REEF FISH CITIZEN SCIENCE DATA ASSESSMENT

Presented To The Regional Report Card Network

AU213005077.001 R81946

Citizen Science Fish Data Assessment

1

23 February 2023

rpsgroup.com

This review of Reef Fish Citizen Science Data is funded by the partnership between the Australian Government's Reef Trust and the Great Barrier Reef Foundation.

Document status

Version	Purpose of document	Authored by	Reviewed by	Approved by	Review date
0	Citizen Science Fish Data	G.Vinall	G.Vinall	G.Vinall	23/2/2023
1	Citizen Science Fish Data	G.Vinal	G.Vinal	G.Vinal	9/6/2023

Approval for issue

Greg Vinall		9/6/23

Prepared by:

RPS

Greg Vinall
Principal
Scientist

135 Abbott Street
Cairns QLD 4870

T +61 7 4031 1336
E Greg.vinall@rpsgroup.com.au

Prepared for:

Reef Catchments Pty Ltd

Brie Sherow
Technical officer

1/85 Gordon St
Mackay QLD 4740

T +61 428 955 490
E brie.sherow@healthyriverstoreef.org.au

Contents

TABLES

FIGURES

This review of Reef Fish Citizen Science Data is funded by the partnership between the Australian Government's Reef Trust and the Great Barrier Reef Foundation.

1 Introduction And Background

The GBR Regional Report Card Network (RRCN) consists of five regional report card partnerships (RRCP): Fitzroy Partnership for River Health, Gladstone Healthy Harbour Partnership (GHHP), Healthy Rivers to Reef Partnership (Mackay-Whitsunday-Isaac) (HRRP), Dry Tropics Partnership For Healthy Waters (DTPHW), and Wet Tropics Waterways (WTW). These partnerships publish annual Regional Report Cards assessing the condition of freshwater, estuarine, and marine systems within their respective regions.

The RRCP's rely on data collected by partner organizations to undertake the analyses presented in their report cards, although challenges exist when data is unavailable for certain indicators. One such challenge has been the acquisition of coral reef fish data to contribute to the assessment of inshore and offshore marine zones. Currently, these assessments rely on coral health and water quality data only, despite a fish metric having been identified at the inception of the report card program as being key to the understanding of reef health.

This project assesses the potential for citizen science data to be utilised in the assessment of inshore and offshore reef condition, the ultimate goal being the development of a coral reef fish metric that (combined with coral health and water quality data) would increase the robustness of reef health assessments through an additional line of evidence.

This report follows a previous, broader assessment of citizen science fish data (RPS, 2022), which identified three data sets as potentially having data of sufficient robustness and resolution to enable indicator development. These included data collected by the following:

- The Reef Life Survey program administered by the Institute of Marine and Antarctic Studies (University of Tasmania).
- The Great Barrier Reef Marine Park Authority "Eye on the Reef" tourism and rapid assessment programs
- The ReefCheck program, a global fish monitoring citizen science program.

Numerous other fish datasets were considered in the initial data assessment undertaken by RPS but were found to be unsuitable for use in indicator development due to various combinations of low frequency sampling, low resolution taxonomy, limited species monitored and/or poor or undocumented sampling rigour and/or quality control.

A key feature of the three datasets assessed herein is strong similarities between sampling protocols, which created the potential for integration of data to increase the spatial and temporal resolution of data. This is important in the context of a complex reef system as it will increase the ability to detect important observations over the natural data variability.

2 Assessment of three community science reef fish datasets

Each of the three data sets was first examined for obvious factors that might limit their use individually or inhibit integration with the other data sets. **Table 1** summarises these results, with further detail provided below.

2.1 Overview of data set attributes

Table 1: Overview of the three-citizen science coral reef fish data sets for sites within the Great Barrier Reef footprint, as relevant to regional report cards.

	Taxa	Duration (yr)	Site consistency	Comment
RLS Data	689	8	No	Comprehensive long-term species-level data set with good spatial and temporal resolution.
GBRMPA	51	25	Unknown*	Long-term genus level data set of key indicator taxa with spatial and temporal deidentification.
Reef Check	11	23	Unknown*	Long term data set recording of a smaller number of key indicator taxa

* The degree of site consistency for GBRMPA data sets is not known as the monitoring locations were deidentified to report card regions. Site consistency is also unknown for the Reef Check program as the data set provided for this assessment did not include precise monitoring locations.

2.1.1 Reef Life Surveys

RLS Data is the most comprehensive and usable of the three datasets. This data set extends back to 2010 but the frequency of data collection data collection and archiving has increased significantly since 2015. The RLS dataset also has supporting parameters in many cases (eg water quality, invertebrate data) that could assist with reef health assessments.

This data set is also the most readily available of the three, being downloadable on demand via a data portal and it's usage being covered by a creative commons licence.

The following limitations regarding the RLS data set were noted:

- Monitoring sites are not consistent between years. The locations surveyed appear to be chosen on an ad hoc basis, presumably opportunistically piggy backing other programs.
- The spatial range of data varies between years but in any given year does not typically include sites within all regions/reporting zones.
- Across all regions and zones the total taxa count is 689. At each survey site two blocks and two depths are sampled. These were summed for the purpose of this assessment.

2.1.2 GBRMPA Eye On The Reef

This data set extends back to 1997. Two monitoring protocols are used: Rapid Assessment and Tourism. The former involves a recording observation of a smaller range of indicator species and was insufficiently detailed for use in developing an indicator.

The Tourism data set offers the best prospects for assessing fish health at a reporting zone level. Within this set, coral trout are differentiated by size (above and below the minimum size limit for harvest) and Mauri wrasse are split by sex. There is a focus on species of tourism value, such as Maori wrasse and Queensland grouper

The GBRMPA Tourism dataset (as provided) has the following limitations:

- Limited to 49 taxa, resulting in greatly reduced taxa site averages when compared with the RLS dataset.
- Data are de-identified from any site information, making it impossible to determine whether replicates have been undertaken or to stratify data by factors (eg green vs general use zones or distance from coastline/rivers).
- Data were predominantly from the Wet Tropics Region, with smaller data sets available for other regions.
- Due to data deidentification, there is no indication whether the data provided are from a few sites sampled on different days or multiple sites sampled less frequently.

2.1.3 Reef Check Australia

The Reef Check data set spans 23 years. However, the data collected are greatly restricted when compared with the other data sets, with only 13 target taxa recorded, primarily iconic species. Like the GBRMPA data set, there is some stratification by fish size for coral trout, and also for Queensland grouper. Taxa are recorded by common name and there was some uncertainty in a number of instances as to the correct identification of taxa. Some taxa appear to be recorded at the family level only.

This data set provides only the common name of the reef systems surveyed, rather than latitude and longitude. As some systems are very large and diverse the ability to stratify data appropriately is lost.

The sampling protocol entails four replicates at each site, which have been summed for this assessment.

The Reef Check dataset (as provided) has the following limitations:

- Limited to 11 taxa, resulting in greatly reduced taxa site averages when compared with the RLS dataset.
- Location data were provided only as the generic names of reef systems, with many of the reefs being extensive and known to support a diversity of habitats. It is understood that higher resolution location data are available in the Reef Check database.
- Some uncertainty regarding taxonomy of fish recorded due to use of generic, rather than scientific names.

2.2 Potential to combine citizen science data sets

The ability to perform rigorous analysis and develop a robust indicator increases with the size of the underlying data set. Greater replication and spatial spread have the effect of smoothing out outlying or unusual observations.

An opportunity identified early in this assessment was the potential to combine the three data sets into a single, much larger data set. It was hoped that replication at some sites and/or an increase in the overall number of sites would bolster the analysis.

The potential for this approach was indicated by similar sampling protocols across the GBRMPA, Reef Check and RLS programs, including significant overlap of the indicator taxa used in the former two programs. It was recognised that integration of the datasets could only be useful if the "lowest common denominator" was sufficient resolution to draw conclusions. In other words, the combined data would assume the spatial and taxonomic resolution of the least detailed data set.

On further assessment it was found that whilst there were similarities in the data sets there were also significant differences that proved problematic when integrating them. These included insufficient overlap between the species of fish monitored by the GBRMPA and Reef Check programs (a table highlighting the overlap is provided in **Appendix A**). The restriction of fish species collected by two of the data sets (GBRMPA and Reef Check) makes combining data difficult and doing so is detrimental to the ability to detect differences amongst Regions/zones due to the variance created by the different datasets.

It was ultimately determined that integrating the data sets would result in losses of spatial, temporal and taxonomic resolution as well as the inability to stratify data appropriately for analysis. The result would be a data set that is not sufficiently detailed to enable a meaningful analysis.

It was concluded the RLS data set, used in isolation offered the best opportunity to develop a useful fish health assessment. This has been explored further below.

However, the RLS data set has it's own limitations as there is little consistency of sampling locations over time with some regions or reporting zones not being surveyed in some years. In other instances the number of monitoring events is relatively low, reducing the usefulness of the data set (low n). Key statistics for recent years are summarised in **Table 2.**

2.3 Other Data Set Considerations

2.3.1 Accessibility of data

Access to the GBRMPA and Reef Check data sets is via arrangement with the individual organisations. If these data sets were to be used on an ongoing basis by the report card network it would be necessary to negotiate ongoing data usage agreements and arrangements to ensure the data were received annually in time for report card preparation. Similar arrangements exist for other data sets and the data acquisition process is likely to be relatively straightforward.

The RLS data are the most accessible, being publicly available under creative commons license via an online portal. Care must be taken to ensure that its usage complies with the terms of the creative commons agreement.

Table 2: Summary of Regions, Zones and years of RLS surveys and the number of sites within each year.

Region	Zone	years	# sites	Taxa count	Avg site taxa count	Taxa (sum) abundance	Avg site Abundance
DT	Black	2015	1	66	66	11838	11838
		2016	2	70	56.5	32927	16463.5
	Offshore	2016	11	274	112.8	192706	17518.7
		2017	6	215	103.8	31961	5326.8
		2020	2	107	73.5	5811	2905.5
MWI	Central	2018	10	148	51.1	16479	1647.9
		2019	3	61	35	5182	1727.3
		2020	3	73	41.7	6580	2193.3
	North	2010	1	48	48	1255	1255
	Offshore	2015	4	139	75.25	76087	19021.75
		2016	5	179	82.2	23970	4794
		2017	4	225	107.5	55235	13808.75
		2018	10	231	75.5	24565	2456.5
		2019	6	178	51.25	24258	3032.25
		2020	2	114	73.5	5377	2688.5
	Whitsunday	2010	4	107	56.25	15943	1566
		2015	1	26	26	1022	1022
		2016	3	122	73.3	11380	3793.3
		2018	10	195	63.2	23078	2098
		2019	4	113	53.75	7043	1266.5
		2020	7	160	58.6	16814	2101.75
		2021	5	141	72	8577	1715.4
WT	Central	2010	5	148	56.6	27511	5502.2
		2015	3	97	42.5	5004	1251
		2016	4	150	52.5	93427	13560.75
		2018	1	62	62	1243	1243
		2020	2	71	51	4583	2291.5
		2021	1	51	51	1608	1608
	North	2015	1	33	33	4399	4399
		2016	6	235	92	39777	6629.5
	Offshore	2010	11	360	127.1	64132	5830.2
		2015	10	258	84.6	92777	9277.7
		2016	9	295	112.1	85704	9522.7
		2017	2	130	80	10690	5345
		2020	6	195	82.2	8383	1397.2
		2021	6	211	98	13254	2209
	South	2010	4	126	60.75	55550	13887.5
		2020	3	75	40.3	7772	2590.7

2.3.2 Monitoring program longevity

A critical consideration for report cards is that indicators are developed using data sets that will continue to be collected for some time into the future, since the discontinuation of a data set would represent not only wasted effort but challenges in the calculation, reporting an interpretation of report card scores when the underlying indicator mix is altered.

In this regard, Reef Check is a not for profit that is approximately 70% Government funded with the remaining 30% coming from sale of goods/services and from private donations. Whilst no program of this type is guaranteed ongoing funding, Reef Check has been operating for 23 years and in the current climate of concern for reef health and anticipated Government (Commonwealth and Qld) investment in citizen science it is likely that the Government portion of Reef Check funding will be ongoing in the near to medium future. Similarly, the GBRMPA "Eye on the Reef" program, under which the Tourism and Rapid fish monitoring programs are run, is likely to receive ongoing funding in the current political climate.

Reef Life Surveys is a not for profit that receives funding from a variety of Government and philanthropic sources. At the time of preparing this report their funding for 2023 is uncertain, raising questions about the availability of data for future report card needs.

2.3.3 Automation of data preparation and analysis

In the format received, each of the data sets required significant cleaning and sorting prior to analysis using the statistical package.

For the data to be useful to the report card partnerships, given current levels of resourcing would require an efficient process for "cleaning", sorting and filtering data prior to integration and/or data analysis. It was found that this could be efficiently achieved by developing an excel spreadsheet with embedded algorithms and formulas to automate the data preparation process.

3 Data Preparation And Analysis

3.1 Data Acquisition And Licensing

Reef Life Survey data are available for public download and use under creative commons license from the "Open Access to Ocean Data" (OAOD) online portal https://portal.aodn.org.au/search.

GBRMPA Eye On The Reef data were provided under a data agreement negotiated with GBRMPA.

Reef Check data can be accessed from the Reef Check Australia website https://www.reefcheckaustralia.org/ However, due to the volume and nature of data used for this analysis, a data agreement was negotiated and Reef Check Australia facilitated the data extraction.

3.2 Data Preparation And Transformation

All data sets were received as comma separated value (.csv) spreadsheets. These required the following actions prior to statistical analysis:

- Cleaning to remove data from survey sites that were outside of the temporal and spatial ranges being assessed, and to remove duplicate and/or incomplete data.
- Organisation of data into a consistent format across all data sets. To facilitate importing into the Primer software, data were arranged with variables in columns and samples in rows.
- Unique identifiers were assigned to each observation and data were grouped by year, report card region and reporting zone.

An excel spreadsheet was developed to automate this procedure, and particularly the processing of the RLS data set, which is very intensive and requires considerable preparation prior to analysis.

Due to frequent outliers, skewed data and complex data distribution, transformation was often applied prior to statistical analysis. This included:

- Logarithmic transformation, was used to stabilise the variance of data with highly skewed distributions (over several orders of magnitude). This transformation compresses the scale of the data and improves statistical analysis.
- Square Root Transformation was used where the data had a positive skew or the variance increased with the mean, particularly where count (as opposed to relative) data were of interest. It reduces the magnitude of extreme values and improves the symmetry and linearity of the data.
- Power transformation was used where the data distribution is unknown or didn't conform to logarithmic or square root transformations.

3.3 Data analysis

3.3.1 Descriptive statistics

Univariate statistics were used to characterise specific variables such as mean, median, range, and standard deviation and to provide insights into the central tendency, spread, and distribution of variables related to fish populations, such as abundance, biomass, size, or species richness.

Descriptive statistics can be visualized in a range of different ways, but in this assessment are largely presented as “Box and Whisker” plots. This style of presentation is effective for simultaneously displaying a range of variables in a relatively easy to interpret format.

3.3.2 Multivariate statistics and visualisations

Multivariate statistics were utilised to examine the relationships between multiple variables simultaneously. Bray-Curtis similarity analysis was used to explore the complex interactions and patterns within the dataset and works by calculating the compositional similarity between samples by comparing the relative abundances of species present.

Permanova analysis has also been utilised to further explore Bray-Curtis similarity. This analysis checks for different species compositions or abundances between sites or sampling events and helps to understand what factors might be causing those differences. It’s useful in determining whether differences are due to factors other than chance.

BIOENV analysis is used to determine the subset of environmental variables that best correlates with the similarities or dissimilarities in species composition data. It was used in this assessment to identify the reef fish families that contributed most to similarities between sites. This analysis can provide insights into the taxonomic groups that may be critical for maintaining healthy and diverse fish populations.

Numerous visualisations of Bray-Curtis outputs are possible, but non-metric multidimensional scaling (nMDS) plots are one of the more common visual tools and have been used in this assessment. nMDS ploys enable a quick visual assessment of the similarity between data sets, groups or sites, with those exhibiting similar taxonomic composition grouping close together on the nMDS plot. Those that are dissimilar group further apart.

Bubble plots have also been used to visualize data where it was beneficial to include additional information that couldn’t be displayed on an nMDS plot. For example, bubble plots have been used to overlay the proportions of various taxonomic groups onto the similarity/dissimilarity matrix.

4 Exploring The Data Set For Potential Indicators

The following provides details of several analyses that have been undertaken in an effort to identify potential options to utilise the available citizen science coral reef fish data. These are considered preliminary assessments to determine whether further exploration of techniques is warranted, rather than exhaustive analyses.

Note that these analyses utilise the RLS dataset only, except where noted. In some instances the GBRMPA and Reef Check data were also analysed to assess whether data based on indicator species may also be included in the analysis. The metrics discussed in this section are summarised in **Appendix B**

4.1 Observed vs Expected Taxa Richness (Unit Effort)

A useful fish indicator for regional report card purposes could be based around the number of species recorded within a reporting zone as a function of the monitoring effort expended.

Species accumulation curves created from long term data provide the number of taxa expected as a function of sampling effort and are shown for the MWI region and WT Region offshore zones in **Figures 1 & 2**. Results have been extrapolated over the remaining regions and zones.

The analysis suggests that a minimum of 4-6 sampling events would be required annually in each of the zones to develop an indicator using this approach. Obviously, a larger number of sampling events is preferrable due to the taxonomic diversity within these systems. In the long-term data set this minimum is typically met in the offshore zones and less frequently on the inshore reef systems.

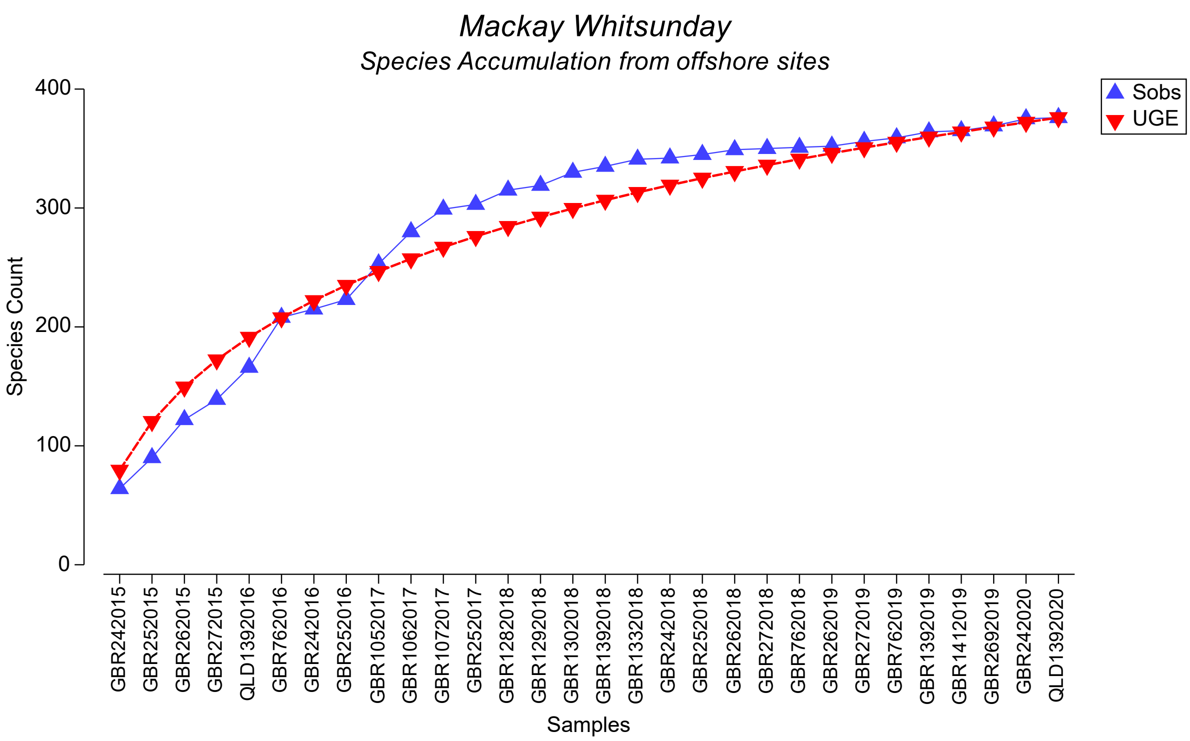

Figure 1: Species accumulation curve for the Mackay Whitsunday Isaac region (RLS data). Sobs indicates actual taxa based on the order in which the sites appear in the database. UGE is a modelled smoothing of the curve.

