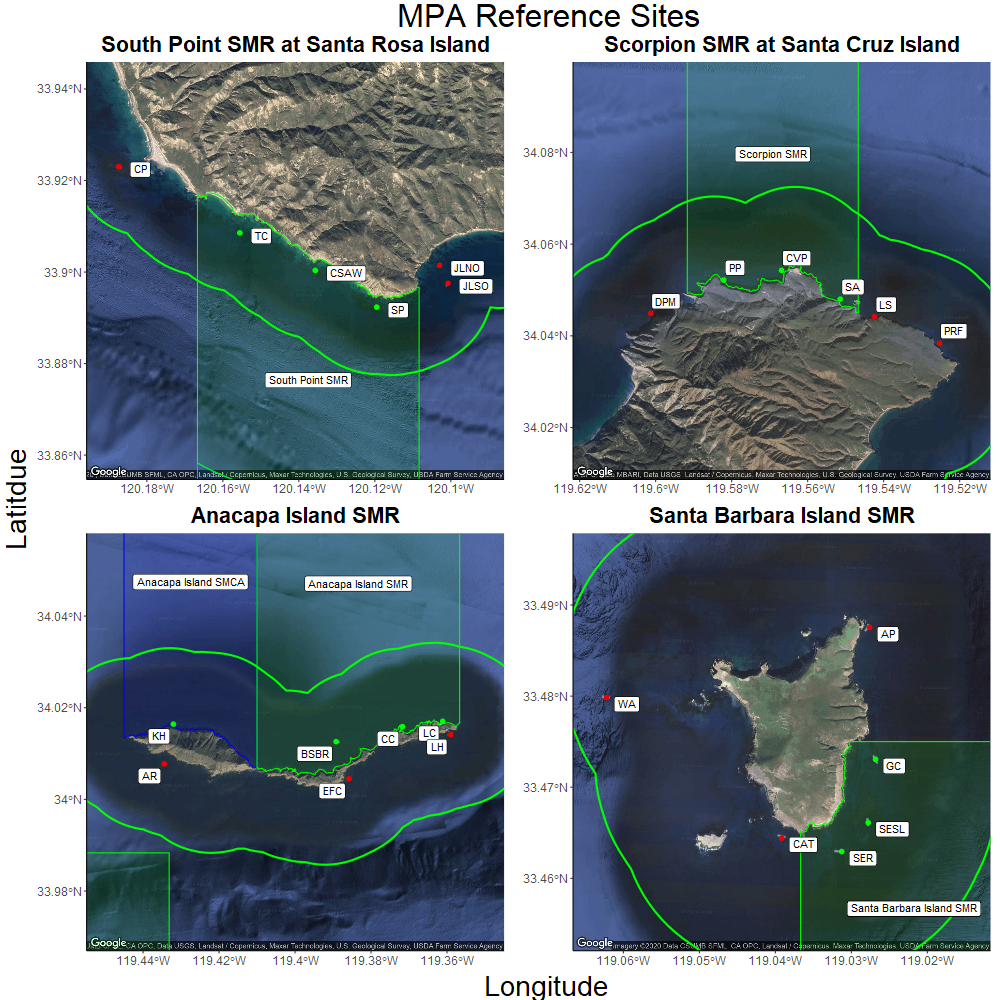
# Methods

## Sampling Locations

Paragraph describing Selected MPA study sites

 Fig 2. South Point SMR at Santa Rosa Island (top left), Scorpion SMR at Santa Cruz Island (top right), Anacapa Island SMR at Anacapa Island (bottom left), and Santa Barbara Island SMR at Santa Barbara Island (bottom right). Within-MPA sites indicated with green points, and outside-MPA “reference” sites indicated with red points. Locations of SMRs relative to the Channel Islands indicated in insets.

## Survey Methods

The Kelp Forest Monitoring Program employs many different survey methods to characterize benthic and fish communities, as well as the surrounding physical environment. Each methodology is aimed at best capturing the most precise and accurate representation of the abundance of the 71 “indicator species” and all additional species of fish. Size measurements of certain invertebrates and size estimates of most species of fish give an idea of the size frequency distribution for a select list of species.

Given that all surveys are conducted on SCUBA at depths ranging from 4 m to 18 m, there are logistical considerations of data collection that must be taken into account when examining the locations, areas, times, and organisms that were sampled. Since the monitoring program’s inception, monitoring has taken place between May and October each year, with the exception of 2020 due to COVID-19. Each site has a permanent 100 m transect cable used to define the center of the 2,000 m² survey area.

