtained from a national bathymetric dataset, prior to undertaking a multivariate analysis. Overall, our results confirmed two longitudinal groups (eastern margin [group A] vs southern and western margin [group B]) and 11 depths represented within two bathymetric zones (shallow [50-200 m] and deep [300-1000 m]). Both the longitudinal groups and depth zones partially explained coral distribution patterns, with depth highly correlated with species variation. Caryophylliids, flabellids, and dendrophylliids contributed the most towards distinguishing longitudinal and depth gradients. Data limitations within our data set resulted to unexplained variance, however, despite these limitations, the study demonstrates that historical museum samples provide a valuable data source that can fill research sampling gaps and improve our understanding of biodiversity patterns of the coral fauna in under sampled marine ecosystems.

**Keywords:** stony Cold-water corals, longitude, depth, gradients, distribution.

 $\blacksquare$ 

# **1. Introduction**

 The distribution of azooxanthellate corals, a group of scleractinian species that lack a symbiotic relationship with photosynthetic dinoflagellates, is influenced by environmental variables at various scales (Guinotte et al., 2006; Hovland, 2008; Roberts et al., 2009; Davies and Guinotte, 2011; Angeletti et al., 2020). Physical and chemical oceanographic factors, as well as geomorphologic settings affect food supply and, consequently, benthopelagic coupling (Roberts et al., 2009). Overall, depth might be used as a variable linked to several oceanographic factors that influence species distributions. For example, coral species have preferred thermal ranges (Davies and Guinotte, 2011), and a global azooxanthellate coral richness trend has been

 documented between 200 and 1000 m deep. This depth range often coincides with shelf and slope features, which may provide suitable substrate for larval settlement and habitats for azooxanthellate coral species to colonise (Cairns, 2007; Roberts et al., 2009). Furthermore, long- term environmental stability appears to also be important for the occurrence/distribution of deep water stony coral species. In addition to the temporal and spatial stability of an environment, it is well established that life history patterns, including reproduction strategies and relationship to substrate, are of utmost importance for a species' distribution (Oakham, 2009). For instance, attached deep water scleractinians require consolidated substrates to survive, whilst unattached forms are found on or in unconsolidated substrates (Roberts et al., 2009; Hovland, 2008).

 Given the difficulty of sampling in deep-water marine systems, the mapping and classification of biodiversity into spatial units (which then act as surrogates for unmapped biodiversity) is a common approach in spatial planning (Waters, 2008; Costello, 2009; Reygondeau and Dunn, 2018; Reygondeau 2019; Richter et al., 2022). Considering the growing concern regarding declining ocean health, voluntary commitments to reach a national 30 % area protection by 2030, and the United Nations call for better ocean governance (United Nations, 2018; 2019; 2020), such spatial classifications are powerful tools to guide conservation and management strategies to 65 support the achievement of the United Nations  $14<sup>th</sup>$  Sustainable Development Goals (SDGs). Although ocean basins have been mapped at broad scales (Sayre et al., 2017) and several global and regional bioregionalisations exist (Ekman, 1953; Briggs, 1974; Spalding et al., 2007; Grant et al., 2006; Cedras et al., 2020; McQuaid et al., 2023), few studies have used species data to describe biological patterns in areas deeper than 200 m (Zezina, 1997; O'Hara et al., 2011; Watling et al., 2013; Cedras et al., 2020; Summers and Watling, 2021; Watling and Lapointe, 2022), particularly  within benthic ecosystems (O'Hara et al., 2011; Summers and Watling, 2021; Watling and Lapointe, 2022). Developing nations, such as South Africa, particularly lack specialised resources to survey deeper waters, further constraining *in situ* research (Bell et al., 2022). Consequently, samples housed in scientific collections may be a valuable source of biological data to evaluate distribution patterns (Spalding et al., 2007; Thandar et al., 1989; Woolley et al., 2020) but should be interpreted with caution.