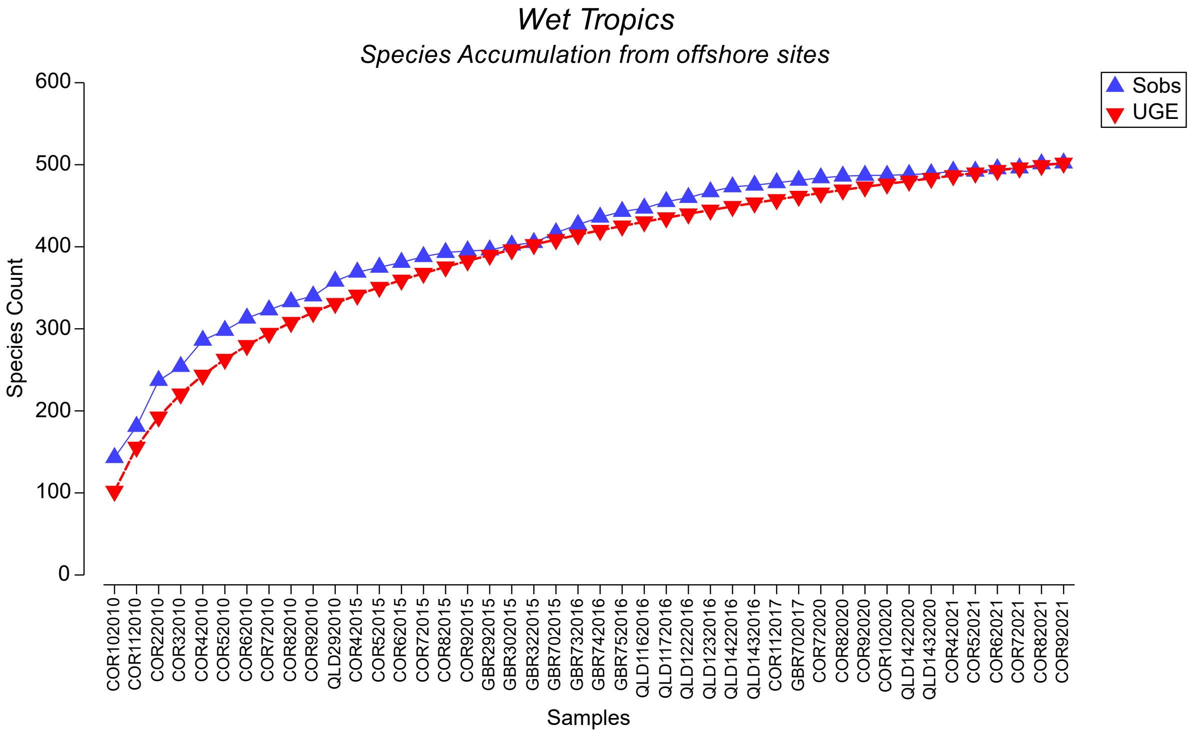

Figure 2: Species accumulation curve for the Wet Tropics region. Sobs indicates actual taxa based on the order in which the sites appear in the database. UGE is a modelled smoothing of the curve.

The analysis indicates that based on the number of sampling events (**Table 2**):

- **Mackay, Whitsundays, Isaac:** Insufficient monitoring was conducted at offshore sites during 2020 and 2021 to assess whether the taxonomic richness was commensurate with sampling effort, although sufficient data may be available from the Whitsunday's zone in each of those years.
- **Dry Tropics:** Insufficient monitoring was conducted during 2020 and 2021 to assess whether the taxonomic richness was commensurate with sampling effort in any zone.
- **Wet Tropics:** The offshore zone has barely sufficient monitoring effort in 2020 and 2021 to enable an assessment of expected species as a function of monitoring effort. Other zones had insufficient monitoring effort.

Based on the 2020 and 2021 data sets it would appear that the RLS program would only be sporadically useful for developing an indicator based on taxa richness as a function of monitoring effort.

4.2 Observed Vs Expected Taxa Richness (Reference Comparison)

Consideration was given to developing a fish indicator based on observed taxa richness/abundance at monitoring sites versus those observed at reference sites/zones.

A major challenge with this approach is identifying relatively unimpacted sites that might qualify as reference sites. In addition, these must be consistently surveyed by one or more of the programs to facilitate annual assessment by reporting zone. None of the programs assessed herein currently monitors any site consistently on an annual basis.

4.2.1 Spatial community shifts

Figure 3 shows non-metric multidimensional scaling (nMDS) of fish community data. The analysis shows some separation between offshore sites (clustering towards the bottom right) and inshore sites. This indicates shifts in community structure at fine longitudinal scales.

A north-south gradient between regions is also apparent, particularly when the offshore and inshore data are viewed in isolation (**Figures 4-5**). Closer examination shows a north-south gradient also occurs within the WT and MWI regions (there were insufficient data to check DT), which indicates changes in community structure on fine spatial scales.

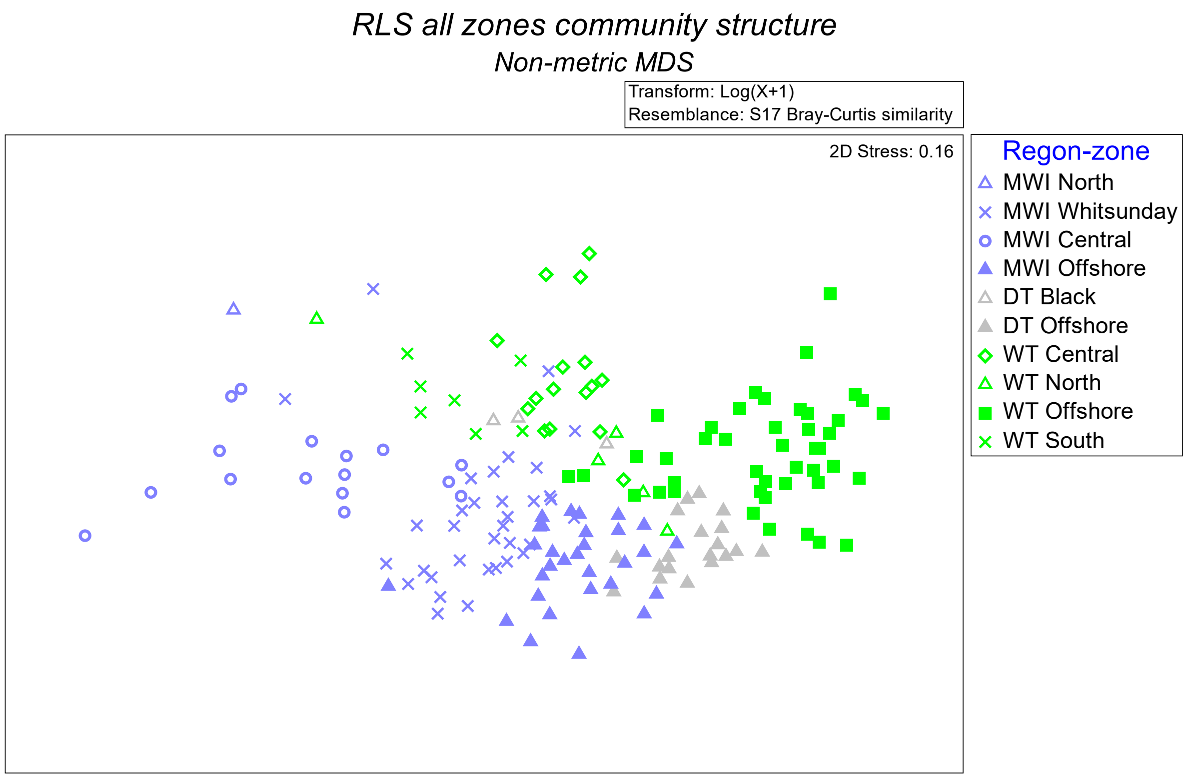

Figure 3: nMDS plot of all data from the RLS data set

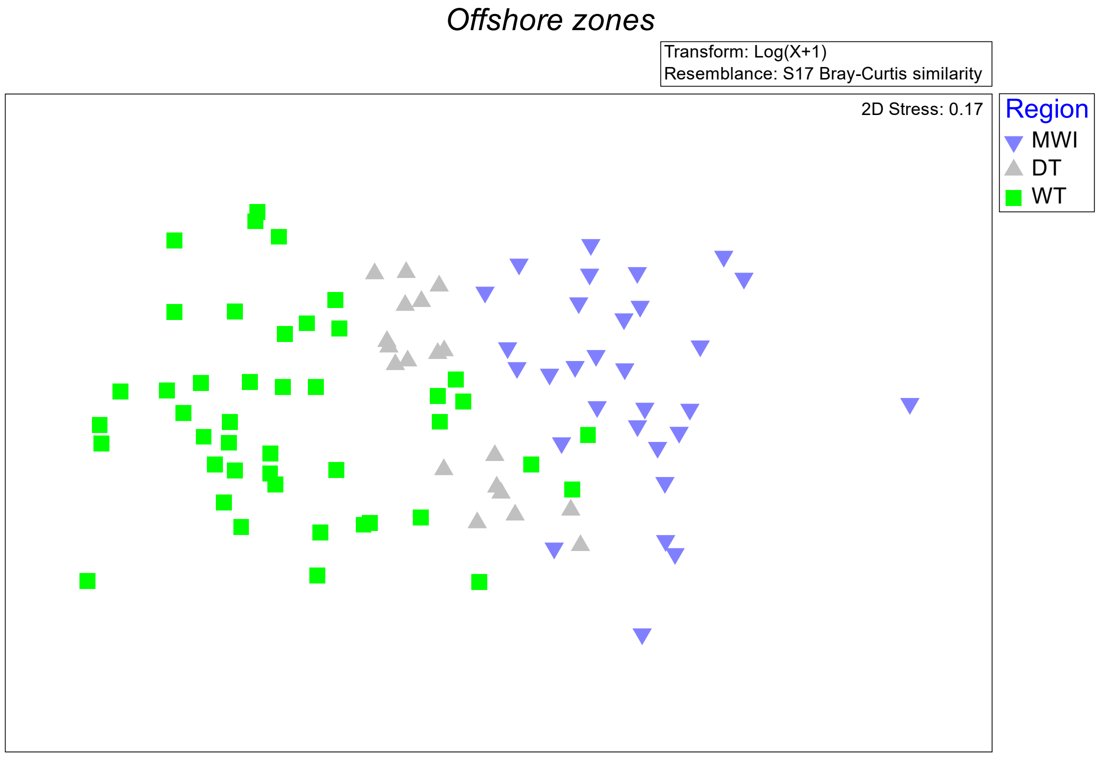

Figure 4: nMDS plot of offshore zone data set for all regions and years

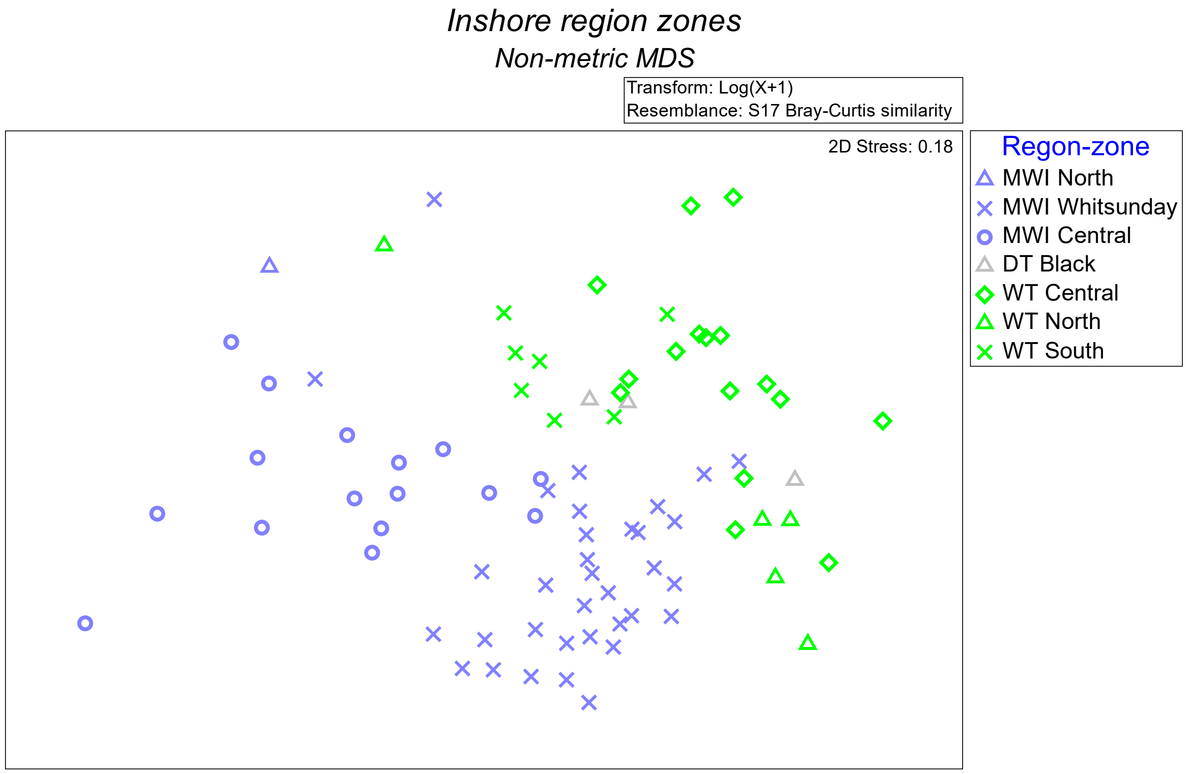

Figure 5: nMDS plot of inshore data for all reporting zones and regions

The identification of reference sites within each report card reporting zone was further hindered by the limited number of sites sampled and the lack of consistency regarding sampling frequency.

4.2.2 Relevance of offshore sites as a benchmark for inshore sites

In the absence of useful inshore reference sites, the potential to use taxa richness and abundance at offshore sites as a benchmark for assessing inshore reefs was explored. Many assumptions underpin this approach, including:

- That offshore sites are in relatively undisturbed condition.
- That the inshore zones would have supported similar communities to the offshore zones prior to anthropogenic disturbance.

These assumptions are recognised as being significant. For example, offshore sites are known to be unevenly impacted by fishing pressure and/or coral bleaching and may not be in "undisturbed" condition. Likewise, inshore sites are frequently pressured by turbidity, nutrients and pollutants such as pesticides. These different influences between reporting zones are likely to shift coral reef fish communities in different ways.

Further, the nMDS results shown in **Figures 3-5** show that there are differences in community structure between inshore and offshore zones that might by naturally occurring or might reflect decades of anthropogenic impacts.

Permanova of taxa richness data (RLS dataset, **Figure 6**) further confirms that there are significant differences between regions, as was identified from the nMDS plot with taxa richness changing along a north-south gradient.

PERMANOVA RESULTS FOR RLS Taxa data presented above.

This is equivalent to a ANOVA as I have used Euclidean distance resemblance. If you want to use any of the significance data let me know ill tidy it up and put meaningful words to it!

Taxa richness

Resemblance: D1 Euclidean distance

Sums of squares type: Type III (partial)
Fixed effects sum to zero for mixed terms
Permutation method: Unrestricted permutation of raw data
Number of permutations: 999

WET TROPICS REPORT CARD AREAS

PERMANOVA table of results

Source	df	SS	MS	Pseudo-F	P(perm)	Unique perms
Zo	3	31052	10351	21.384	0.001	999
Res	68	32915	484.04			
Total	71	63967				

Pairwise tests

Groups	t	P(perm)	Unique perms
Central, North	3.8237	0.004	226
Central, Offshore	6.8503	0.001	360
Central, South	1.4	0.21	150
North, Offshore	0.0038533	1	235
North, South	3.1011	0.014	215
Offshore, South	5.5243	0.001	288

DRY TROPICS REPORT CARD AREAS

PERMANOVA table of results

Source	df	SS	MS	Pseudo-F	P(perm)	Unique perms	P(MC)
Zo	1	5524.3	5524.3	21.778	0.001	162	0.001
Res	20	5073.2	253.66				
Total	21	10597					

MACKAY WHITSUNDAYS REPORT CARD AREA

PERMANOVA table of results

Source	df	SS	MS	Pseudo-F	P(perm)	Unique perms
Zo	3	12041	4013.8	15.514	0.001	996
Res	77	19921	258.71			

PAIR-WISE TESTS

Groups	t	P(perm)	Unique perms	P(MC)
Central, North	0.10746	0.758	14	0.899
Central, Offshore	6.1759	0.001	313	0.001
Central, Whitsunday	4.2324	0.001	245	0.001
North, Offshore	1.6872	0.139	24	0.118
North, Whitsunday	1.1775	0.21	25	0.251
Offshore, Whitsunday	3.5775	0.001	288	0.001

Figure 6: Permanova results for full RLS data set.

Despite these acknowledged limitations, with no potential inshore no inshore reference site the validity of using of offshore site as a benchmark was nonetheless explored.

A univariate approach displaying the data as boxplots assisted with understanding the taxa richness and total abundance of the full RLS data set (**Figures 7-8**).

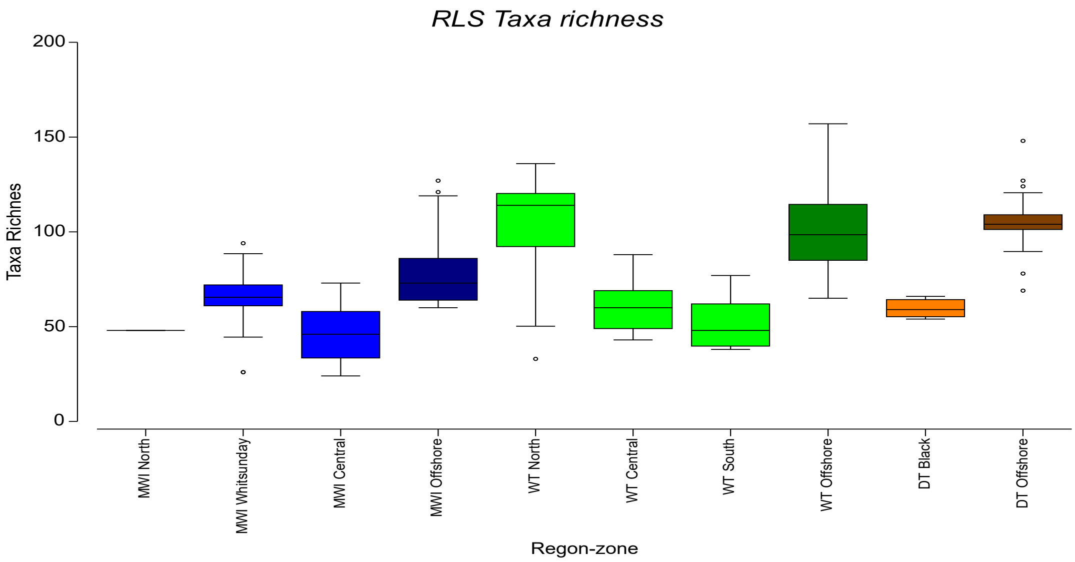

Figure 7 Boxplots of RLS data set taxa richness, all regions and reporting zones.

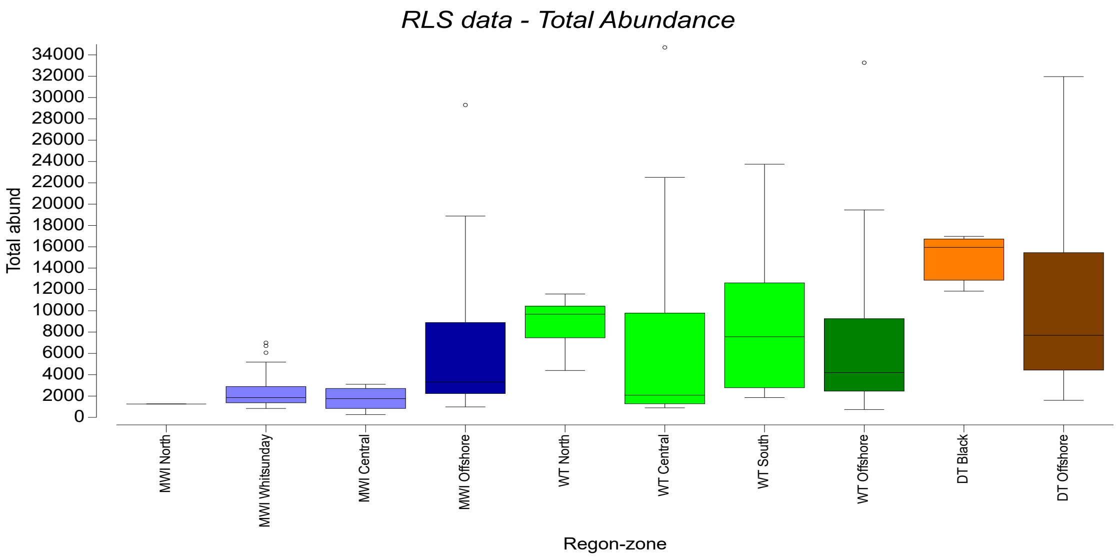

Figure 8: Boxplots of RLS data set total abundance, all regions and reporting zones

Taxa richness in the MWI region was similar between the Whitsunday Inshore and Offshore Zones and in the WT Region between the Northern Inshore and Offshore Zones (**Figure 7**). This demonstrates that inshore reefs can exhibit similar taxa richness to offshore reefs, despite the nMDS indicating differences in community structure (**Figure 4**).