There are seven survey methods which will be are use for this report: 1 m² quadrats, 5 m² quadrats, Band Transects, Roving Diver Fish Counts (RDFC), Fish Size Frequency (FSF), Random Point Contacts (RPC), and Natural Habitat Size Frequency (NHSF). The 1 m², 5 m², and Band surveys all characterize the abundance or percent cover of sedentary and mobile invertebrate species, cryptic fish, and algae on the benthos. NHSF surveys, yield a size distribution for select invertebrates as well as giant kelp. These size distributions are then used to estimate biomass with a length to weight formula unique to each species (see L-W calculation table with sources). RDFC surveys and FSF surveys characterize the abundance and size frequency of fishes found in all habitats (cryptic, water column and canopy dwelling). From these data, fish biomass can be estimated with a length to weight formula unique to each species (see L-W calculation table with sources …MAKE ME…). 21 species of fish are chosen to represent the fish populations for biomass calculations. These 21 species were chosen if they met one or more of the following criteria:

1. Prolific
2. Commonly targeted
3. Charismatic
4. Herbivorous

In 1 m² Quadrats, two divers sample a 2 m² area, at twelve equidistant, random meter numbers along the 100 m main transect line. Since each quadrat covers an area of 2 m², a total of 24 m² are covered between the two divers. Mean densities can be derived for each species on 1 m² Quadrats by dividing the count by the area surveyed. We multiply the mean density by 2,000 to get an estimated site level count. See table \_\_\_\_\_ for a list of species monitored by 1 m² Quadrats.

In 5 m² Quadrats, two divers sample forty 1 m wide by 5 m long quadrats, side by side, parallel and tangential to the 100 m main transect line. Since each quadrat covers an area of 5 m², a total of 200 m² are covered between the two divers. Mean densities can be derived for each species on 5 m² Quadrats by dividing the count by the area surveyed. We multiply the mean density by 2,000 to get an estimated site level count. See table \_\_\_\_\_ for a list of species monitored by 5 m² Quadrats.

In Band Transects, two divers sample a 60 m² area, at twelve equidistant, random meter numbers along the 100 m main transect line. Since each quadrat covers an area of 60 m², a total of 720 m² are covered between the two divers. Mean densities can be derived for each species on Band Transects by dividing the count by the area surveyed. We multiply the mean density by 2,000 to get an estimated site level count. See table \_\_\_\_\_ for a list of species monitored by Band Transects.

In RPCs, one surface supplied diver randomly samples 40 points, at fifteen equidistant, random meter numbers along the 100 m main transect line. Since each random meter has 40 points, a total of 600 points are covered by the diver. Percent cover can be derived for each species on RPCs by dividing the count by the total number of points and multiplying by 100. See table \_\_\_\_\_ for a list of species monitored by RPCs.

In RDFC, a minimum of three divers sample the entire transect footprint of 2,000 m², from the benthos to the canopy, in 30 minutes. Since each diver covers the entire site, a total of 2,000 m² are covered by each diver. Mean densities can be derived for each species on RDFC by dividing the sum of the counts by the total number of divers conducting the survey. See table \_\_\_\_\_ for a list of species monitored by RDFC.

In FSF, one diver samples within the entire transect footprint of 2,000 m², from the benthos to the canopy, with no time limit. Since the diver covers the entire site, a total of 2,000 m² is covered by the diver. Size frequency distributions for each species on FSF are estimated sizes to the nearest cm. All fish species can be entered for FSF, though typically small cryptic fish such as goby sizes are not estimated.

In NHSF, any number of divers sample randomly within transect footprint of 2,000 m², with no time limit. Since the diver samples randomly within the entire site so there is no specific area covered. Size frequency distributions for most species on NHSF are measured with vernier calipers to the nearest mm. Giant kelp stipes are counted per plant at 1 m above the benthos. Gorgonians are measured at the widest point to the nearest cm. See table \_\_\_\_\_ for a list of species monitored by NHSF

Some species are counted on multiple sampling protocols. Typically these species are added to a new protocol to better asses their density based on an evaluation of the previous methods ability to detect changes for those species as well as to select the “best” protocol to effectively search their habitat. These species continue to be counted on the previous protocol to provide consistency. For this analysis we selected the “best” survey method for each species for each year to include in our count data. The species which are also counted on RPCs were left in for the Simpson index calculation, but were not used for the random forest model.