 Despite the fact that the first marine collections along South Africa's shores dates back to the 1700s (Day, 1977; Griffiths et al., 2010), ocean resource management is still constrained by the poor state of knowledge of key invertebrate species (Sink et al., 2019). Endeavouring to address such species data gaps, local research advancements have recently been initiated by examining natural history collections (Biccard 2012; Laird 2013; Filander, 2014; Olbers 2016; Boonzaaier, 2017; Landschoff 2011). Some of these studies have been integrated into the ecosystem map developed by Sink et al. (2019; 2023) for the National Biodiversity Assessment (NBA). The NBA used pelagic and benthic data, including biological available information (i.e., macrofauna, epifauna, and fish) to produce an expert-driven ecosystem type map for national reporting frameworks. It comprises four hierarchical levels that represent six ecoregions (two deep ocean and four confined to the continental shelf), five depth zones (shore, shelf, slope, plateau, and abyss), and different substrate types. Absent, however, from this national spatial classification map 90 is a holistic consideration of the South African azooxanthellate scleractinian fauna, given that this taxonomic group was only recently reviewed (Filander et al., 2021), and its distribution patterns had not yet been investigated. The NBA does however report on some distribution records of

 potential Vulnerable Marine Ecosystem indicator taxa, which includes records of two reef-building azooxanthellate coral taxa (i.e., Dendrophylliidae Gray, 1847 and Caryophylliidae Dana, 1846).

 Earlier international studies (Cairns, 2007) have grouped the available literature on azooxanthellate Scleractinia into broad geographic regions. Although not a biodiversity analysis, Cairns (2007) produced an output that served as a starting point for emerging taxonomists in the field. Cairns and Keller (1993) did, nonetheless, summarise depth affiliations within the southwest Indian Ocean, in which South African taxa reported off the eastern and southern margins were represented. Apart from these two publications (Cairns and Keller, 1993; Cairns, 2007), the South African azooxanthellate Scleractinia distribution pattern have not been investigated in light of its 103 relationship to physical variables. Therefore, this paper aims to examines the diversity measures of the South African azooxanthellate coral fauna, with respect to sample-specific environmental Ş. gradients.

**2. Material and Methods**

 Data considered for this study were based primarily on a subset of species distribution records for the South African azooxanthellate scleractinian fauna recently reported by Filander et al. (2021). The subset was compiled by including those occurrence data with coordinate information and a sample number, resulting in 761 occurrence records (**Figure 1** and **Appendix A**: Occurrence data). 112 These coral occurrence data were predominately collected during six historical dredge surveys undertaken between 1898 and 1990 (i.e., RV *Anton Bruun*, Benguela IV, RV *Meiring Naude,* RV *Pieter Faure*, Sardinops, and University of Cape Town Ecological Surveys). The recent surveys 115 undertaken in the  $21^{st}$  century are represented by two trawl (NANSEN and Department of



# **2.1. Assumptions and sampling biases**

 Over 80% of the resulting data is of historical origin, and therefore poses some limitations. One of these limitations is sampling coverage bias, given that past national marine surveys focused mainly 139 on nearshore areas due to their accessibility, whilst sampling in areas beyond the continental shelf

 relied on international surveys (the *Pieter Faure* expeditions being an exception) (Griffiths et al., 2010). Nevertheless, these historical surveys represent decades of sampling effort but were not 142 systematic and provide **presence-only data.** Secondly, depth and co-ordinate information are the only two variables commonly associated with such datasets, but may be unreliable in some instances (i.e., the *Pieter Faure* collection). Thus, the use of the occurrence data in the multi- variate analysis required three assumptions that may not necessarily be a true reflection of the data 146 attributes. These include presenting the data as presence-absence, inferring the occurrence of the historical records to the modern day, and applying longitude and depth as a proxy for ocean basin and water mass properties; respectively. The above-mentioned two-part data geo-referencing methodology was therefore also undertaken to standardize the associated collection specific parameters for the application of longitude and depth data as abiotic variables.

 Furthermore, the data preparation methodology does not follow the interpolation of the presence-153 absence matrix (i.e., if species occur between two extreme points, then occurrence is assumed in between) as conducted in preceding marine benthic invertebrate studies based on museum specimens (Filander, 2014; Olbers, 2016; Boonzaaier, 2017). This approach would have yielded unrealistic conclusions in the absence of fine-scale substrate data sources- as substrate type is one of the primary drivers of coral settlement. Whereas the 2018 NBA (Sink et al., 2019) sub-divided substrate into ecosystem types, the multibeam data represented less than 1% of the South African seabed. Cawthra et al. (2021) review on existing South African core samples collected from areas deeper that 130 m highlights the importance of high resolution hydroacoustic surveys to better 161 contextualize published core localities. It is for this reason that the substrate level was not considered to support interpolation techniques.

Ę.