All other inshore survey sites had lower taxonomic diversity than the corresponding offshore sites, which is consistent with the previously discussed observations of community shifts over short spatial distances. Again, it is not known whether this is due to anthropogenic influences or is simply a natural occurrence associated with structure and habitat.

In terms of total abundance (**Figure 8**) the RLS data shows the offshore zone in the MWI region to have greater abundance than the inshore zones. In the wet tropics total abundance was comparable between all zones and in the dry tropics the Black Inshore zone had higher total abundance than the offshore zone.

Whilst the RLS data set has previously been identified as the most useful, the GBRMPA and Reef Check data were also plotted as boxplots, largely to visually confirm the suspected loss of analytical resolution resulting from the limited species recorded in these programs (**Figures 9-12**).

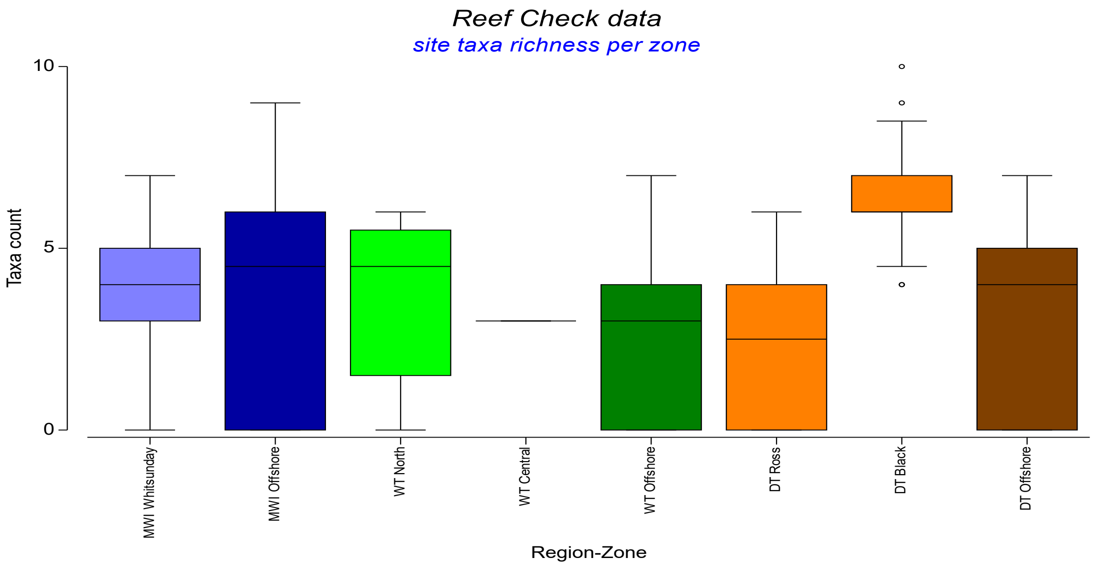

Figure 9**:** Boxplots of Reef Check data set taxa richness, all regions and reporting zones.

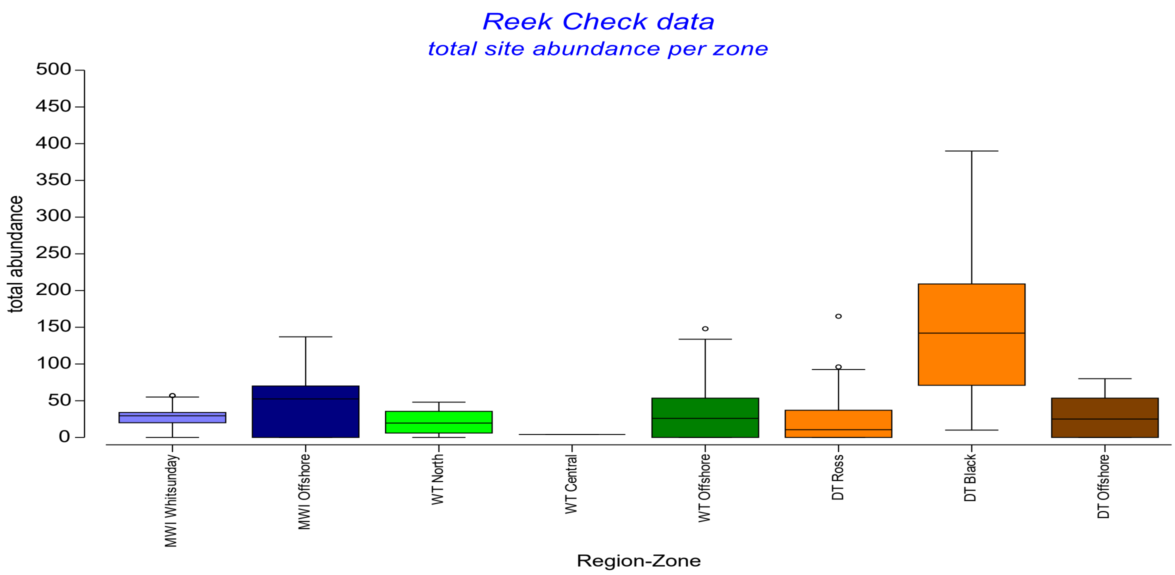

Figure 10: Boxplots of Reef Check data set total abundance, all regions and reporting zones.

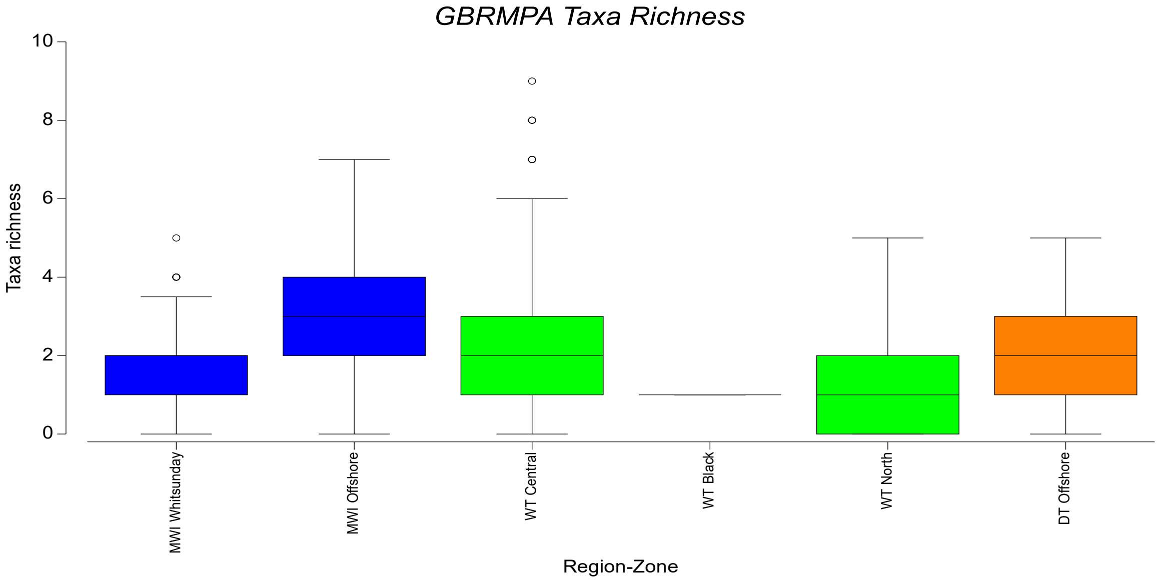

Figure 11: Boxplots of GBRMPA data set taxa richness, all regions and reporting zones.

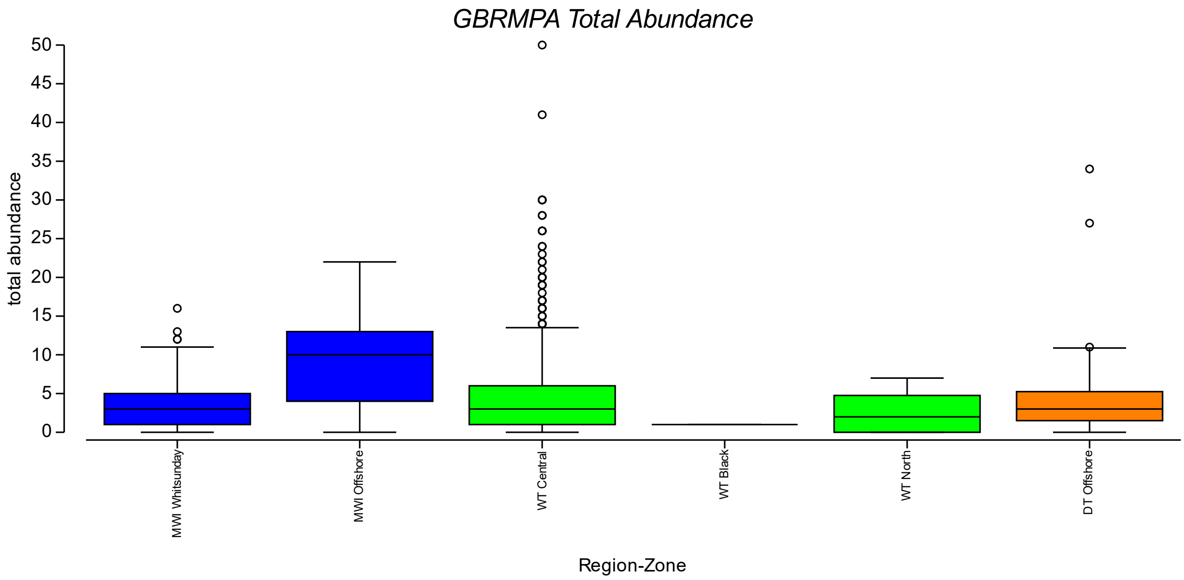

Figure 12: Boxplots of GBRMPA data set total abundance, all regions and reporting zones.

These figures show that greatly reduced taxa richness occurs in the GBRMPA and Reef Check data sets due to the limited taxa recorded. As expected, these survey methods substantially reduced the ability to detect changes in the data sets that would be useful for an indicator.

Notably, the richness and abundance results for Dry Tropics were reversed when the Reef Check data set was used (**Figures 9-10**). That is, when using RLS data sets the offshore zone in the Dry Tropics was observed to have higher diversity of taxa with lower total abundance than the inshore (Black) reporting zone. When using Reef Check data the opposite was observed. There were no inshore sites in the GBRMPA data set during 2020-21 to compare with offshore. These observations illustrate the loss of resolution caused by limiting the recoded taxa during surveys and support the argument to use only the RLS data set moving forward.

4.3 Taxonomic Richness and Abundance

The RLS data set provides a comprehensive species list that opens the possibility of using fish community data to derive an indicator.

This approach is not without limitations, as the data set does not include metrics of habitat condition, hence it is not easy to identify communities associated with healthy reef ecosystems and/or the shifts in community structure as a response to changing reef health.

For example, there is no indication of whether monitoring sites may have recently been affected by cyclones, COTS outbreaks, turbid/freshwater (flood) flow, bleaching events and so on. So once again, there is no baseline or benchmark that can be used to develop this style of indicator.

Again, a lack of consistency with regards the survey sites results in some zones receiving little to no monitoring effort in a given year. Species accumulation curves (**Figures 1-2**) suggest an additional 4-6 (or more) surveys may be required in each zone annually for sufficient data to be collected. It seems unlikely that upscaling the monitoring to this level would be a realistic option.

BIOENV analysis was used to identify specific environmental factors that are most closely related to the presence or absence of key taxa. When this analysis is applied to the RLS data set at the Family level (**Figure 13**) the top five correlations are herbivore families (Chaetidontidae, Pomacentridae, Acanthuridae, Scaridae and Labridae). The first four of these have been identified as families that maintain coral health (eg Sweatman 2007, Krishnan et al 2013).

BIOENV overall Best results for each number of groups (1- 4 family combinations)

No.Groups	Corr.	Selections
1	0.887	Pomacentridae
2	0.953	Labridae,Pomacentridae
3	0.963	Acanthuridae,Labridae,Pomacentridae
4	0.971	Chaetodontidae,Labridae,Pomacentridae,Scaridae

5 best overall Best results combinations

No.Groups	Corr.	Selections
4	0.971	Chaetodontidae,Labridae,Pomacentridae,Scaridae
4	0.971	Acanthuridae,Labridae,Pomacentridae,Scaridae
4	0.969	Caesionidae,Labridae,Pomacentridae,Scaridae
4	0.969	Acanthuridae,Caesionidae,Labridae,Pomacentridae
4	0.969	Acanthuridae,Gobiidae,Labridae,Pomacentridae

BIOENV - Individual Family best correlations

1	0.887	Pomacentridae
1	0.804	Labridae
1	0.532	Scaridae
1	0.531	Chaetodontidae
1	0.427	Acanthuridae
1	0.401	Gobiidae
1	0.401	Nemipteridae
1	0.380	Pomacanthidae
1	0.372	Serranidae
1	0.330	Blenniidae

Figure 13: BIOENV results for the RLS data set.

This indicates that fish community data could be explored further, particularly if corresponding environmental (habitat) data was available.

4.4 Taxa richness and total abundance

Taxa richness and total abundance boxplots over the entire duration of the RLS Whitsundays dataset are shown in **Figures 14-15**. These indicate that taxa richness is typically higher in the Offshore Zone than the Central Zone. Offshore and Whitsundays zone taxa richness are similar in some years, but generally the offshore sone has higher richness. They also indicate that taxa richness within each reporting zone is variable temporally.

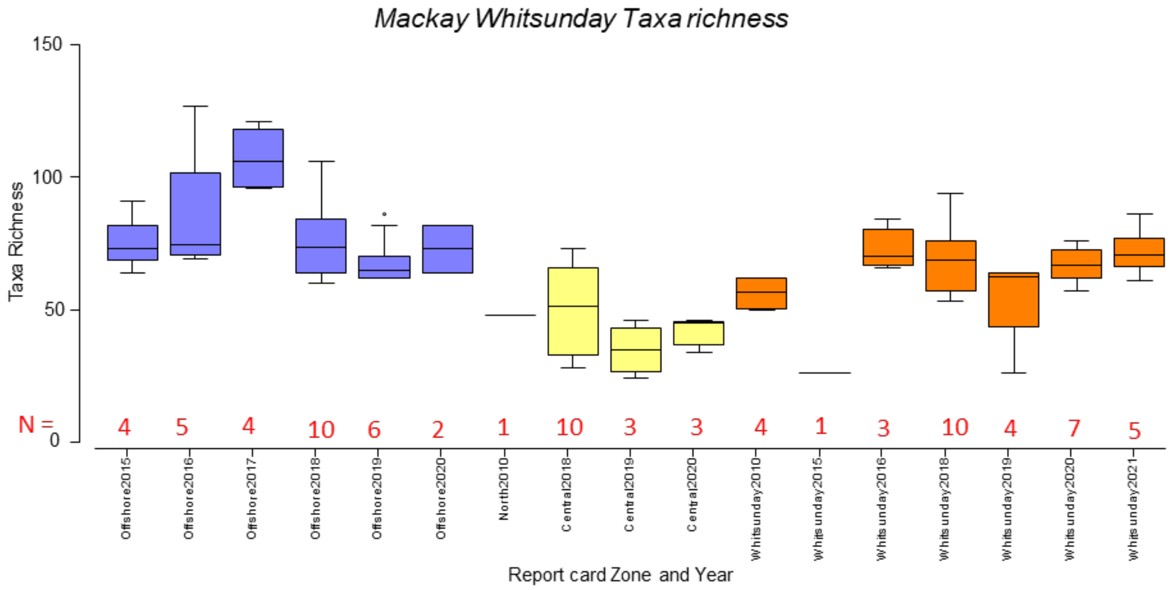

Figure 14: Boxplot of taxa richness at MWI reporting zones using RLS data from 2015-21. n represents the number of monitoring events.

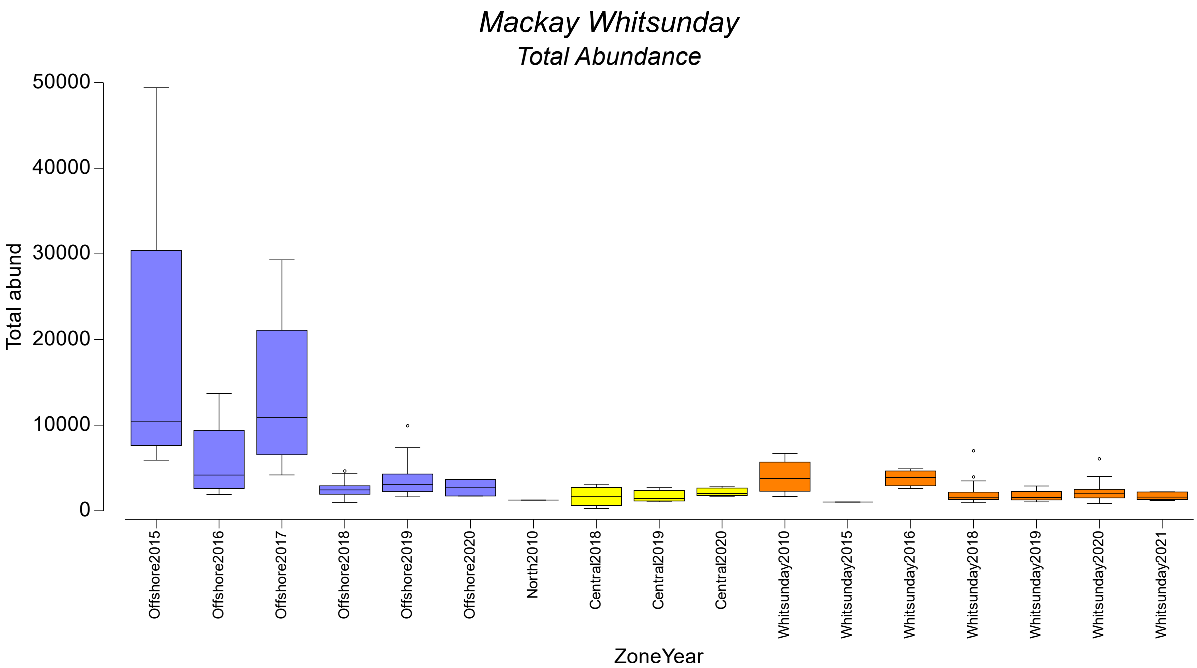

Figure 15: Boxplot of total abundance at MWI reporting zones using RLS data from 2015-21.

The total abundance boxplot (**Figure 15**) is dominated by very high abundances from 2015-2017, with subsequent years having lower total abundance and less variance in the data.

nMDS (transformed)(**Figure 16**) and bubble plots (**Figure 17**) on the full MWI data set show a large degree of separation, with offshore sites grouping to the left, Whitsunday sites largely in the centre and central zone sites grouped to the right. This indicates that in terms of community structure the taxa contributing to richness are different between reporting zones.

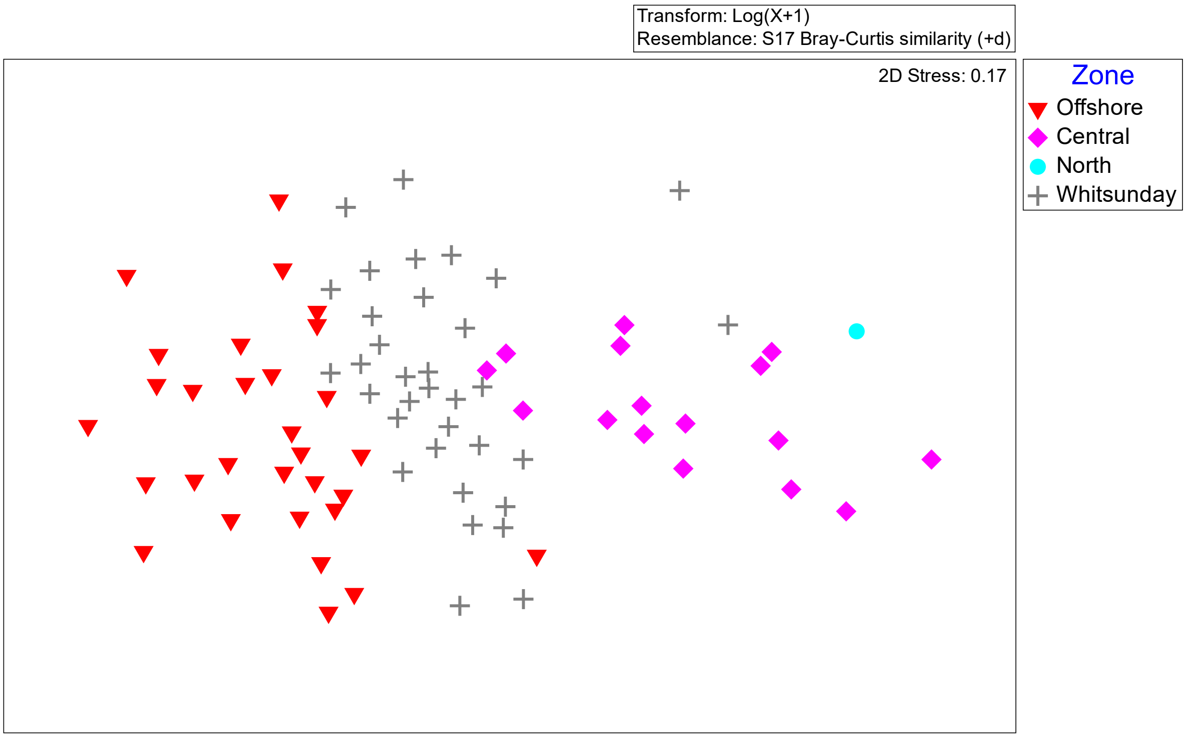

Figure 16: nMDS plot for MWI taxa richness data since 2015

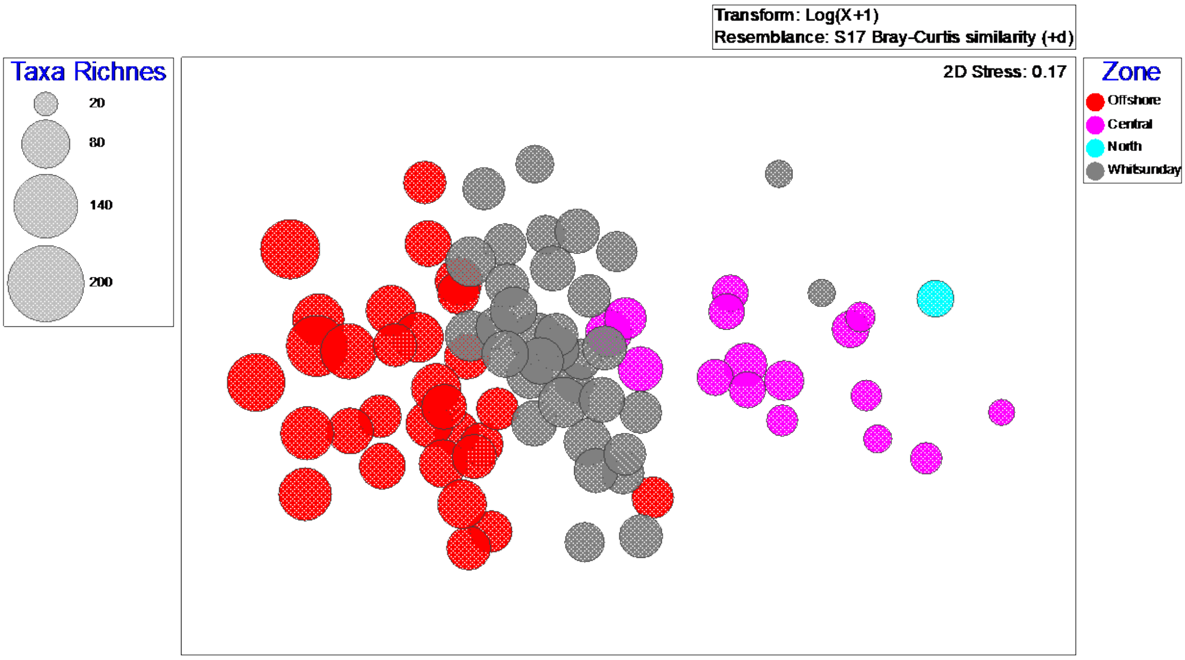

Figure 17: Bubble plot for MWI taxa richness data since 2015

The separation of reporting zones on the basis of total abundance was similar to that for taxa richness, though a little more overlap and greater temporal variability within monitoring sites was observed in the inshore zones.

When the full WT data set is displayed as a boxplot of taxa richness (**Figure 18**) and taxa abundance (**Figure 19**) it is found that the Offshore and Northern Zones have similar taxa richness. The Central and Southern zones have comparable richness with each other, but lower than the Offshore and Northern zones.

Total abundance data is highly variable within the WT Central and Southern zones, and appears to be greater that at the offshore zones early in the data set but declines to well below the abundance observed at offshore zones later in the data set.

nMDS and bubble plots for the full RLS WT data set (**Figures 20 and 21**) show that the offshore sites appear to have distinctly different fish communities to the inshore sites. The inshore zones also exhibited grouping but were not as clearly separated as the offshore zone. Again, this suggest each reporting zone may have different taxa richness than other reporting zones.

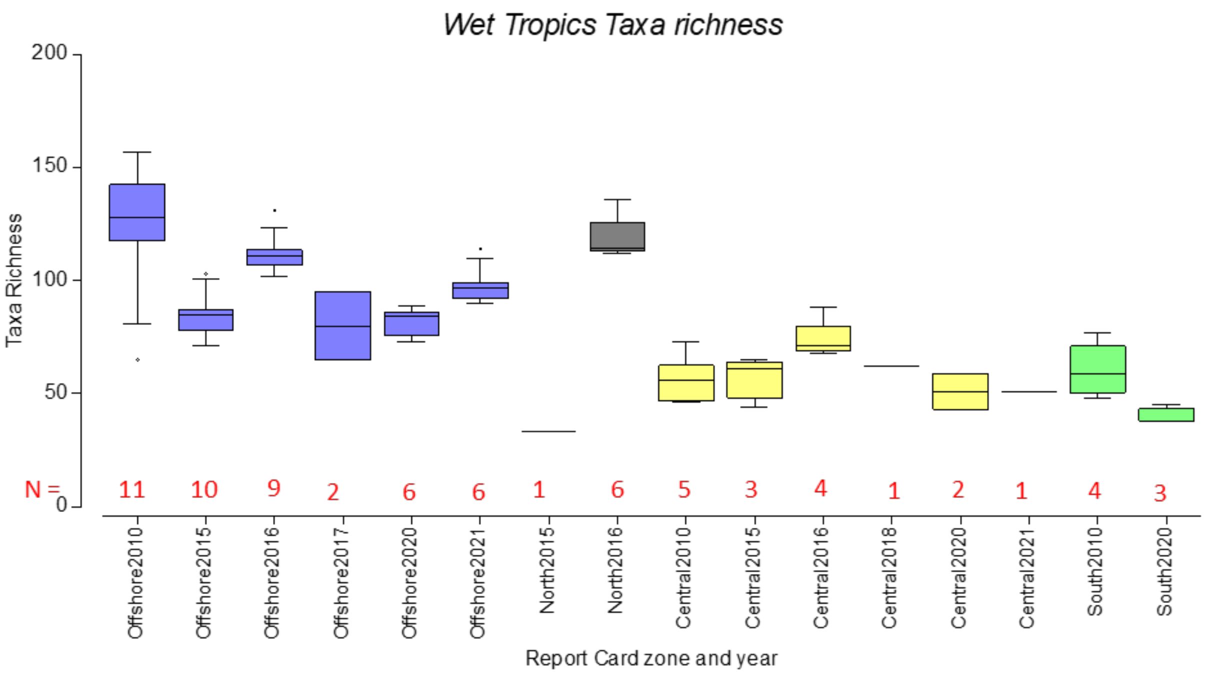

Figure 18: Boxplot of WT taxa richness data since 2015

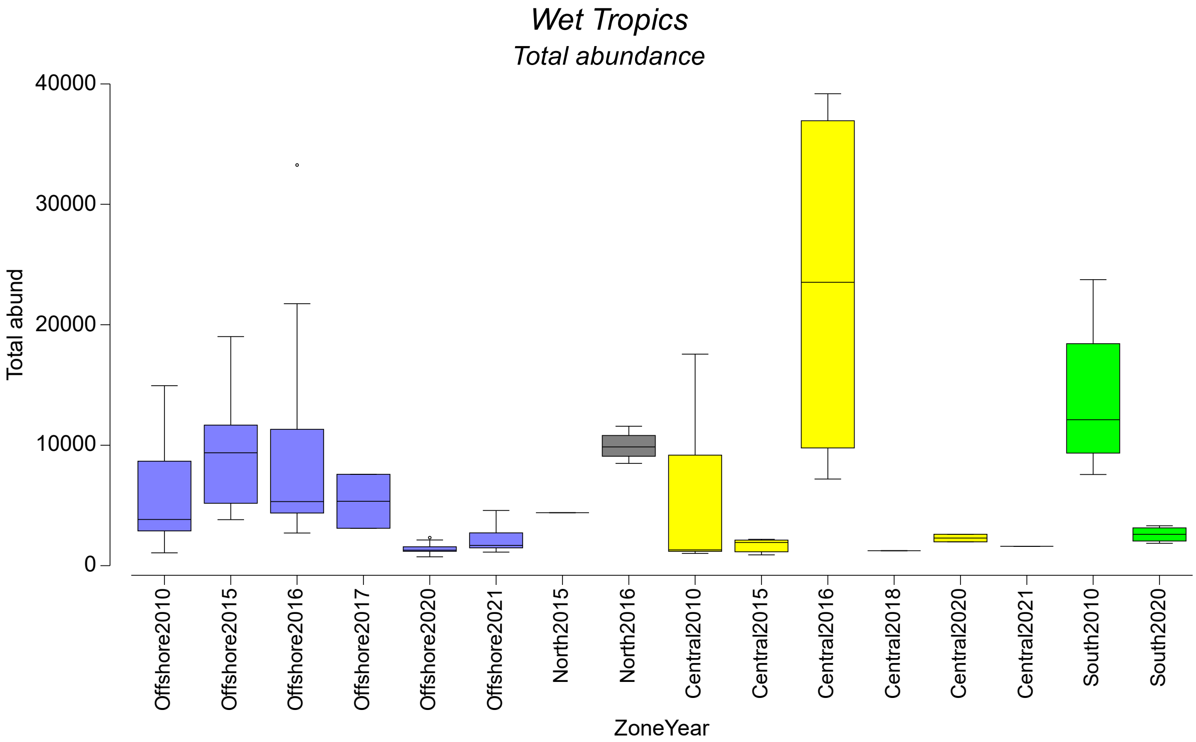

Figure 19: Boxplot of WT total abundance data since 2015

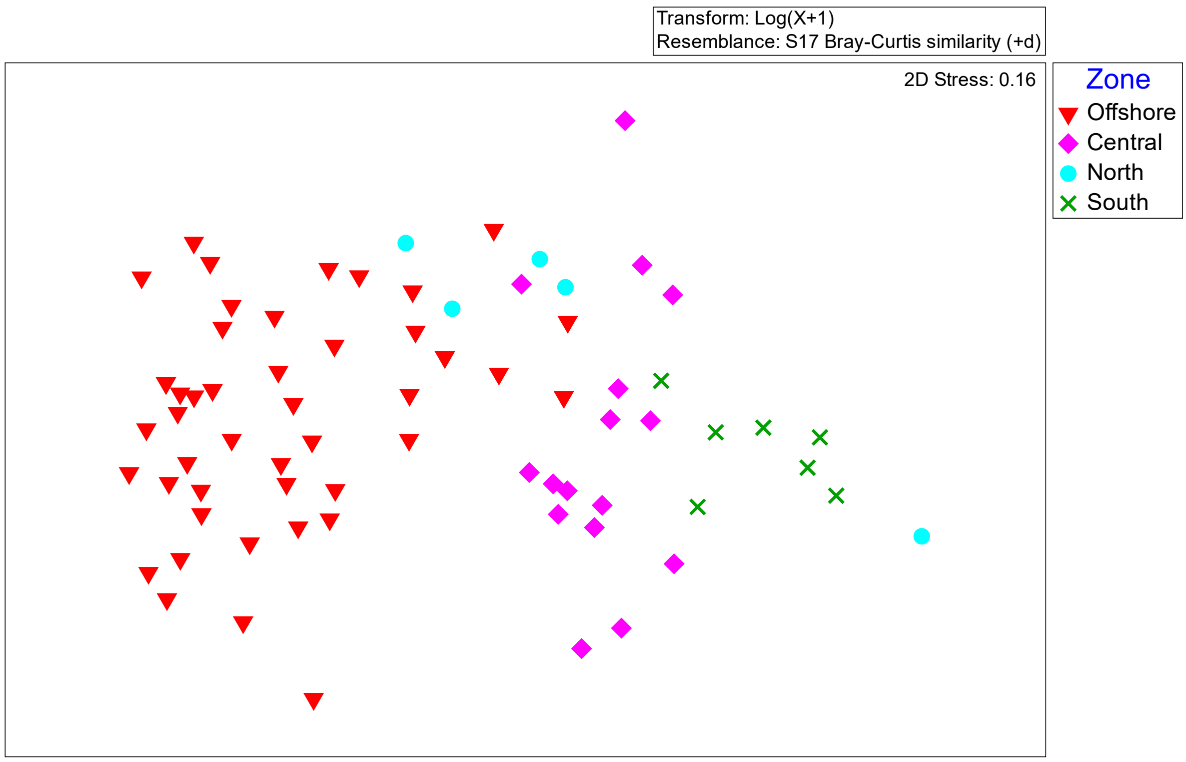

Figure 20: nMDS plot for WT total abundance data since 2015

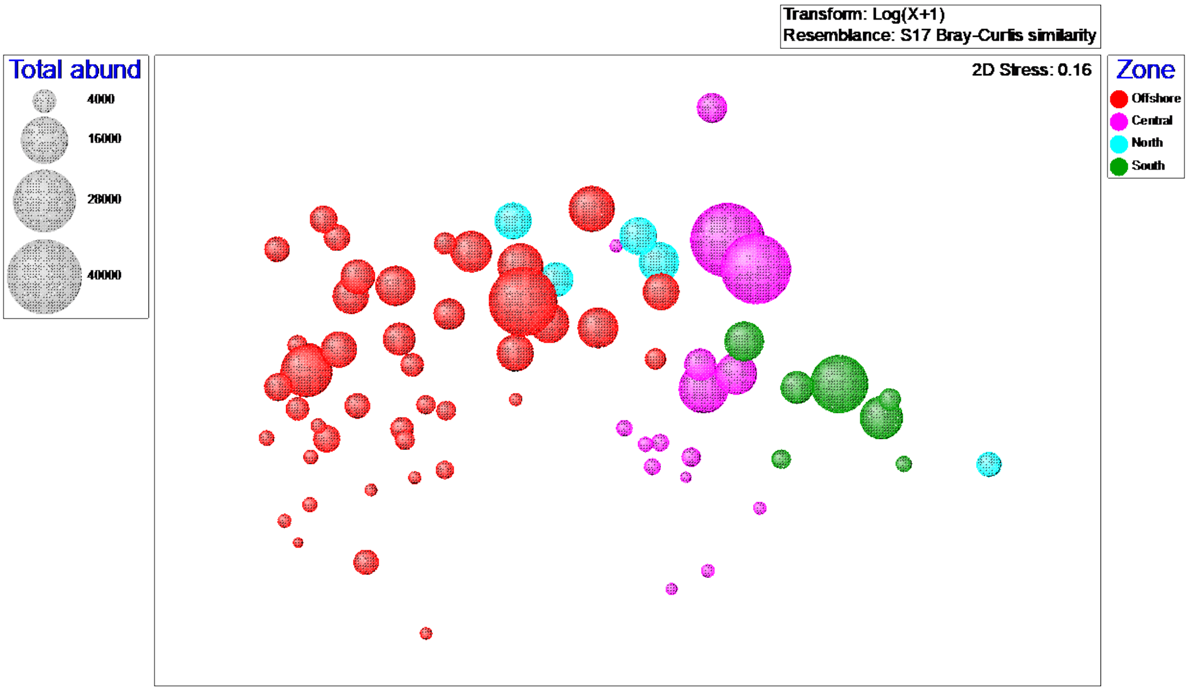

Figure 21: Bubble plot for MWI taxa richness data since 2015.

4.5 Taxonomic Proportions Analysis

Taxonomic proportions analysis also requires the selection of an appropriate reference site/zone. In this instance offshore zone data was used as a reference with the assumptions that fish communities on the offshore reefs are likely to be the least impacted and that unimpacted inshore fish communities would be similar to offshore communities. The taxa richness and/or abundance is then assessed as a proportion of the offshore richness and/or abundance. The limitations of these assumptions are recognised as significant and have been previously discussed.

Comparing taxonomic proportions within a zone using long term data from the same zone was considered but was not viable for the purposes of this study as it is more applicable for assessing temporal trends in proportions rather than an annual snapshot of community structure or health. This is because the long-term median of taxa richness or abundance within a zone may be impacted by natural or anthropogenic influences.

The proportions analysis was undertaken for the MWI and DT regions using both univariate and multivariate techniques. The approach was not explored for the DT data set due to insufficient data being available.

Taxonomic richness proportion results for MWI and WT are shown in **Figures 22** and **24**, respectively. This analysis was repeated using the median taxa richness for offshore sites, for all years (**Figures 23** and **25**). The use of medians helps to reduce the influence of unusual or outlying taxa richness and total abundance data. The analysis is complicated by the fact that not all zones were sampled in all years, which must be considered when reading the boxplots.

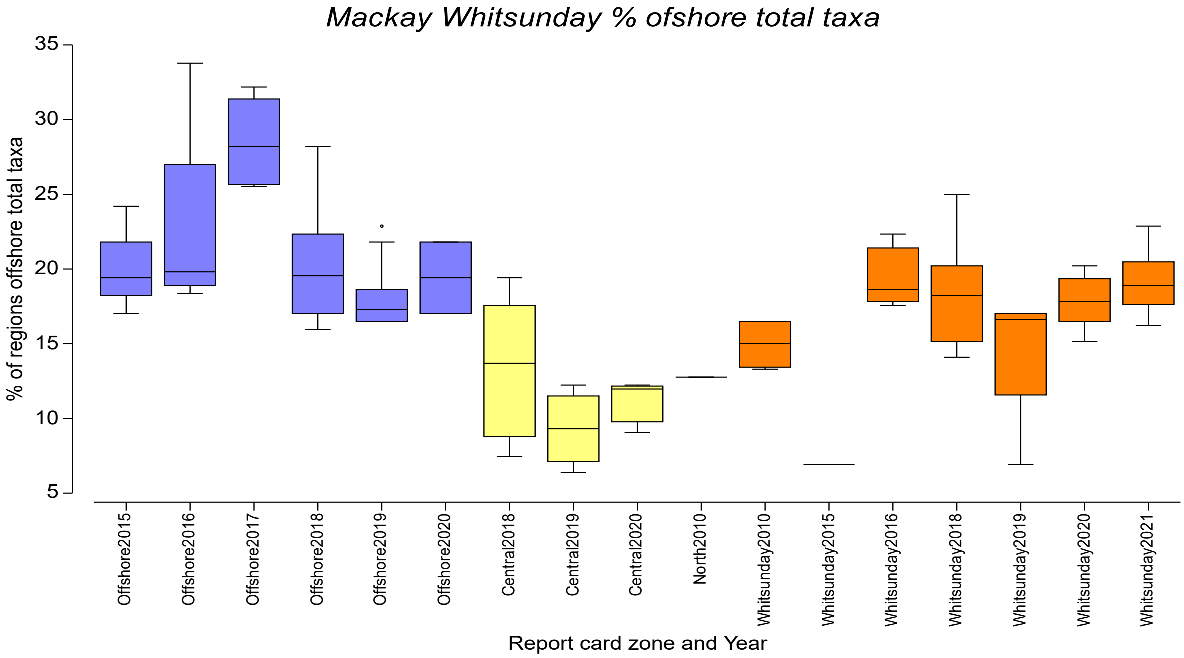

Figure 22: Boxplot of observed taxa as a proportion of all taxa recorded in MWI offshore zone (all years).