E

 Lastly, the average taxonomic distinctiveness (ATD) diversity measure was based on the 164 established phylogenetic clades (Kitahara et al., 2010; Stolarski et al., 2011). However, owing to limited resolution regarding species relationships below family level, phylogenetic scores were not assigned to taxonomic levels lower than family. Additionally, the existing phylogenetic reconstructions still lack sufficient representation of azooxanthellate coral species occurring in South Africa. For instance, less than 20% of South African coral species have been sequenced and included in existing molecular trees (Kitahara et al., 2010; Stolarski et al., 2011). It is important to note that ATD is a diversity calculation method that considers the distance between each species and its closest relative outside the group. This calculation is then divided by the number of species within the group being evaluated. The resulting ATD value provides an estimation of the group's evolutionary uniqueness, with higher values indicating greater distinctiveness. Consequently, an alphabetically arranged method would yield inaccurate results.

# **2.2. Data analysis**

 A presence-absence matrix (**Appendix A**: presence-absence) of the coral occurrence data was compiled and all analyses were undertaken using the PRIMER 7 software package (Clarke and Warwick, 2001; Clarke et al., 2014), with the PERMANOVA+ add on (Anderson et al., 2008). The matrix, consisting of 488 columns (stations/samples) and 95 rows (species), was converted to a resemblance matrix. The associated higher taxonomic classifications of these resulting species identifications were thereafter extracted from the World Register of Marine Species batch match online function (WoRMS Editorial Board, 2021) (**Appendix A**: Taxonomic attributes). Owing to the patchy nature of the data set, in which 30 species were represented by only one sample and 22 species by less than ten samples (**Appendix A**: Number of records per species), the Gamma+

Е

 dissimilarity matrix was selected (Clark and Warwick, 1999; Clark et al., 2006). This measure used the cophenetic distances derived from the phylogeny established in Kitahara et al. (2010) and Stolarski et al. (2011) (e.g. "Basal", "Complex", and "Robust") (**Appendix A**: Taxonomic attributes). Such a procedure allowed for biotic distances among samples to be quantified even when they had zero or very few species in common.

 The sample-specific data also required data preparation, which followed the biological data assessment. Longitude and depth are the two sample-specific variables considered to determine the environmental settings of the South African maritime domain (**Appendix A**: Sample-specific 195 abiotic data). For instance if a sample was recorded at a  $31^\circ$  longitude, then it was collected in the Indian Ocean and influenced by the Agulhas Current. Additionally, water mass temperature (T) and salinity (S) properties can be confirmed by depth, in which upper layer waters were identified between the surface and 500 m deep, intermediate between 500 and 1500 m, and abyss greater that 1500 m (Emery, 2003). Each abiotic parameter was firstly classified accordingly (see below), prior to running an one-way (unordered) similarity analysis (ANOSIM) test to evaluate variation of corals species within the longitudinal and depth groups (**Appendix A**: Sample-specific abiotic data). The ANOSIM test requires groupings and measures the mean of ranked dissimilarities between groups to the mean of ranked dissimilarities within groups. A standard approach was undertaken to investigate change in species attributes along the longitudinal and depth gradient, whereby a SIMPER analysis was run to evaluate contributing taxa.

 To classify the longitudinal data as a factor to be tested on the biological data, an auto select k-R cluster mean analysis was run on the normalized longitudinal data. A draftsman's plot was

produced to identify the number of longitudinal groups present and validate the cluster groups

210 present (**Appendix B**: Figure 2). The depth classification starts at a 50 m isobath and progresses

211 at 100 m increments. These depth readings were further classified according to zone, i.e., shallow

 (50-200 m) *vs* deep (300-1000 m). The depth values and subsequent depth zones were also added as a factor for testing. Similarity percentage (SIMPER) tests were run independently to distinguish coral species contributing to the longitudinal groups identified by the k-R mean analysis and depth groups. Sampling effort (i.e., denoted by N), species richness (denoted by S), Shannon index 216 (denoted by H'log<sup>e</sup>), and taxonomic distinctiveness (denoted by delta+) across the longitudinal and depth groups was quantified. The former was investigated by assigning coral records to 50x50 km grids created with the fishnet ArcGis function, whereby the grid size was guided by the boundary breaks of the k-R mean cluster groups.