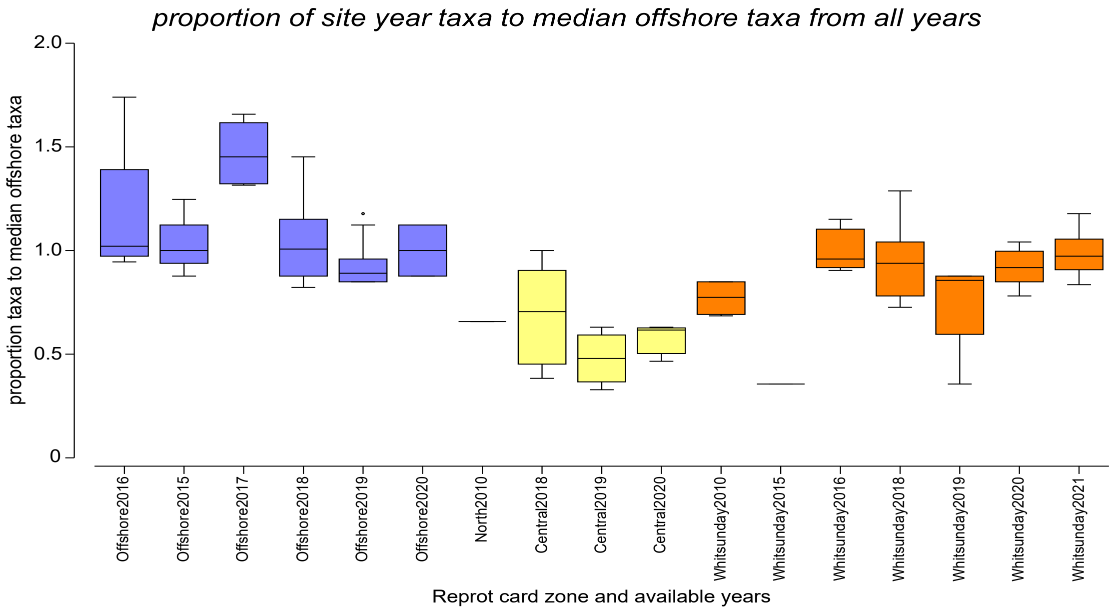

Figure 23: Boxplot of observed taxa as a proportion of median taxa richness recorded in MWI offshore zone (all years).

The MWI data indicates that from 2018-2020 the taxa recorded at offshore zone sites and Whitsunday Zone sites as a proportion of all offshore taxa for MWI sites were comparable. Central Zone sites over the same period exhibited distinctly different taxa proportions. The use of median offshore taxa richness as reference data reduced variance but did not alter the outcome of the analysis.

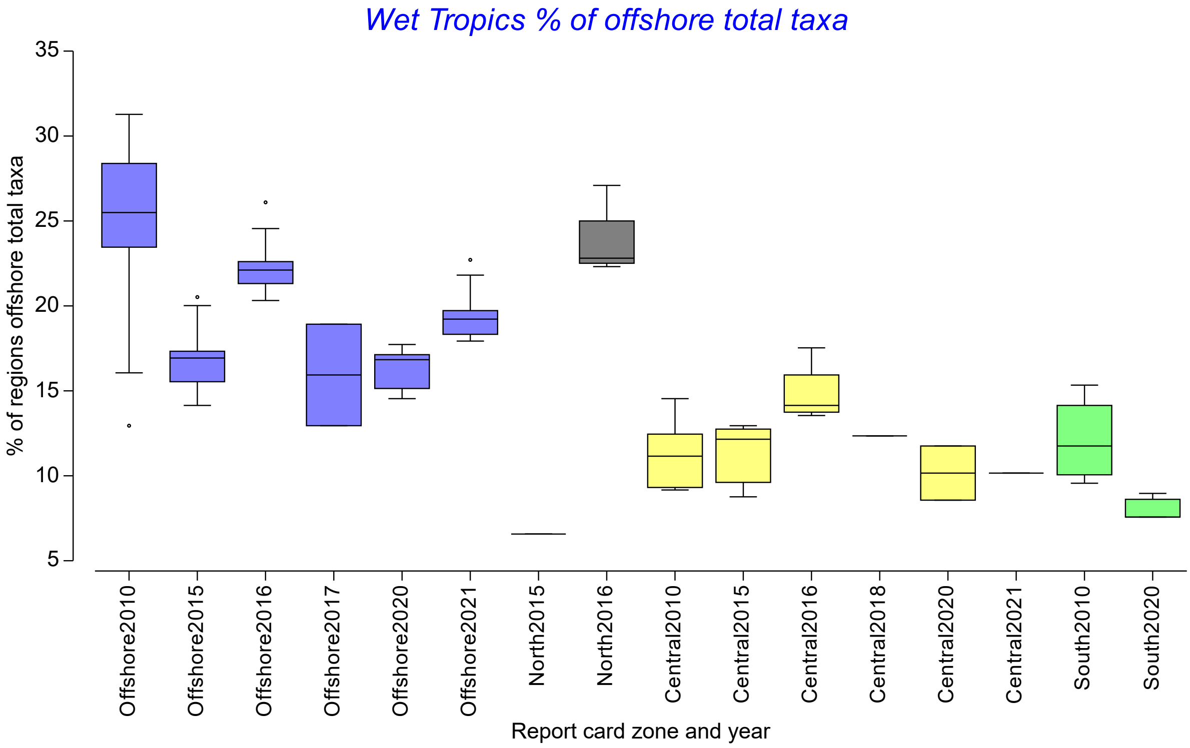

Figure 24: Boxplot of observed inshore zone taxa as a proportion of all taxa recorded in offshore zone (WT all years).

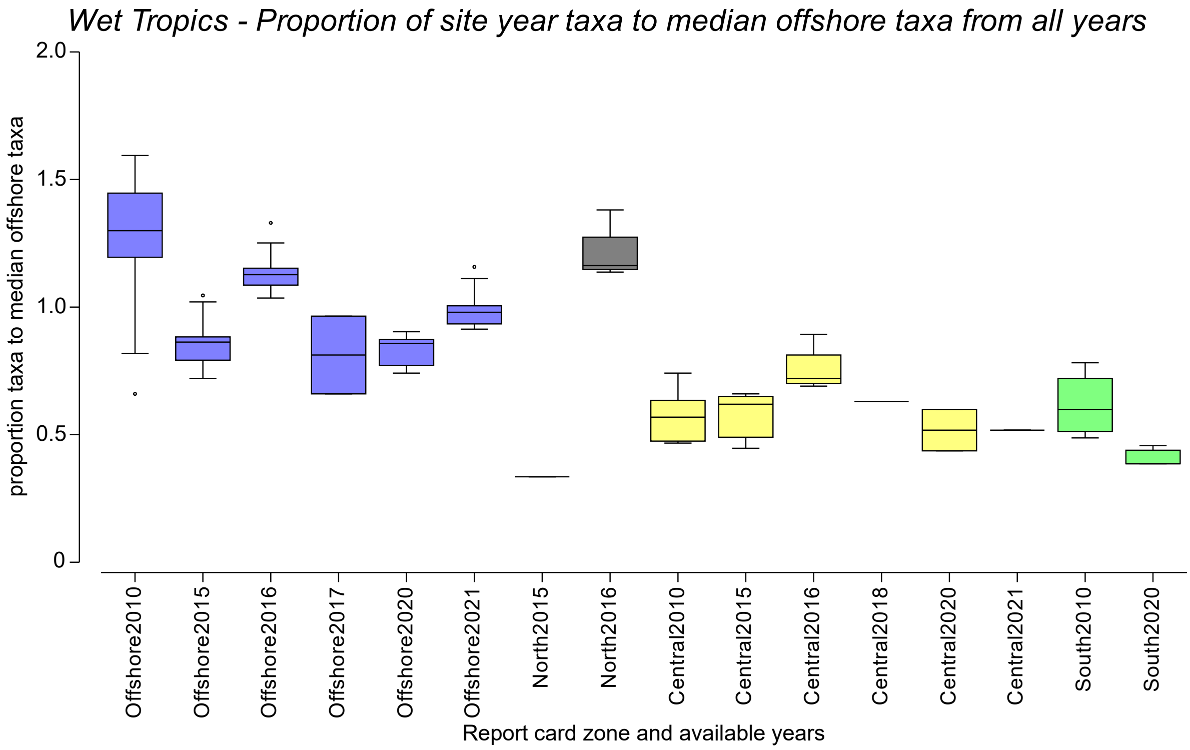

Figure 25: Boxplot of observed taxa within reporting zones as a proportion of median taxa recorded in offshore zone (all years).

Wet tropics data shows that in 2015 a single Northern Zone site exhibited a very low proportion of taxa richness compared with total offshore taxa, but further sampling in 2016 resulted in a similar proportion of taxa as the offshore sites in the same year. Central and Southern Zones exhibited lower proportions of the total recorded offshore taxa. The use of median offshore data reduced variance and resulted in the 2016 Northern Zone having a higher proportion of recorded taxa than the Offshore Zone in the same year. The central and southern zones remained lower than their respective Offshore Zones

4.6 Fish Community Structure (Herbivores)

Given literature and community analysis of family influences the proportion of herbivores has been explored as a proportion of total taxa and of the median offshore herbivore taxa for both the MWI and WT regions. This analysis was undertaken at both the species and family level.

4.6.1 MWI herbivores

Figures 26 shows the herbivore taxa richness as a proportion of total taxa richness at each MWI site. The number of herbivore taxa was found to differ between offshore and inshore zones prior to 2018, with the Whitsundays Zone having a larger number of herbivore taxa than the Offshore Zone. Post 2018, the herbivore taxa richness at the Whitsundays Zone declined and became comparable with the Offshore sites. No surveys were conducted in the central zone prior to 2018, with post 2018 results indicating a lower proportion of herbivores in this zone when compared with Offshore and Whitsunday's zones.

Figure 27 shows the herbivore richness as a proportion of the median offshore zone herbivore taxa richness. This approach reduces the influence of surveys that record unusually high or low herbivore taxa richness that might be associated with localised or short-term influences. The results and observations are similar to those described above for **Figure 26.**

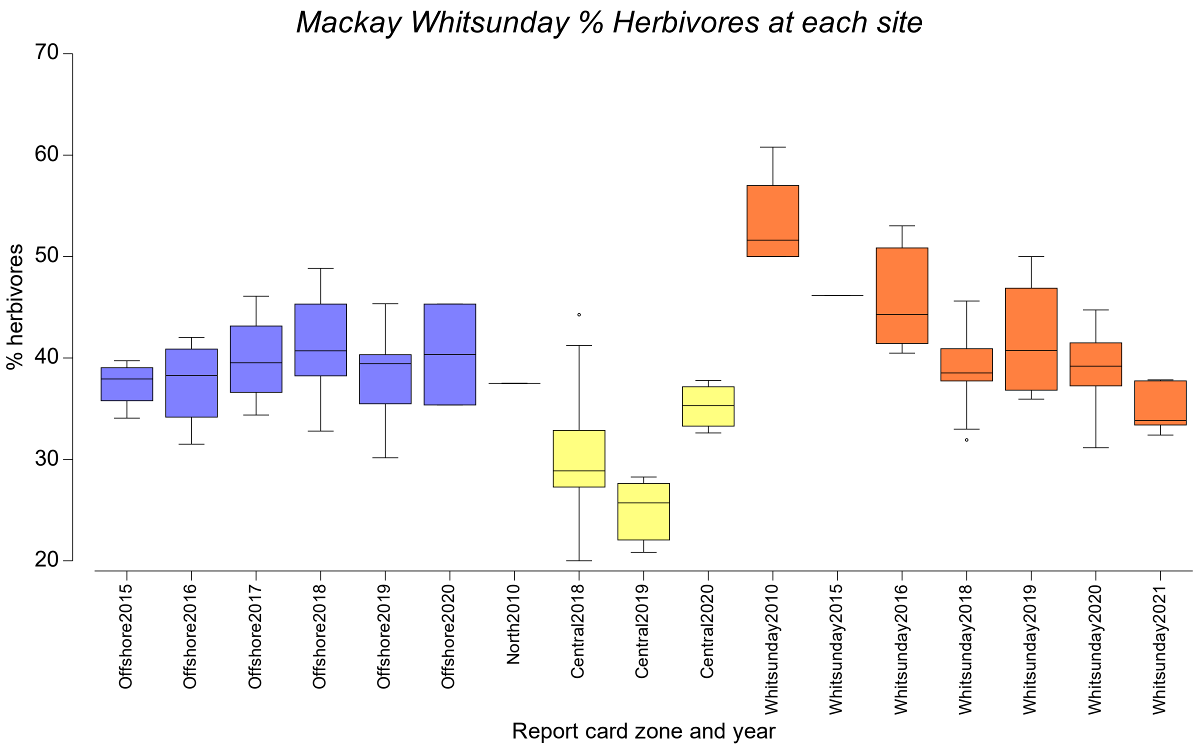

Figure 26: Boxplot of herbivore taxa as a proportion of total taxa within the MWI region.

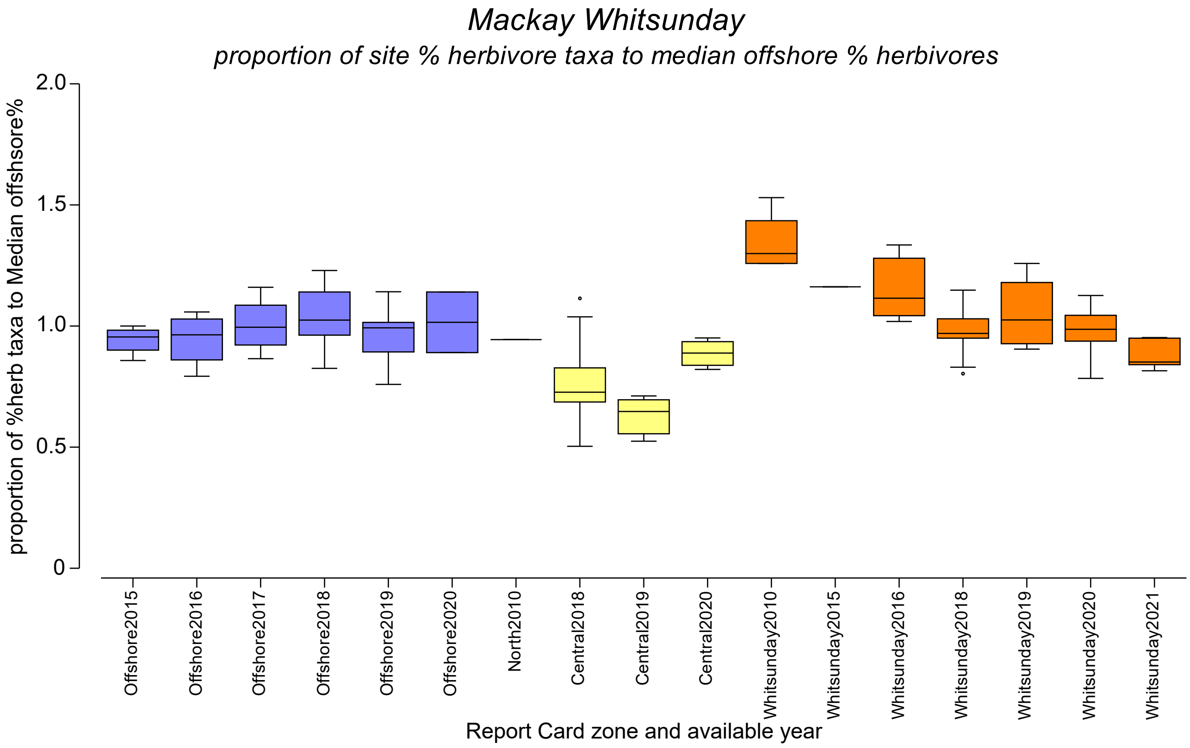

Figure 27: Boxplot of herbivore taxa as a proportion of median offshore taxa within the MWI region

Repeating this analysis using herbivore families as a proportion of total families (**Figure 28**) and herbivore families as a proportion of median offshore herbivore families (**Figure 29**) significantly reduced the ability to discern differences in the data sets for each zone. This would result in a significantly less sensitive indicator of fish community health.

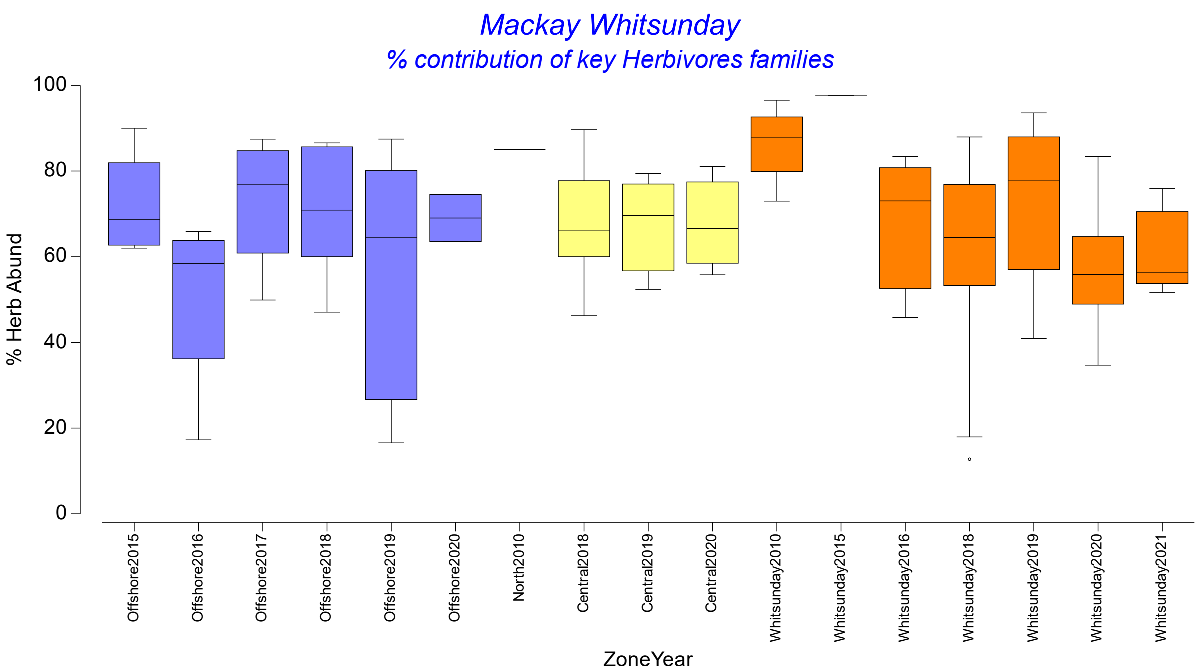

Figure 28: Boxplot of families as a proportion of total taxa within the MWI region.