 Subsequently, a RELATE routine was undertaken to evaluate if the combined longitude and depth spatial gradients correspond with those inferred from the coral species patterns. Here we used the Gamma + matrix in relation to the associated depth and longitude information, which was normalised into a Euclidean distance resemblance matrix. The RELATE technique calculated a Spearman's ρ rank correlation coefficient between all elements of the coral assemblage and environmental variable resemblance matrices, followed by a permutation test. Following this, a biota and/or environment matching (BEST) test was conducted to confirm which variable contributed the most to sample statistic given by the RELATE results. A species accumulation model was lastly produced to assess how well the current observed azooxanthellate stony coral data represents South Africa's predicted coral diversity.

**3. Results**

# **3.1. Longitudinal gradient**

234 The k-R (non-hierarchal) cluster analysis yielded two longitudinal groups  $(R = 0.94)$ , whereby group A encompasses samples from the eastern margin of South Africa and group B are samples from the southern and western margins (**Figure 3**). The SIMPER results suggested that biological samples represented by each of the two longitudinal clusters had an overall low average similarity in species composition (**Appendix C**: Long group SIMPER spp results). Communities in group A, with an average similarity of 3.4 %, were characterised by a total of 11 species from five families (i.e., Dendrophylliidae, Caryophylliidae, Micrabaciidae Vaughan, 1905, Turbinoliidae Milne-Edwards & Haime, 1848, and Flabellidae Bourne, 1905). Group B had a total of eight species contributing to the group identity with a slightly higher average similarity (4.6%), whereby only two families (i.e., Dendrophylliidae and Caryophylliidae) were represented.

 **Figure 3**. The 50x50 km gridded cells with samples in relation to the longitudinal groups defined by the k-R cluster analysis. Group A represents samples collected off the eastern margin and group B are samples collected off the southern and western margins.

 Furthermore, there was a distinction between families contributing the most to the cluster identities. Three dendrophylliids (*Balanophyllia capensis* Verrill, 1865)*, Pourtalopsammia togata*  (van der Horst, 1927), and *Heteropsammia cochlea* (Spengler, 1781) contributed the most to the similarity within group A; and the caryophylliids (*Trochocyathus* sp. 2 sensu Filander et al. (2021), *Caryophyllia stellula* Cairns, 1998*,* and *Desmophyllum dianthus* (Esper, 1794)) defined group B. When comparing the two groups, a 98% average dissimilarity was observed when investigating species contributing to a minimum cut off of 70%, whereby different species influenced group  distinction. *Heteropsammia cochlea*, *Letepsammia franki* (Owens, 1994)*, Flabellum pavoninum*  (Lesson, 1831)*,* and *Labyrinthocyathus delicatus* (Marenzeller, 1904) were represented only in group A. Among these species, *L. delicatus* is restricted to the Indian Ocean, whilst *L. franki, F. pavoninum* and *H. cochlea* are widely distributed in the Indo-Pacific. Four species contributed the most to group B, of which two are restricted to South African waters (*Ednapsammia columnapravia* Filander, 2021 and *Dendrophyllia* sp. 1 sensu Filander et al. (2021)), one considered cosmopolitan (*Desmophyllum pertusum* (Linnaeus, 1758))*,* and the remaining reported from the Atlantic and Indian ocean basins (*Enallopsammia pusilla* (Alcock, 1902)) (**Appendix C**: Long Group SIMPER spp results).

 Overall, the number of samples between the two groups varied, whereby group A (eastern margin) had over two times more samples than group B (southern and eastern margin) (**Table 1**). Contrary to this, the related area (i.e., number of grids) representing these samples was larger in group B than in group A (**Table 1**). Diversity followed the same pattern of higher measures in group A as compared with group B.

# **Table 1**. Summary of sampling effort in relation to longitudinal gradient



# **3.2. Depth gradients**

 A direct relationship between the number of samples (N), species richness (S), and depth was observed (**Figure 4**). The highest number of samples and observed species richness occurred between depths of 50 and 200 m, with the greatest species richness and sample count recorded at a depth of 50 m. The same two measures (S and N) fluctuated at the deep bathymetric zone (i.e., 300-1000 m). Within this deep zone, the highest coral diversity measures (S and N) were recorded at 1000 m and the lowest at 800 m. The overall pattern suggests that species richness is influenced by depth and sampling effort, but other factors may also play a role in shaping patterns of species diversity across different depths. Average taxonomic distinctiveness (denoted by delta +), which takes into account species phylogeny, did not show a clear pattern in coral diversity with depth and species diversity was relatively constant from 50 to 200 m. However, according to this measure, coral diversity was slightly higher at 1000 m despite the usage of a smaller number of samples from this depth (42 samples compared to 269 samples at 50 m). In other words, eight taxonomic families were recorded at 1000 m, while only seven were recorded at 50 m. In contrast, however, the conservative Shannon diversity index mirrored the pattern of species richness with depth (**Figure 4**).

 **Figure 4**. The relationship between number of coral samples (N) and species richness (S) in conjunction with the 292 average taxonomic distinctiveness (delta+) and Shannon diversity (H'log<sup>e</sup>) index measures across depth gradients. The x-axis shows samples represented in depth values in metres and the y-axis shows values that represent diversity 294 measures in arbitrary units. The inset shows the Shannon diversity  $(H'log^e)$  index repeated on a Y axis of 1-4.

 The SIMPER results of the coral species data according to family suggested that the caryophylliids, dendrophylliids, and flabellids were the main contributing taxa towards both the shallow (50-200 m) and deep (300-1000m) stations. Whilst all three families collectively contributed towards the  zone comparison (i.e., shallow *vs* deep) at a 70% cut, the Caryophylliidae representatives were more abundant in the deep stations compared with the Dendrophylliidae and the Flabellidae in the shallow stations (**Appendix C**: Depth zones SIMPER family results).

 When comparing the two bathymetric zones at lower taxonomical rankings, three species were exclusively recorded in the shallower zone (*Rhizopsammia compacta* Sheppard & Sheppard, 1991, *Truncatoflabellum inconstans* (Marenzeller, 1904), and *Dendrophyllia cornigera* (Lamarck, 1816)), whereas the deeper zones were characterized by three of the four species restricted to the longitudinal group B: *E. columnapravia, Dendrophyllia* sp. 1, and *E. pusilla.* 

# **3.2. The correlation of sample-specific variables (longitude and depth groups) to coral distribution patterns**

 The one-way ANOSIM results, which evaluated the rank differences in the coral pattern that may be explained by the longitudinal k-R clusters and depth groupings independently, showed that the 313 two longitudinal clusters (R=0.05, p = 0.001), and the eleven depth groupings (R = 0.072, p = 0.001) of the biological assemblages differed significantly from one another (**Appendix D**: Figure 315 5). Depth also showed significant differences ( $R = 0.105$ ,  $p = 0.001$ ) when grouped as shallow (50- 200 m) and deep zones (300-500 m) (**Appendix D**: Figure 6). Nonetheless, even though some overlap was observed in the coral structure when testing longitude and depth; the null hypothesis (of no differences in rank groups) can be rejected.

 The RELATE results showed a marginal correlation (Rho-value = 0.087) but a significant 321 difference (p-value  $= 0.001$ ) when comparing the coral patterns modelled by the Gamma+

 resemblance matrix to that of the Euclidean distance matrix (i.e., environmental variables - longitude and depth; **Figure 7**). It is important to note that the null hypothesis in the RELATE function is that there is no correlation. Thus, although the correlation is closer to zero (unexplained variance), the p-value confirms that longitude and depth are good predictors for the coral distribution patterns. The BEST results further confirmed that depth had an independent correlation value of 0.094, whilst both environmental parameters (longitude and depth) accounted for a correlation value of 0.097.

**Figure 7**. Simulated distribution/histogram of the test statistic Rho under the null hypothesis that there is no

correlation between the modelled coral patterns and that of the environmental variables Rho = 0.041.

 The majority of the species accumulation curves, which show how the number of species detected (i.e., observed or sampled) increases with increasing sampling effort (i.e., the number of individuals or samples collected), did not reach a plateau (**Figure 8**). All seven estimated curves, along with the observed or sampled species, started with a steep slope and indicated a rapid increase in the number of species observed with increasing sampling effort. Only two (MM and UGE) of the seven estimator curves followed the species observed pattern (Sobs), which appears to be levelling off as the sampling effort increases (**Figure 8**).

**Figure 8**. Species richness accumulation curve showing the species observed (Sobs= blue upright triangle) in

relation to five estimators (Chao 1= red downward triangle, Chao 2= green square, Jacknife 1= pink diamond,

Jacknife 2= blue circle, Bootstrap= grey cross). Two pairs of curves overlap, whereby the UGE estimator curve

follows the same pattern as the Sobs and the Chao1 has the same pattern as Chao 2.