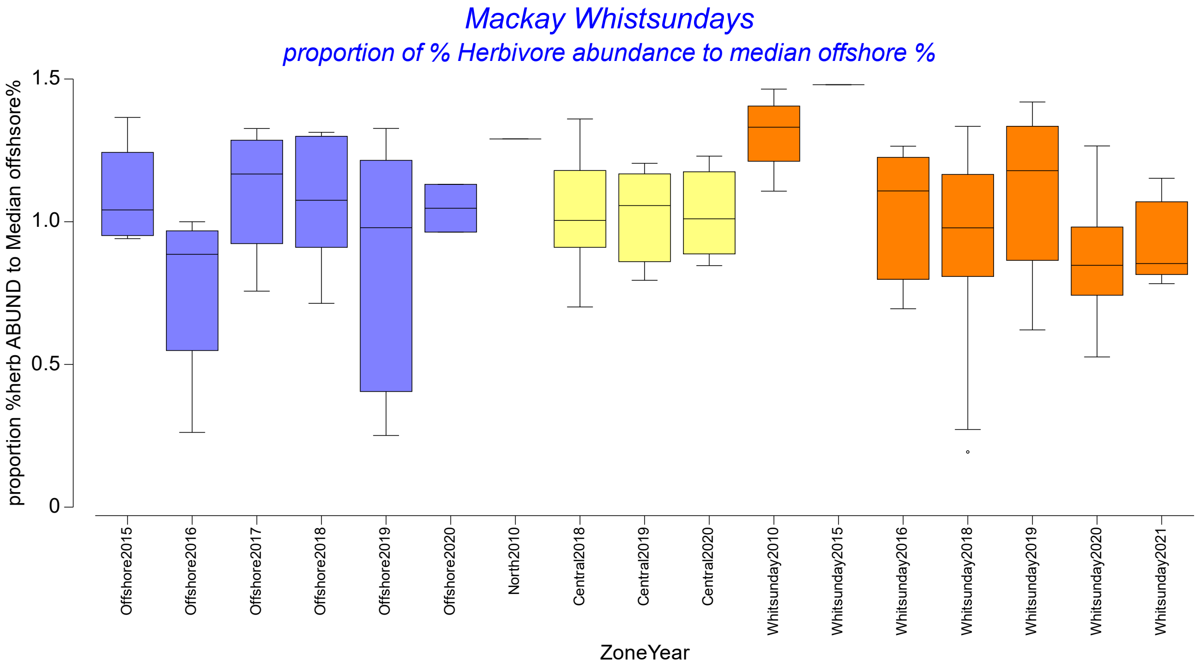

Figure 29: Boxplot of herbivore abundance as a proportion of median offshore taxa within the MWI region.

4.6.2 WT Herbivores

Figures 30 shows the herbivore taxa richness as a proportion of total taxa richness at each WT reporting zone. Interpretation is challenging due to the inconsistent sampling frequencies; However the data can be seen to differ both between and within reporting zones. The proportion of herbivore taxa generally appears to be higher in the Offshore and Central Zones, although the latter also exhibits more variability between sites sampled in the same year. The North and South Zones appear to have a smaller proportion of herbivorous taxa.

Figure 31 shows the herbivore taxa richness in each zone as a proportion of the median offshore zone herbivore taxa richness. This approach reduces the influence of surveys in which unusually high or low herbivore taxa richness that might be associated with localised or short-term influences were recorded. The results and observations are similar to those described above for **Figure 30**, indicating that extreme results were not significantly impacting on the overall analysis.

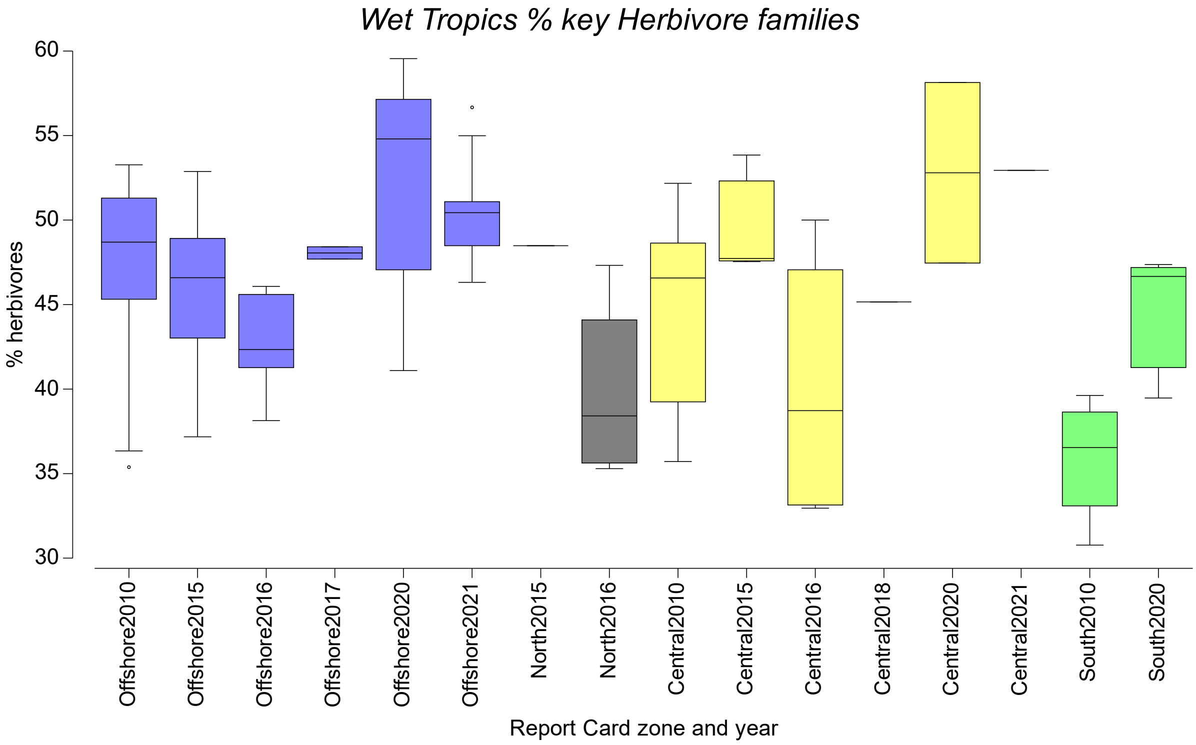

Figure 30: Boxplot of herbivore taxa as a proportion of total taxa within the WT region

Repeating this analysis using herbivore families as a proportion of total families (**Figure 32**) and herbivore families as a proportion of median offshore herbivore families (**Figure 33**) significantly reduced the ability to discern differences in the data sets for each zone. This would result in a significantly less sensitive indicator of fish community health.

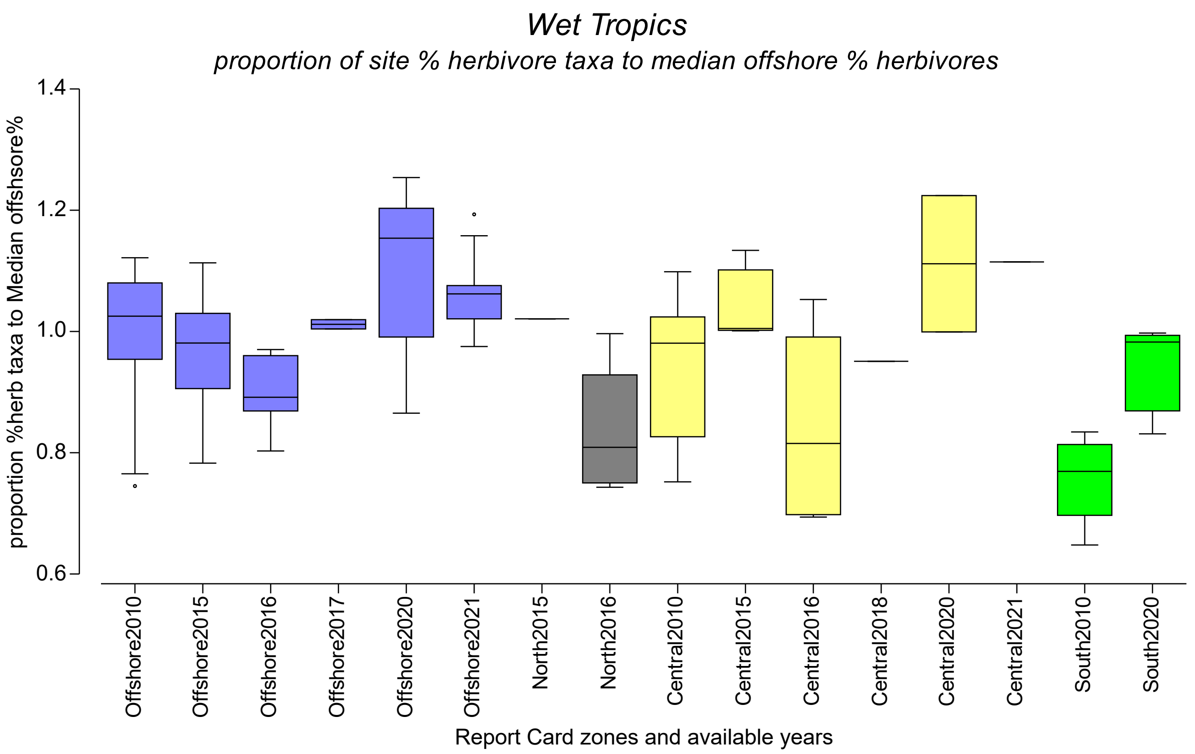

Figure 31: Boxplot of herbivore taxa as a proportion of median offshore taxa within the WT region

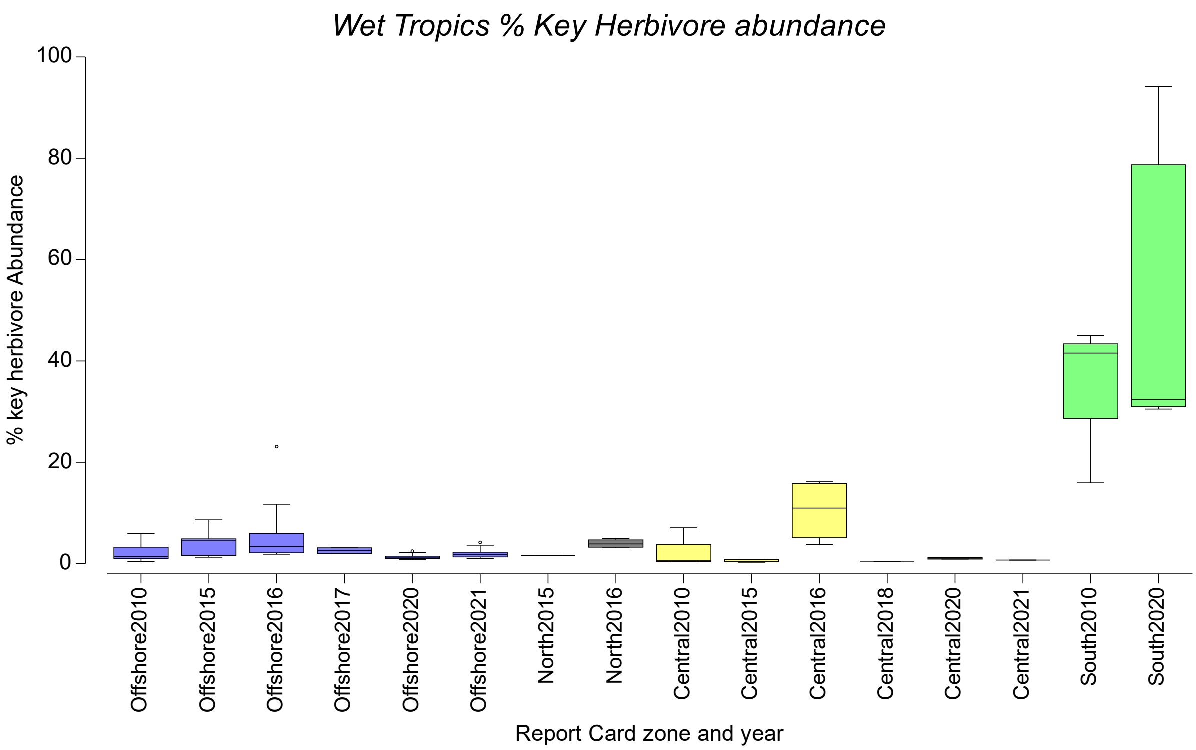

Figure 32: Boxplot of families as a proportion of total taxa within the WT region.

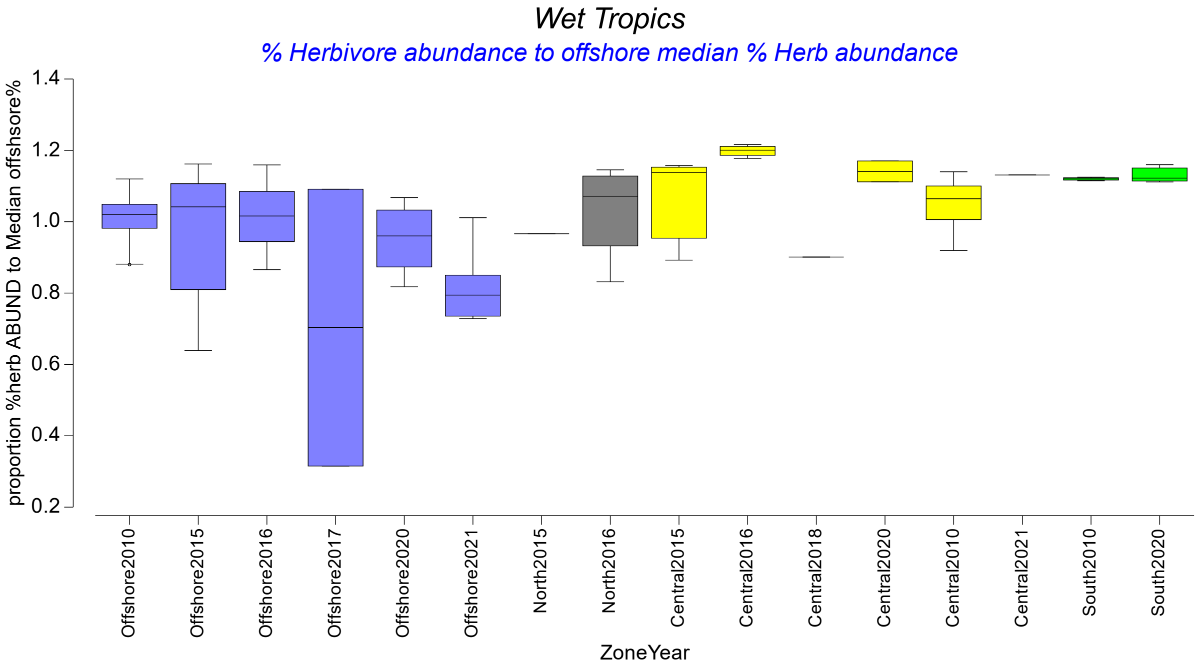

Figure 33: Boxplot of families as a proportion of total taxa within the WT region.

4.7 Community structure analysis

4.7.1 Community Structure Using GBRMPA Data

nMDS was used to further assess how the GBRMPA rapid assessment and tourism data sets differ as a result of the differences in the number of taxa surveyed. Note that due to the amount of sites with zero observations the BrayCurtis dissimilarity was run with a dummy value (1).

Figure 34 shows the results of this nMDS on presence/absence data and shows strong grouping of the two data sets. **Figure 35** utilises square-root transformation of data to reduce the influence of large dissimilarity values that might dominate the analysis. This nMDS resulted in the majority of tourism data points grouping to the right of the plot, with the rapid assessment and a small number of tourism points grouping to the left.

The tight grouping of the rapid assessment surveys indicated less variance in the data set, which may be due to the smaller number of taxa recorded using the rapid assessment method.

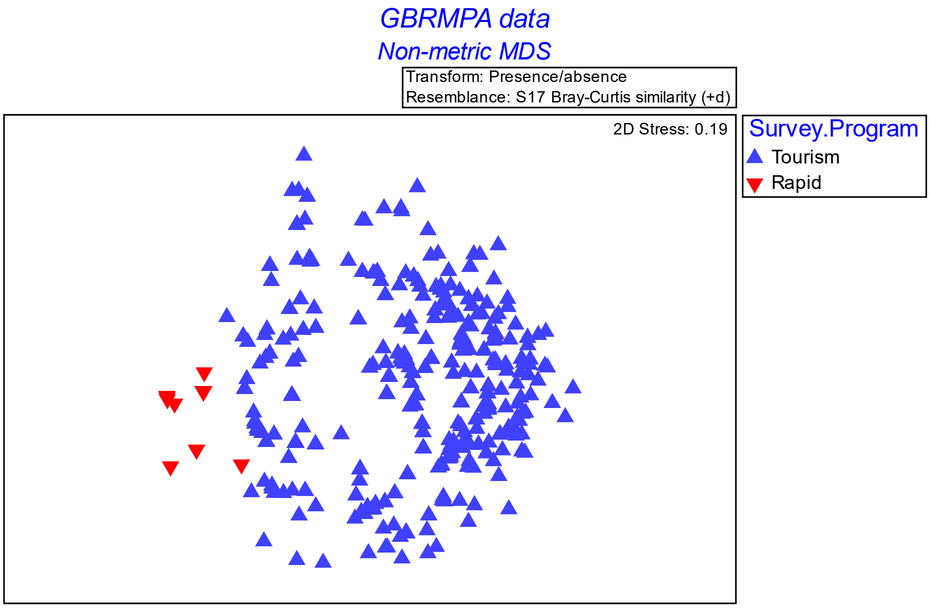

Figure 34: nMDS (presence/absence) of GBRMPA rapid assessment and tourism data.

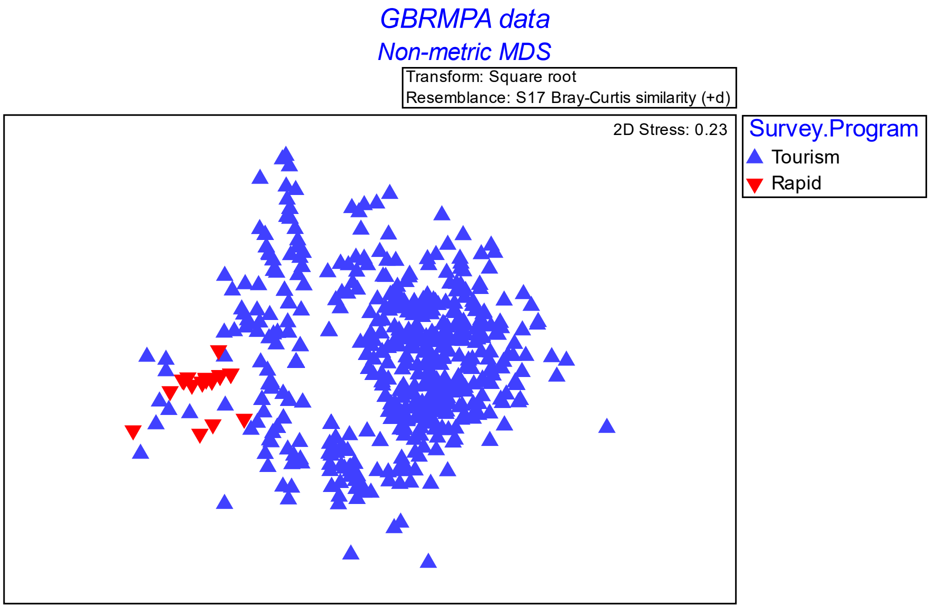

Figure 35: nMDS (square root) of GBRMPA rapid assessment and tourism data.

Removal of the rapid assessment data resulted in high stress on a two-dimensional plot, hence a three-dimensional plot was instead used. Axis 1 and 2 (**Figure 36**) show best separation of region zones.

Data were dominated by the WT Central Zone, which is suspected to be due to a high proportion of diving tourism in this zone. It is likely that data from this zone are dominated by a low number of reefs that are sampled numerous times each year!

Vectors show direction of influence or species with a Pearson correlation >0.3

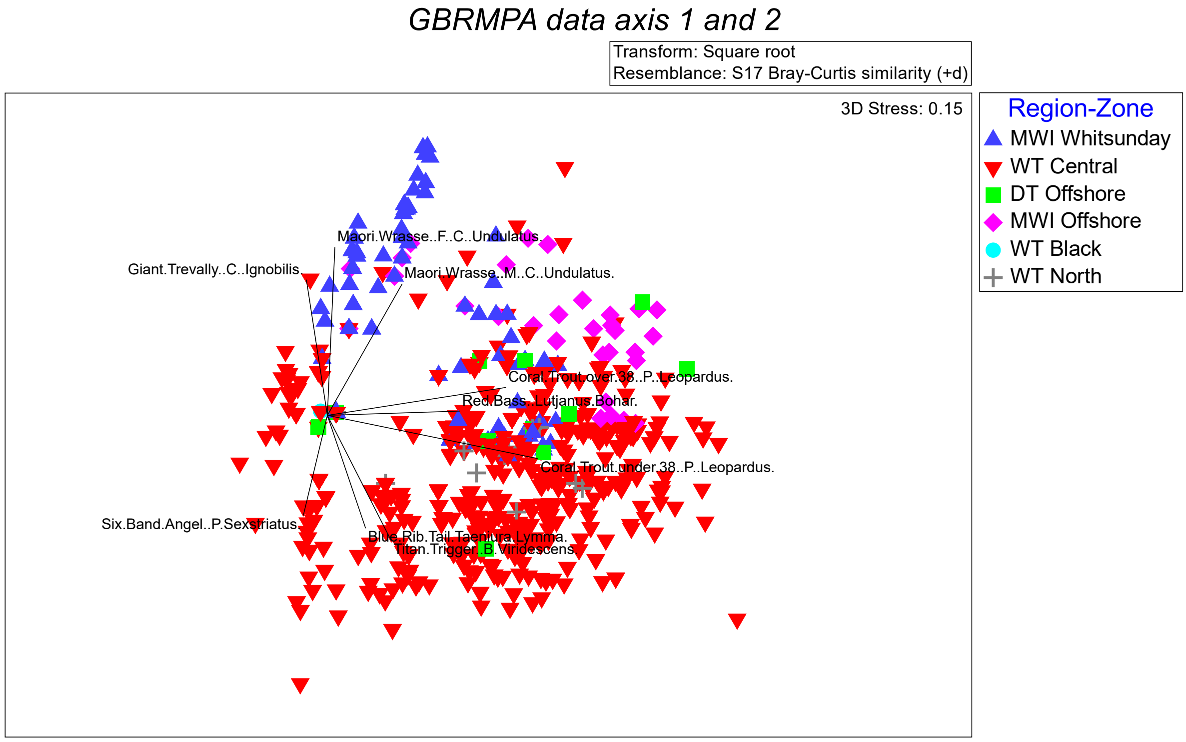

Figure 36: nMDS (square root) of GBRMPA tourism data set, axes 1 and 2.