## **4. DISCUSSION**

 The multivariate analyses suggests that the sample-specific associated data (e.g., longitude and depth) are significant predictor variables for azooxanthellate Scleractinia coral diversity. Nonetheless, unexplained variance exists. Diversity measures were assessed, in which the number of samples showed an inversely proportional relationship with species richness. Contrary to this observation, taxonomic distinctiveness (a diversity measure independent of the number of samples) revealed an opposing pattern to that of the univariate Shannon index measure. Thus, taxonomic distinctiveness accounted for the uneven species distribution across the South African continental maritime domain.

 An increasing species turnover along the west to east gradient was observed in our analysis. Such distributional patterns have long been reported for other South African marine invertebrates (e.g., Lang, 2012; Filander, 2014; Boonzaaier, 2017), suggesting that different oceanographic conditions are influencing the South African marine fauna. The accompanying current regimes may also govern these contrasting species profiles across the region. Thus, although the two longitudinal boundaries (Group A = eastern margin *vs* Group B= western margin) established by the k-R mean cluster analysis do not conform to the previously proposed oceanographic boundaries (Longhurst, 2007; Spalding et al., 2012;), the ANOSIM suggested a species pattern that may be explained by the two longitudinal groups. These margins correspond to varying oceanographic variables and currents, whereby the eastern margin (group A) is situated within the oligotrophic waters of the Indian oceanic basin and influenced by the western boundary Agulhas current. Interestingly, group B encompasses the southern and western margins located in both the Indian and Atlantic basins respectively. At the southern margin, the Agulhas current retroflects, moving away from the shelf,

 and introduces Indo-Pacific waters into the Atlantic Ocean, the latter being regulated by the northward flowing Benguela current (Shannon, 1985).

 The SIMPER results detailed a clear taxonomic/ family and species distinctions within these two longitudinal groups. Dendrophylliids contributed the most to Group A samples and caryophylliids to Group B. Additionally, the exclusivity in species found between Group A (*H. cochlea*, *L. franki, F. pavinonum,* and *L. delicatus*) and Group B (*E. columnapravia, Dendrophyllia* sp. 1*, D. pertusum,* and *E. pusilla*) (see longitudinal gradient results) corroborates with the proposal that species have a temperature threshold (Roberts et al., 2009; Cairns, 2007). The physiological characteristics of azooxanthellate coral species are indeed influenced by the properties of ambient water temperature (Gori et al., 2016; Castellan et al., 2019). For example, an *ex-situ* experiment undertaken on the reef-building corals *D. pertusum* and *Madropora oculata* revealed that they 381 respond differently when exposed to three temperatures (12, 9.0, and 6.0  $\degree$ C; Naumann et al., 2014). The respiration response rates varied; *M. oculata* declined whereas *D. pertusum* was not affected by temperatures being lowered. Two other physiological responses (i.e., calcification and dissolved organic carbon) were measured, and neither showed a consistent trend when comparing the two species. Thus, species belonging to different families or even congeners are expected to exhibit varying thermal tolerance.

 The recovered species longitudinal pattern of low sampling effort in Group A (eastern margin) but higher number of records and diversity observed herein, was particularly surprising as the western margin (which contributes to Group B) has a higher historic sampling effort (Griffiths et al., 2010). The greater presence of coral species in the eastern Agulhas region (Group A) may be explained

 by the heterogenous seabed substrate types provided by the increased abundance of mesophotic reefs, submarine canyons, and mosaic ecosystem types (Sink et al., 2019). Whilst the incising submarine canyons along the eastern continental margin (Green et al., 2007; Green, 2008; 2009; Green, 2011) may also give rise to a heterogenous environment, localised canyon substrate type studies need to be undertaken to confirm such hypothesis (Filander et al., 2022). Even though the Benguela Current in the South Atlantic (influencing the western passive margin) is unique in its interactions with the western boundary Agulhas current (Longhurst, 2007), this region has substrate predominately unconsolidated, resulting in a more homogenous environment (Dingle,1979; Cawthra et al., 2021; Filander et al., 2022). Additionally, dissolved oxygen levels have been proposed to affect scleractinian growth (Hanz et al., 2019) and the Southern Benguela Upwelling region does include a low-oxygen area/cell off St Helena Bay (Lamont et al., 2015). Though nearshore, such cells are reported to show spatial variability and may modify offshore upwelling water masses. These oxygen parameters superimposed with unconsolidated bottom 405 types and a slow current  $(< 3 \text{ m/s})$  may be a constraint for coral presence. The presence of coral is however influenced by multiple factors operating at different scales, and it is crucial to consider species-specific regional adaptation abilities to environmental gradients (i.e., dissolved oxygen) - even for cosmopolitan species (Orejas et al., 2021). Nonetheless, the prominence of anthropogenic activities that interact with the seabed in the Southern Benguela Upwelling area (Atkinson et al., 2011; Majiedt et al., 2019) cannot be overlooked and may also influence the low number of species records in the area.