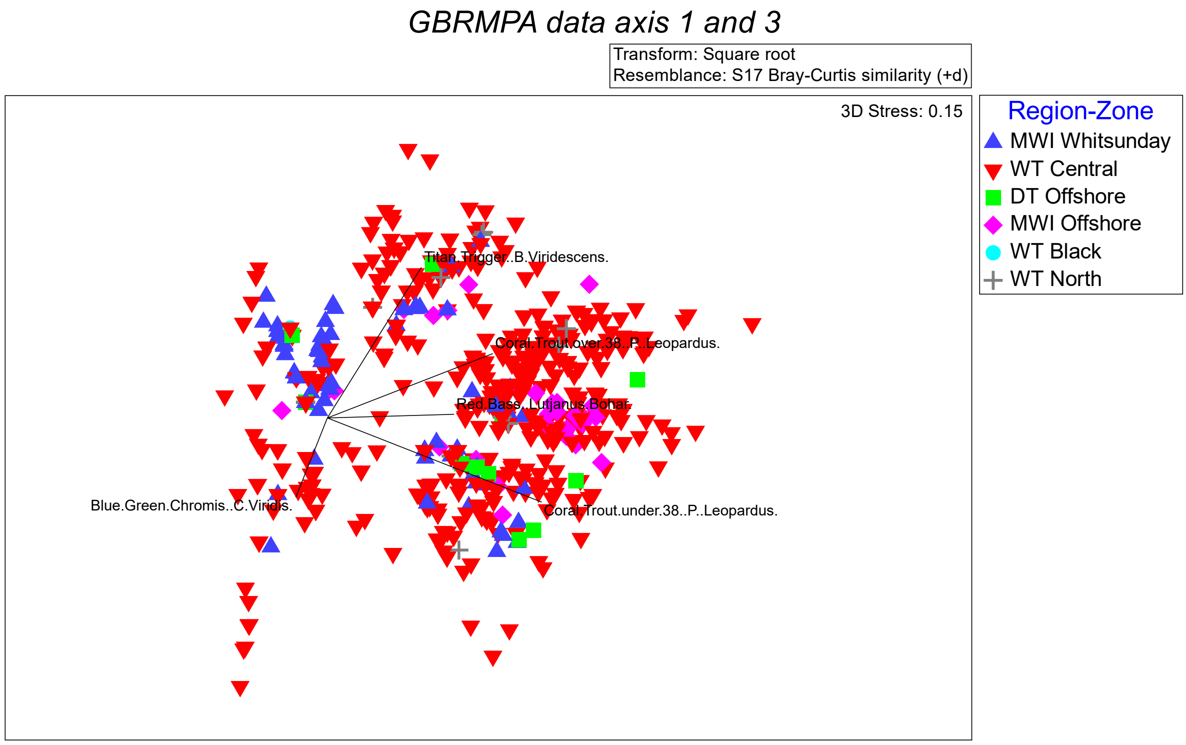

Figure 37: nMDS (square root) of GBRMPA tourism data set, axes 1 and 3.

The BVSTEP analysis is useful for data sets with high-dimensional data and multiple predictor variables. It was used in this case to identify the species best representing the nMDS abundance (sqrt transform) analysis. Smallest subset with highest correlation of 0.654 was 9 taxa combined:

- Blue Green Chromis (*C. viridis*)
- Six Band Angel (*P. sexstriatus*)
- Coral Trout under 38cm (*P. leopardus*)
- Coral Trout over 38cm (*P. leopardus*)
- Titan Triggerfish (*B. viridescens*)
- Red Bass (*L. bohar*)
- Bigeye Trevally (*C. sexfasciatus*)
- Blue Rib Tail (*T. lymma*)
- Painted Sweetlip (*D. picta*)

Of these, only three taxa (*P. leopardus, B. viridicens* and *L. bohar*) are recorded during GBRMPA rapid assessments, further confirming the limitations of this data set.

4.7.2 Community Structure Using RLS Data

An nMDS plot of RLS data was undertaken using log transformation to even out the effects of highly dissimilar or outlying data. The analysis shows that sites groups quite strongly by reporting zone, with occasional overlapping or outlying data (**Figure 38**).

All offshore sites grouped to the right, with Wet Tropics sites in the upper right corner, Mackay, Whitsundays, Isaac sites in the bottom left corner and Dry Tropics sites falling between the other two. There are also strong indications of a north-south gradient, with WT sites grouping largely to the top right, MWI sites grouping to the bottom left and DT site once again falling between the other two.

A bubble plot shows increasing species diversity of the five key herbivore families within the offshore zones (**Figure 39**). This challenges the use of methods that assume similarity between inshore and offshore zones and supports the need for species level data.

When the abundance of species within the five herbivore families is presented as a bubble plot the results are less clear. However, abundance is generally higher in the Wet Tropics and in offshore zones.

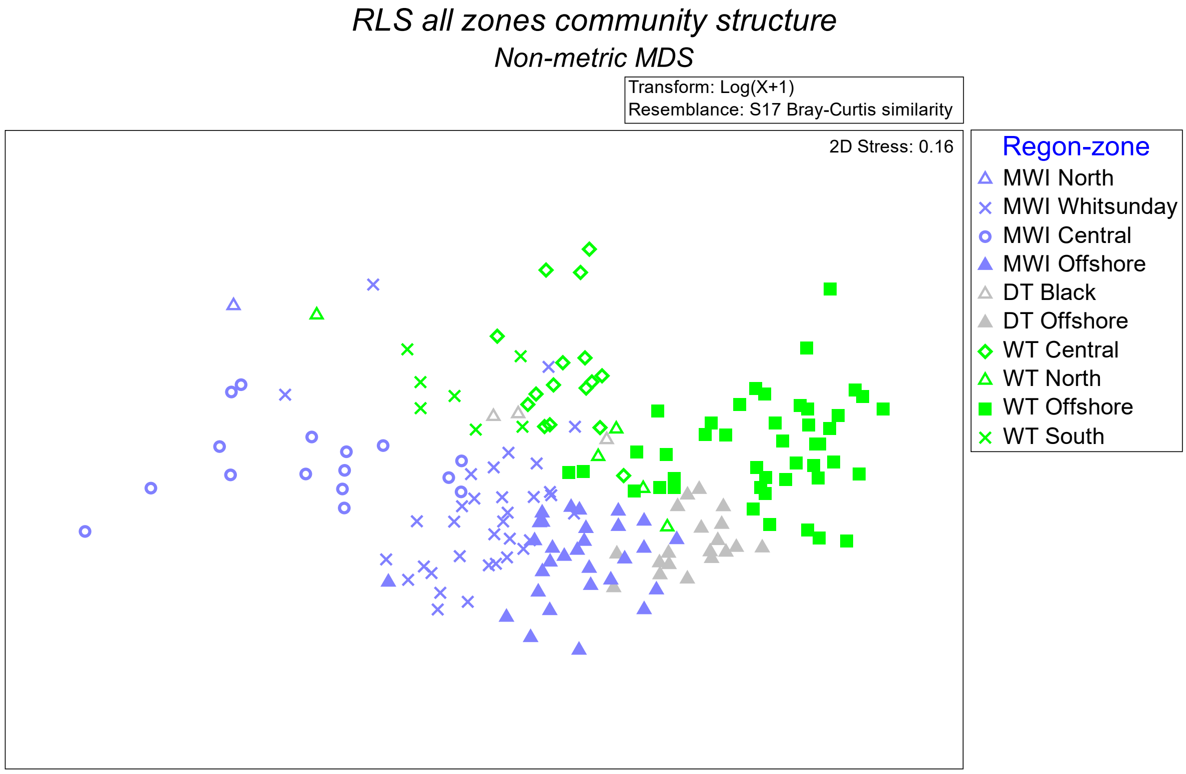

Figure 38: nMDS (log) of RLS data.

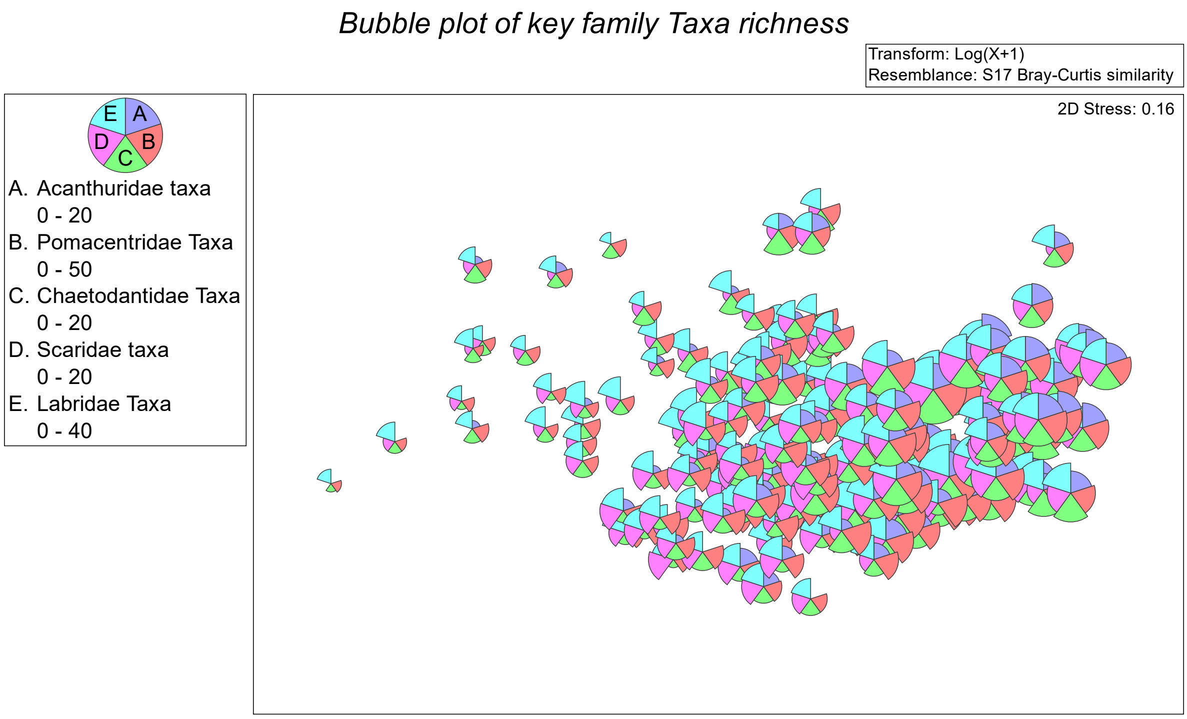

Figure 39: Bubble plot showing taxa richness within five key herbivore families.

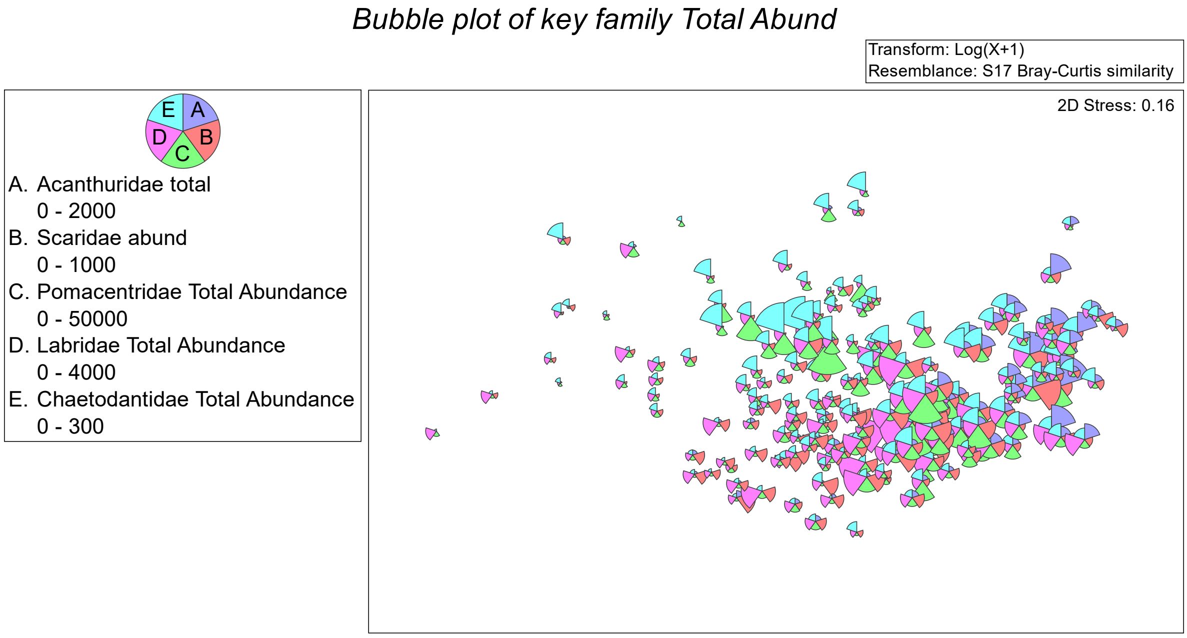

Figure 40: Bubble plot showing total abundance of species within five key herbivore families.

5 Discussion

5.1 Integration of data sets

The integration of multiple data sets is frequently challenging, particularly with respect to a coral reef system where temporal, spatial, habitat and other factors are highly variable. Different sampling methods further complicate this process.

In this study the data collection methods for the three data sets (RLS, GBRMPA and Reef Check) sets are somewhat similar, being diver-swum transects with the observed fish taxa recorded and abundance estimated. The key differences that challenge data integration relate mainly to the recording protocols:

- Reef Life Survey participants record all fish taxa observed to species level and estimate the abundance of each;
- GBRMPA Eye On The Reef tourism participants record the abundance of 49 key taxa and a more limited number of taxa (11) are recorded during rapid assessment, and
- Reef Check participants record the abundance of 13 key taxa, which are a combination of species and family level groupings.

Despite these differences in recording protocols it was identified that integration of data might overcome the issue of some report card reporting zones having few, or no, surveys conducted in a given year.

By its nature, integration of data sets involves standardising the data to the lowest common denominator and hence sacrificing some of the information contained in the more detailed dataset. The first stage in this assessment was therefore to determine whether the lowest common denominator would provide sufficient information on which an indicator metric might be developed.

The most obvious loss of data resolution would be reducing the large number of taxa recorded in the RLS surveys to the much more limited range of taxa recorded in the GBRMPA and Reef Check surveys.

Boxplots illustrate the loss of analytical resolution resulting from recording limited taxa (**Figures 9-12**) and nMDS plots clearly indicated that within the GBRMPA data set the reduced taxa list recorded in rapid assessment surveys limit the ability to understand community structure when compared with the tourism data set (**Figures 34 and 35**). Further, when community structure is considered across all sites (**Figure 38**) it is clear that latitudinal and longitudinal gradients exist that need to be considered and would require a more complete taxa list than is available from the GBRMPA and Reef Check data sets.

Other impediments to the integration of the data sets include the aggregation of GBRMPA data by reporting zone and the low resolution spatial data associated with Reef Check (generic reef names only).

5.2 Usefulness of individual data sets

Whilst the low taxonomic richness of the GBRMPA and Reef Check data sets is a key limiter of their usefulness for report carding purposes, there are indications of factors that might make the individual use of these data sets challenging.

The selection of taxa recorded during surveys is a mixture of iconic species of tourism interest, protected species recovering from fishing or other pressures and those that perform specific ecological functions.

It is notable that some iconic species (for example, Queensland grouper, *E. lanceolatus*) that regularly occur in the GBRMPA and Reef Check data sets have not been recorded at all in the RLS data set. It is suspected that the GBRMPA and Reef Check surveys, being run from commercial dive boats, tend to occur on higher quality reefs where tourist operators know their customers have a higher chance of encountering these species. This frames the data well for assessing .

This issue is further supported by observations that much of the GBRMPA data tends to come from the Wet Tropics Central Zone, and likely from multiple visits to a relatively small number of sites (**Figures 36** and **37**).

In addition to the full taxonomic list, the RLS data set appears to provide a more complete catalogue of taxa across a wider selection of (and likely more diverse) reef habitat.

Analysis of the RLS data set has revealed that the offshore zones generally tend to exhibit higher diversity and abundance of fish than the inshore zones and have distinctly different community structure than the inshore zones. There is also strong evidence of shifts in community structure along a north-south gradient, hence the larger taxa list of the RLS data set becomes important for developing potential indicators.

The usefulness of all data sets is limited by a lack of consistent sampling (annual/biannual) at specific sites. This results in uneven survey effort across the zones as some zones receive little or no survey effort in some years and there tends to be higher sampling effort on offshore reefs in general. Combined with the previously discussed differences in community structure between zones this limits the ability to undertake useful analysis of any of the data sets individually.

5.3 Potential to develop a coral reef fish indicator

For the reasons outlined above, the focus when exploring potential report card indicators has been on the RLS data set. The indicators explored include:

- Observed vs expected taxa richness (unit effort)
- Observed vs expected taxa richness (reference sites)
- Taxa proportions analysis
- Fish community structure

These are discussed in more detail below.

5.3.1 Observed vs expected taxa richness/abundance (unit effort)

Expected taxa richness can be determined for individual reporting zones from the long-term data set, taking for example, median richness as the benchmark and comparing observed richness to generate a report card score.

The key challenge to this approach is the infrequency of sampling and inconsistency of survey sites in each year, particularly within the inshore zones. Species accumulation curves for Wet Tropics and Mackay Whitsunday Isaac offshore zones are still relatively steep, indicating that the full species list has not yet been reached after a decade of sampling (in the case of Wet Tropics).

The curves suggest an additional 4-6 surveys per reporting zone per annum would ideally be required to have confidence in an indicator based on richness/abundance by unit effort. Put simply, insufficient data have been collected by the RLS program thus far to confidently determine expected taxa richness or abundance to use as a benchmark for an indicator.

5.3.2 Observed vs expected taxa richness/abundance (reference)

Comparing observations at a test site with those at a control site is a commonly adopted approach for assessing ecological communities. The challenge is to identify a suitable reference site or sites that are undisturbed and are representative of the communities likely to have existed at the test site prior to disturbance.

However, data analysis indicates that fish community structure varies between reporting zones, even for zones within the same region (**Figure 38)**. Metrics based on expected taxa, taxa richness or community structure would therefore need to be benchmarked against reference data relevant to the specific reporting zone of interest.

Choosing suitable reference sites within each reporting zone is challenging. It is likely that most, if not all inshore zones are to some extent affected by anthropogenic influences and that pristine, unimpacted reference sites may not be available. If such sites do exist, they would need to be surveyed annually to support a report card metric. Ideally, two reference sites would be required in each reporting zone to facilitate assessment in the event of one reference site being affected by cyclone, COTS or other impact. This would require a substantial increase in survey effort and a change of monitoring approach to ensure reference sites were consistently surveyed. It is considered unrealistic in the context of the current RLS program.

An alternative scenario is to use aggregated data, such as the median or 80th percentile of taxa richness /abundance in each reporting zone. However, this approach provides a result that is relative to the overall condition of fish communities within a reporting zone. For example, if all sites within a reporting zone have experienced a decline in taxa richness and/or abundance then the median or 80th percentile for the zone may be low and may not be an appropriate benchmark for determining community health. In addition, the small number of surveys conducted in most inshore zones precludes this approach.