 The southern margin, which contributes to Group B, is a unique area that exhibits minimal interaction with other landmasses and, as such, high endemism has been noted (Griffiths et al.,

 2010). In this region, the Agulhas current injects Indo-Pacific waters into the Atlantic, down to depths of 2000 m in the form of anticyclonic rings (Beal et al., 2011), before retroflecting eastwards towards the Southern Indian Ocean Gyre and the Antarctic circumpolar current (Spalding et al., 2012). Schouten et al. (2000) noted that the location of the retroflection loop is variable, but still within the southern region. Nonetheless, the Agulhas transport is estimated to 420 increase from 65 Sv ( $1Sv - 10^6$  m<sup>3</sup>s<sup>-1</sup>) at 32°S to 95 Sv at the southern tip of South Africa, as it breaks away from the shelf (Gordon et al., 1992; Duncombe Rae, 1991). Thus, the unpredictable behaviour and velocity of the Agulhas current make this area challenging for sampling and, therefore, the low number of records here may be attributed to limited sampling effort.

 The analysis of depth gradients allowed patterns of species richness in relation to the depth to be better understood. These results complement the longitudinal gradients whereby the univariant biodiversity measures peaked at 50 m, which corresponds to the accessible eastern margin of the South African maritime domain. In addition to the shelf being shallower (~ 50-150 m) and more accessible, the western boundary Agulhas current (characteristic of this area) has been linked to the highly diverse biological properties in the Southwest Indian Ocean, where eddies can trap and transport material over long distances (Halo et al., 2014). These complex oceanographic eddies can upwell deep nutrient-rich waters through surface divergence mechanisms (Halo et al., 2014), creating environments that favour the continuous inflow of potential food sources. Thus, these observations may provide grounds for a hypothesis to explain why azooxanthellate corals have a higher presence within this area.

437 The multivariate taxonomic average distinctiveness measure (denoted by delta +) showed diversity 438 to be highest at 1000 m, in which eight of the eleven known South African coral families are represented. This result aligns with the knowledge that the global azooxanthellate stony coral pattern (Cairns, 2007) has overall higher species diversity between the 200 and 1000 m. Irrespective, the SIMPER analysis distinguished three major families to contribute to bathymetric zone delineation. The deeper depths (300-1000 m) were characterized by caryophylliids and 443 flabellids, and the shallow zone (50-200 m) by dendrophylliids. These results conforms with the 444 known depth affiliations of these families, in which Dendrophylliidae species occurrence is reported to peak at shallower depths (50 to 300 m) (Cairns, 2001) and extant species of Caryophylliidae and Flabellidae are more prominent in the deeper waters (more than 200 m) (Kitahara, 2011).

 The two sample-specific (i.e., depth and longitude) data sets were applied in combination to extrapolate ocean basin properties (nutrient content, salinity, temperature, etc.), which characterise the oceanographic settings influencing South African marine fauna (i.e., the colder Benguela current along the western margin, and the warmer Agulhas current along the southern and eastern margin). In this context, the permutation models (ANOSIM and RELATE) imply that longitude and depth are good predictors for coral distribution patterns. However, the close to zero R-values (R<0.5) suggests a non-linear relationship even though significant variability is evident in the species composition within the factorial groups. Whilst depth is noted to be one of the main drivers for coral distribution (as shown by BEST results), it is important to recognize that this parameter encompasses several other properties, such as the Aragonite Saturation Horizon (ASH) that is the depth below which calcium carbonate becomes unstable and tends to dissolve (Jiang et al., 2015; 460 Guinotte et al., 2006). Such a zone has been estimated at 700-1500 m depth range south of  $\sim 20^{\circ}$ S (Jiang et al., 2015). Eight of the eleven known South African coral families are recorded within this depth range, suggesting these species are surviving within an aragonite saturation state. Interestingly, coral species have been previously reported to withstand saturating conditions. For example, a study undertaken in the Caribbean basin showed the depth of the aragonite saturation horizon to be strongly related to coral assemblage variation, whereby *M. oculata* and *S. variabilis* occur in patchy distributions at or above the saturation zone (Auscavitch et al., 2020). The response of coral species to water properties, such as the ASH, are in no way consistent, highlighting the need for further research to comprehend the underlying environmental drivers of coral distribution.