The potential to use the data from the offshore zones, which represent a large proportion of the RLS data set, as a benchmark against which the inshore zones could be assessed has also been examined. The underlying assumption is that offshore zones are less impacted by anthropogenic influences than inshore zones and that all things being equal the fish communities on unimpacted inshore reefs would be similar to those offshore. Unfortunately, this assumption cannot be verified, and it is not known whether the previously described differences between taxa richness, abundance and community structure inshore and offshore are due to impacts on the inshore reef ecosystems or would have occurred naturally as a result of habitat differences.

Community structure shifts moving north to south may suggest the latter is the more likely and would invalidate this assumption.

5.3.3 Taxa proportions analysis

Taxonomic proportions analysis using aggregated data overcomes the need to find individual reference sites that are representative of the zones of interest and to some extent mitigates the risk of single reference sites being lost due to disturbance events (eg coral bleaching, COTS).

However, the need to benchmark the inshore data against reference sites outside of the zone of interest again relies on comparisons with Offshore Zone data and the inherent assumptions and limitations that accompany this approach.

5.3.4 Fish community structure (herbivores)

This is once again a proportions analysis and is subject to many of the limitations and assumptions of the taxa proportions analysis.

Multiple sources have associated healthy coral reef ecosystems with a high abundance of herbivorous fish species. Systems with a reduced herbivore community tend to exhibit higher coverage of macroalgae and associated declines in coral health (eg Bellwood, *et. al*. (2004), Burkepile & Hay (2010), Mumby, *et. al*. (2007), Hughes, *et. al*. (2010) to name a few). However, the relationship is not without some ambiguity, as other studies have shown that a decline in coral health did not necessarily result in a decline in herbivore biomass.

The literature suggests that a healthy and diverse herbivore biomass is necessary for coral health, but that the presence of a healthy and diverse herbivore biomass in itself is not indicative of a healthy coral reef system.

If herbivore proportions were to be used as a report card metric the fish data may need to be considered in conjunction with coral health data, adding to the complexity and labour intensiveness of the analysis. Further, an additional dependency is added; The availability of recent, coral data, ideally from the same reef at which the fish data were collected.

5.3.5 Fish community structure

Mumby, *et. al*. (2008) found that the trophic structure of coral reef fish communities such an increase in the abundance of herbivorous fish or a decline in the abundance of top predators, can be an indicator of ecosystem degradation. Other examples are the loss of predatory fish, which may indicate overfishing, or an increase in smaller, faster growing taxa , which might indicate habitat degradation.

This approach has been tested to a limited extent using herbivore data, due to the availability of a highly relevant reference regarding key herbivore families of interest (Sweatman, H, 2007). The assessment indicates that fish community structure is probably only feasible using the RLS data set, as the nMDS plots show the substantial loss of analytical power when the GBRMPA and Reef Check data sets are used, due to the reduced number of taxa recorded.

The issue of benchmarking once again arises, as it can be seen from the nMDS plot (**Figure 38**) that sites tend to group by reporting zone. Examining shifts in community structure towards or away from those of high-quality reference sites could be useful in developing an indicator, but again, suitable reference sites are unlikely to exist or be sampled with sufficient regularity for this approach to be used.

As a result of these limitations it was deemed that structure analysis using other trophic groups, or indeed proportions analysis using other trophic groups would not add further value to this assessment.

5.4 Parameters Considered But Not Assessed

Consideration was given to stratifying the RLS data set by factors such as distance from coastline, distance from river mouth or conservation/general use zoning. However, it was clear that the issues of inconsistent sampling effort and survey site selection that were pervasive throughout the above assessments would be further exacerbated by stratifying the data in this manner. It was deemed very unlikely to provide further information or options given low survey numbers and not worthy of the additional work required to stratify the data.

6 Conclusions

6.1 Indicator Development

The purpose of this evaluation was firstly to assess whether the three data sets (RLS, GBRMPA and Reef Check) together or independently contain sufficient data to support the development of a coral reef fish indicator for the regional report cards. If so, the aim was to identify a likely indicator metric for further development and establish data management protocols such that annual indicator calculation would be efficient.

It is important to note that the fish surveys undertaken through the RLS, GBRMPA and Reef Check programs were designed and implemented before the regional report cards existed and as such were never intended to support the types of data analysis required for developing report card indicators. As a result, attempts to utilise the data for this purpose were met with some limiting factors:

- The analytical (and data integration) options were constrained by the data collection protocols, and in particular the recording of a limited number of key taxa in GBRMPA and Reef Check surveys.
- A suspicion (supported by the analysis) that GBRMPA and Reef Check surveys favour high quality reef habitat where iconic species are likely to be overrepresented.
- Inconsistent survey effort resulting in temporal and spatial patchiness of data and lack of regular survey replication at any given site. This is likely reflective of the opportunistic nature of citizen science surveys and the lack of consistent funding often experienced by citizen science programs.
- Insufficient sampling effort within any data set, particularly at inshore zones, to provide sufficient confidence for developing metrics.
- Inability to integrate the three data sets to overcome any of the above limitations, due to the overall loss of taxonomic data suffered when data sets are restricted to only indicator taxa.

It was determined that the RLS data set offered the best prospects of developing a fish indicator, being a complete inventory of observed taxa and their abundance and less influenced by tourism objectives.

Despite the remaining limitations still applying to the RLS data set (ie inconsistent survey effort and insufficient survey events), it was decided to explore some of the more commonly used coral reef fish metrics using statistical techniques. The statistical assessment presented here is not purported to be exhaustive, as there are dozens more options that could be explored. Rather, we have identified and tested some of the more likely options for a report card indicator and tested the value of proceeding further, plus identified future opportunities for the data sets.

In terms of report card indicator development, the following observations have been made regarding the RLS data set:

- The best prospects of developing an indicator are offered by the RLS data set alone. However, the GBRMPA and Reef Check data sets may provide useful narrative to support the report cards.

- There are clear differences in fish community structure, taxa richness and total abundance between north and south, inshore and offshore. This can be evident even within reporting zones.
- The development of indicators around taxa richness, abundance or proportions are hampered by four factors:
    - Insufficient survey effort to support this style of indicator, with many zones not being surveyed each year.
    - The degree of variation between inshore and offshore zones makes the use of offshore data as benchmarks of inshore community health dubious.
    - The nature of inshore zones being that identifying reference sites that reflect high quality undisturbed communities unlikely.
    - In the event that reference sites could be identified the infrequent sampling is likely to result in a lack of reference data on many occasions.
- The development of report card indicators based on trophic structure shows some promise, but is again limited by the survey effort and inconsistent monitoring frequencies.

6.2 Integration With Future AIMS/JCU Fish Data

The RLS data set, and to a lesser extent the GBRMPA and Reef Check data sets, have potential to support, enhance and add further value to the long-term fish monitoring program being developed by the AIMS/JCU partnership in a number of ways. The report card network has ongoing dialogue with the leaders of this project and it is anticipated that indicator development will commence as data become available for analysis. The findings of this assessment will be useful to inform the indicator development process and will likely present opportunities to more fully utilise the citizen science data.

The following opportunities will be worth exploring once the nature of the monitoring program becomes known and given different sampling methodologies would likely require a couple of years of program overlap in order to calibrate data.

6.2.1 Temporal context

Depending on the nature of the metrics developed by the AIMS/JCU consortium, the long-term citizen science data sets assessed herein have the potential to provide information regarding variability in taxa richness, total abundance and community structure prior to commencement of the new monitoring program. For example, advances in technology have seen an increase in the deployment of baited remote underwater video (BRUV) by citizen science groups and there is potential for these programs to augment the AIMS/JCU data spatially or temporally.

6.2.2 Spatial integration

Should AIMS/JCU choose metrics based on the abundance of indicator taxa and should those taxa align with those utilised by GBRMPA or Reef Check there may be potential to combine data, which could overcome some limitations of the citizen science data sets and provide a more useful report card metric.

If the AIMS/JCU program instead focusses on community structure, trophic structure or other metrics there is potential for the RLS data set to become more useful in future. Once again, the two data sets could possibly be combined to increase the total survey effort, number of sites per reporting zone and so on, notwithstanding the differences in survey methodology between the two programs.

6.2.3 Communications outcomes

Given existing networks and connectivity with regional communities, the report card partnerships should be viewed as an important conduit for the dissemination of information from the AIMS/JCU program. The use of citizen science fish data to support and value-add to the AIMS/JCU data set would further enhance the communications opportunities.

6.3 Other Opportunities

6.3.1 Supporting narrative

Whilst the citizen science data sets are not currently sufficient for indicator development, they do offer some opportunities to enhance narrative around the other report card results or to create other communications tools and products of value to the partnerships. Some interesting observations that could be used for these purposes include:

- After 8 years of RLS surveys of reef fish the species accumulation curves have not yet plateaued. This shows that long-term monitoring is critical to understanding taxa richness and community structures.
- nMDS plots illustrate fine scale spatial changes in fish communities, highlighting the complexities of coral reef fish indicator development and the incredible diversity of the reef system. Each reporting zone within the regions is essentially unique in terms of the fish communities it supports.
- Matching herbivore abundance/proportion data with coral health data may provide opportunities to highlight the dependence of coral reefs on herbivore communities

There are numerous other interesting stories that could be pulled from the existing analysis or from additional analysis of the now cleaned data set.

This sort of information is useful for report cards, stewardship reports, websites and might also potentially be displayed as environmental dashboards.

6.3.2 Refinement of citizen science programs

This data analysis, whist cursory, has identified limitations of the citizen science data sets that could potentially be addressed in future fish surveys. Some opportunities that could be explored are:

- Coordination between citizen science programs might reduce the incidence of duplication of effort in some reporting zones and lack of surveys in others both spatially and temporally.
- Achieving consistent (annual, or at least biannual) of a specific site or sites within each report card reporting zone.
- Refinement, standarisation and potentially increase of, taxa recorded during GBRMPA and Reef Check surveys.

7 References

Bellwood, D. R., Hughes, T. P., Folke, C., & Nyström, M. (2004). Confronting the coral reef crisis. Nature, 429(6994), 827-833. doi: 10.1038/nature02691

Burkepile, D. E., & Hay, M. E. (2010). The trophic roles of herbivorous fishes on coral reefs. Trends in Ecology & Evolution, 25(6), 317-325. doi: 10.1016/j.tree.2010.01.010

Carpenter, K. E., & Alevizon, E. (1999). A review of coral reef monitoring indicators. In Coral reefs and fisheries in the Pacific island countries (pp. 31-47). Apia, Samoa: SPREP.

Donahue, M. J., & Graham, N. A. J. (2010). Ecosystem structure and herbivore abundance on Caribbean coral reefs. Marine Ecology Progress Series, 411, 63-72.

Done, T. J. (1992). Phase shifts in coral reef communities and their ecological significance. Coral Reefs, 11(2), 121-132.

Feary, D. A., Saenz-Agudelo, P., Wachenfeld, D. R., Graham, N. A. J., & Berumen, M. L. (2014). Herbivore biomass and functional group composition on coral reefs: The influence of coral loss. Marine Ecology Progress Series, 505, 49-62.

Hoogenboom, M. O., van Rijssel, J. C., & Vermeij, M. J. A. (2015). Large herbivores protect coral reefs by controlling algal overgrowth. Scientific Reports, 5, 9691.

Hughes, T. P., Graham, N. A., Jackson, J. B., Mumby, P. J., & Steneck, R. S. (2010). Rising to the challenge of sustaining coral reef resilience. Trends in Ecology & Evolution, 25(11), 633-642. doi: 10.1016/j.tree.2010.07.011

Krishnan, P., George, G., Immanuel, T., & Bitopan-Malakar, M. (2013). Studies on the Recovery of Bleached Corals. In K. Venkataraman et al. (Eds.), Ecology and Conservation of Tropical Marine Faunal Communities: Fishes as Indicators of Reef Health (Chapter 25). Springer-Verlag Berlin Heidelberg. doi: 10.1007/978-3-642-38200-0_25.

Mumby, P. J., Dahlgren, C. P., Harborne, A. R., Kappel, C. V., Micheli, F., Brumbaugh, D. R., & Steneck, R. S. (2008). Assessing the health of coral reefs: the value of different physical and biological metrics. Coral Reefs, 27(4), 765-778. doi: 10.1007/s00338-008-0429-z

Mumby, P. J., Hastings, A., & Edwards, H. J. (2007). Thresholds and the resilience of Caribbean coral reefs. Nature, 450(7166), 98-101. doi: 10.1038/nature06252

Sweatman, H. (2007) Coral reef health indicators and thresholds of concern. Unpublished report to the Marine and Tropical Sciences Research Facility. Reef and Rainforest Research Centre Limited, Cairns (31 pp.).

Szczebak, J. M., Babcock, R. C., & Ayling, A. M. (2013). Characterising herbivorous fish functional groups and their impact on coral reefs. Coral Reefs, 32(3), 753-765.

APPENDIX A

Data Overlap Between Reef Check and GBRMPA

Group	Common Name	Family	Genus	Species	Reef Check	GBRMPA (tourism)
Angelfish	Six-Band Angel	Pomacanthidae	Pomacanthus	sexstriatus		✓
Butterflyfish	-	Chaetodontidae	all	all	✓	
	Coral Rabbitfish	Chaetodontidae	Siganus	corallinus		✓
	Moorish Idol	Chaetodontidae	Zanclus	cornutus		✓
Damselfish	Blue-Green Chromis	Pomacentridae	Chromis	viridis		✓
Drums and Croakers	Queenfish	Sciaenidae	Scomberoides	commersonnianu s		✓
Eels	Moray eel	Muraenidae	all	all	✓	✓
Emperors	-	Lethrinidae	all	all	✓	
Groupers and Rockcods	Barramundi cod	Epinephelidae	Cromileptes	altivelus	✓	✓
	Common Coral Trout	Serranidae	Plectropomus	leopardus		✓
	Coral Cod	Epinephelidae	Cephalopholis	miniata		✓
	Coral Trout	Serranidae	all	all	✓	
	Flowery Cod	Serranidae	Epinephelus	fuscoguttatus		✓
	Greater grouper	-	-	-	✓	
	Peacock Rockcod	Serranidae	Cephalopholis	argus		✓
	Queensland Grouper	Serranidae	Epinephelus	lanceolatus	✓	✓
Grunters and Sweetlips	Giant sweetlip	Haemulidae	Plectorhinchus	albovittatus		✓
	Painted sweetlip	Lethrinidae	Diagramma	pictum		✓
	Sweetlips	-	-	-	✓	
	Yellowband sweetlip	Haemulidae	Plectorhinchus	lineatu		✓
Mackerels	Shark Mackerel	Scombridae	Grammatorcynus	bicarinatus		✓
Parrotfish	Bumphead parrotfish	Scaridae	Bolbometopon	muricatum	✓	✓
	Parrotfish	Scaridae	all	all	✓	
Pufferfish	Starry Puffer	Tetraodontidae	Arothron	stellatus		✓
Sharks and Rays	Blacktip Reef Shark	Carcharhinidae	Carcharhinus	melanopterus		✓
	Blue-Rib Tail Stingray	Dasyatidae	Taeniura	lymma		✓
	Cowtail Ray	Potamotrygonidae	Potamotrygon	sephen		✓
	Epaulette Shark	Hemiscylliidae	Hemiscyllium	ocellatum		✓
	Grey Reef Whaler	Carcharhinidae	Carcharhinus	amblyrhynchus		✓
	Kuhl's Stingray	Dasyatidae	Dasyatis	kuhlii		✓
	Leopard Shark	Triakidae	Stegostoma	fasciatum		✓
	Manta Ray	Mobulidae	Manta	birostris		✓
	Shovelnose Ray	Rhinobatidae	Rhinobatos	typus		✓
	Spotted Eagle Ray	Myliobatidae	Aetobatus	narinari		✓
Snakeheads	Milkfish	Chanidae	Chanos	chanos		✓
Snappers	-	-	all	all	✓	
	Paddletail snapper	Lutjanidae	Lutjanus	gibbus		✓
	Pink snapper	Sparidae	Pagrus	auratus		✓
	Red Bass	Lutjanidae	Lutjanus	bohar		✓
Surgeonfish	Lined Surgeonfish	Acanthuridae	Acanthurus	Lineatus		✓
	Surgeonfish	Acanthuridae	-			✓
Trevallies	Bigeye Trevally	Carangidae	Caranx	sexfasciatus		✓
	Bluefin Trevally	Carangidae	Caranx	melampygus		✓

	Giant Trevally	Carangidae	Caranx	Ignobilis		✓
	Golden Trevally	Carangidae	Gnathanodon	speciosus		✓
Wrasses and Tuskfish	-	-	-	-	✓	
	Checker Wrasse	Labridae	Halichoeres	hortulanus		✓
	Cleaner Wrasse	Labridae	Labroides	dimidiatus		✓
	Harlequin tuskfish	Labridae	Choerodon	fasciatus		✓
	Humphead/Maori Wrasse	Labridae	Cheilinus	undulatus	✓	✓
	Moon Wrasse	Labridae	Thalassoma	lunare		✓
	Sixbar Wrasse	Labridae	Thalassoma	hardwicke		✓
	Slingjaw Wrasse	Labridae	Epibulus	Insidiator		✓

APPENDIX B

SUMMARY OF METRICS AND ANALYSES

Analysis	Data	Description	Analysis/Visualisation	Comments/Outcome
Sampling efficiency				
Species Accumulation	RLS	Measure of the effectiveness/efficiency of sampling	Univariate. Species curve visualisation.	• Requires all taxa to be recorded, so couldn't be performed on GBRMPA and Reef Check data. • Curves have not plateaued over 5 years (MWI) and 10 years (WT), indicating that the number of sampling sites and/or frequency of sampling are generally not sufficient for the diversity of the system.
Observed v Expected Taxa Richness				
Spatial Community Shifts	RLS	Uses nMDS (Bray-Curtis similarity) to see whether fish communities are similar between locations.	Bray-Curtis similarity (log transform) nMDS visualisation	• Shows that coral reef fish communities are different between the WT, DT and MWI regions, both for inshore and offshore reefs. • Creates challenges finding "reference" sites for benchmarking a fish indicator.
Observed vs expected (unit effort)	RLS, GBRMPA, RC	Standardises the data against the time/effort to conduct surveys and allows sites to be compared regardless of uneven sampling effort	Univariate. Boxplot visualisation. Bray-Curtis similarity (log transform) Permanova. Tabular presentation	• Simplifies comparison between sites/sampling events. • More applicable for monitoring trends over time, rather than "snapshot" health assessment.
Observed vs expected (control/reference sites)	RLS, GBRMPA, RC	Examines whether relatively unimpacted offshore reefs could be benchmarks for the health of inshore reefs. Assesses whether only recording key species is sufficient for an indicator based on taxa richness.	Univariate. Boxplot visualisation. Bray-Curtis similarity (log transform) Permanova. Tabular presentation nMDS and boxplot visualisation	• Strategy was explored in the absence of suitable inshore reef reference sites. Numerous limitations of comparing offshore and inshore reef fish communities are acknowledged. • Data sets based on key indicator species only were confirmed unsuitable for this style of analysis • Highlights challenges of "reference site" approach to benchmarking fish taxa richness/abundance
Taxa Richness and Total Abundance	RLS	Assesses whether there are differences in the diversity of species or total fish abundance spatially and temporally,	Univariate. Boxplot visualisation.	• Provides assessment of the variability f diversity and abundance spatially and temporally. • Highlights naturally high spatial and temporal variability. • Highlights extreme outliers.

Analysis	Data	Description	Analysis/Visualisation	Comments/Outcome
Fish Community Composition				
Taxa Proportions Analysis	RLS	Assesses key indicator taxa as a proportion of total taxa richness	BioEnv. Tabular visualisation	• Reduces need to use control or reference sites. • Requires clear understanding of the relationship of indicator taxa to reef health, challenging in very diverse ecosystems
Taxa Proportions Analysis - herbivores	RLS	Assesses herbivore taxa as a proportion of total taxa richness	Univariate. Boxplot visualisation.	• Reefs with healthy coral have a high proportion of herbivorous fish species. • Some studies found that poor coral health does not result less herbivorous fish.
Fish Community Structure	RLS, GBRMPA	Assesses similarity between sites and over time by functional role.	Bray-Curtis similarity (log transform) nMDS and bubbleplot visualisation	• Changes in fish community structure can indicate alterations in habitat quality and availability. • May be more suitable for monitoring sites over time.