 Although the azooxanthellate coral data reported herein represent an accumulation of samples over 30 years and are the best available representation of the South African fauna, all species richness estimator models did not plateau, demonstrating that the area is still not well sampled and may be much more diverse than currently known. The shape and slope of the curve typically provides information on the species richness, evenness, and heterogeneity of the community being sampled, as well as the adequacy of the sampling effort. In other words, if the species accumulation curve keeps increasing with additional sampling effort then full extent of species diversity within the study area has not yet been captured. Without a doubt, additional sampling coverage will provide clearer conclusions on national coral diversity trends.

# **4. Conclusion and recommendations**

 This study examined the best available data for the South African azooxanthellate coral fauna and presented a pre-processing methodology that caters for historical samples. Differences in  azooxanthellate coral species across South Africa's diverse and dynamic oceanographic conditions were revealed, whereby species turnover increased on a west to east axis. A species depth gradient was additionally observed, in which the multivariate diversity measure complemented the existing knowledge on taxa trends. These patterns were evident despite the data limitations related to museum samples. Whilst the lack of museum associated abiotic data still exists, the methodology to standardize co-ordinate information and depth may be considered in other data sets with similar attributes to inform further research to elucidate diversity patterns. In general, historical collections (which represent years of sampling effort) provide a valuable biological data source but require thorough validation. Despite the sparsity and unbalanced nature of the data, knowledge has been advanced and sampling gaps identified. A purposeful application for this existing coral data set will be its integration into multi-taxa biogeography analyses that will support ecosystem description and delineation. The data set will also be valuable for spatial prioritisation and marine spatial planning, particularly alongside taxa that share similar abiotic requirements.

# **Acknowledgements**

 A sincere acknowledgement goes to Dr David Herbert (Department of Natural Sciences, National Museum Wales) who assisted with associated station data for the *Meiring Naude* and *Sardinops* collections. Dr Carl Palmer (Alliance of Collaboration in Climate and Earth System Science) is thanked for proof-reading the first manuscript draft, and Dr Victoria Goodall (Nelson Mandela University) for reviewing the data analysis section. Mr Ashley Johnson and Dr Lauren Williams (Department of Fisheries, Forestry, and the Environment) provided words of encouragement and ArcGIS technical support; respectively.



















































Click here to access/download Supporting Information [Appendix D\\_Figure 5. TOP.tif](https://www.editorialmanager.com/pone/download.aspx?id=33425026&guid=0882da1a-624e-4e97-b1a3-71970a504cdc&scheme=1)

Click here to access/download Supporting Information [Appendix D\\_Figure 5. Bottom.tif](https://www.editorialmanager.com/pone/download.aspx?id=33425027&guid=974ae4d8-0b5a-4d2c-a16b-392caa81587a&scheme=1)

Click here to access/download Supporting Information [Appendix B\\_Figure 2.tif](https://www.editorialmanager.com/pone/download.aspx?id=33425007&guid=767f5736-e9e0-49b2-b633-3c410a90807e&scheme=1)

Click here to access/download Supporting Information [Appendix D\\_Figure 6.tif](https://www.editorialmanager.com/pone/download.aspx?id=33425028&guid=caa62a46-b4b0-40eb-9b78-5be8e4d6bd44&scheme=1)

Click here to access/download Supporting Information [Appendix A\\_biopaper\\_plos paper.xlsx](https://www.editorialmanager.com/pone/download.aspx?id=33425039&guid=c7c0416a-26c6-413d-9e36-68cd89606c18&scheme=1)

excel

Click here to access/download Supporting Information [Appendix C\\_biopaper\\_plos paper.xlsx](https://www.editorialmanager.com/pone/download.aspx?id=33425041&guid=a80726e3-13e8-4fee-8c09-da917355ec60&scheme=1)