Table XX. Protocols used for species monitored on more than one protocol.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Taxa/Common Name | Scientific Name | 1 m quadrats | 5 m quadrats | band transects | RPCs |
| Oar weed | Laminaria farlowii | 1982-2019 | NA | NA | X |
| Giant kelp | Macrocystis pyrifera | 1985-1995 | 1996-2019 | NA | X |
| California sea palm | Pterygophora californica | 1982-2019 | NA | NA | X |
| Southern sea palm | Eisenia arborea | 1982-2019 | NA | NA | X |
| Devil weed | Sargassum horneri | 2010-2019 | X | X | X |
| Wakame | Undaria pinnatifida | 2016-2019 | X | X | X |
| Giant-spined sea star | Pisaster giganteus | 1982-1995 | 1996-2013 | 2014-2019 | NA |
| Ochre sea star | Pisaster ochraceus | NA | 1996-2013 | 2014-2019 | NA |

Year range is the years the protocol was selected for the species, X indicates it is also counted on that protocol, and NA indicates that species is not monitored on that protocol.

Table XX. Regularly monitored species, substrate, and associated monitoring technique(s). Monitoring techniques are described in Davis et al., 1997, and Kushner and Sprague, in progress.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group | Taxa/Common Name | Scientific Name | 1 m | 5 m | Bands | RPCs | NHSF | VFT | RDFC |
| Algae | Miscellaneous green algae | NA | NA | NA | NA | X | NA | NA | NA |
| Algae | Miscellaneous red algae | NA | NA | NA | NA | X | NA | NA | NA |
| Algae | Articulated coralline algae | NA | NA | NA | NA | X | NA | NA | NA |
| Algae | Encrusting coralline algae | NA | NA | NA | NA | X | NA | NA | NA |
| Algae | Agar weed | Gelidium spp. | NA | NA | NA | X | NA | NA | NA |
| Algae | Sea tongue | Gigartina spp. | NA | NA | NA | X | NA | NA | NA |
| Algae | Miscellaneous brown algae | NA | NA | NA | NA | X | NA | NA | NA |
| Algae | Acid weed | Desmarestia spp. | NA | NA | NA | X | NA | NA | NA |
| Algae | Oar weed | Laminaria farlowii | X | NA | NA | X | NA | NA | NA |
| Algae | Bladder chain kelp | Cystoseira spp. | NA | NA | NA | X | NA | NA | NA |
| Algae | Giant kelp | Macrocystis pyrifera | X | X | NA | X | NA | NA | NA |
| Algae | California sea palm | Pterygophora californica | X | NA | NA | X | NA | NA | NA |
| Algae | Southern sea palm | Eisenia arborea | X | NA | NA | X | NA | NA | NA |
| Algae | Devil weed | Sargassum horneri | X | X | X | X | NA | NA | NA |
| Algae | Wakame | Undaria pinnatifida | X | X | X | X | NA | NA | NA |
| Algae | Miscellaneous plants | NA | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Miscellaneous sponges | NA | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Orange puffball sponge | Tethya aurantia | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Southern staghorn bryozoan | Diaperoecia californica | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Miscellaneous bryozoans | NA | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | California hydrocoral | Stylaster californica | NA | NA | X | NA | X | NA | NA |
| Invertebrates | White-spotted rose anemone | Tealia lofotensis | NA | NA | X | NA | NA | NA | NA |
| Invertebrates | Red gorgonian | Lophogorgia chilensis | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Brown gorgonian | Muricea fruticosa | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Californian golden gorgonian | Muricea californica | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Strawberry anemone | Corynactis californica | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Orange cup coral | Balanophyllia elegans | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Cup coral | Astrangia lajollaensis | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Ornate tube worm | Diopatra ornata | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Colonial sand-tube worm | Phragmatopoma californica | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Scaled-tube snail | Serpulorbis s1 m Quadsuamigerus | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Chestnut cowrie | Cypraea spadicea | X | NA | NA | NA | NA | NA | NA |
| Invertebrates | Wavy turban snail | Megastraea undosa | X | NA | NA | NA | X | NA | NA |
| Invertebrates | Red turban snail | Astraea gibberosa | X | NA | NA | NA | X | NA | NA |
| Invertebrates | Bat star | Patiria miniata | X | NA | NA | NA | X | NA | NA |
| Invertebrates | Giant-spined sea star | Pisaster giganteus | X | X | X | NA | X | NA | NA |
| Invertebrates | Ochre sea star | Pisaster ochraceus | NA | X | X | NA | X | NA | NA |
| Invertebrates | Sunflower star | Pycnopodia helianthoides | X | NA | X | NA | X | NA | NA |
| Invertebrates | White sea urchin | Lytechinus anamesus | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Red sea urchin | Strongylocentrotus franciscanus | X | NA | NA | NA | X | NA | NA |
| Invertebrates | Purple sea urchin | Strongylocentrotus purpuratus | X | NA | NA | NA | X | NA | NA |
| Invertebrates | Warty sea cucumber | Parastichopus parvimensis | X | NA | NA | NA | NA | NA | NA |
| Invertebrates | Aggregated red sea cucumber | Pachythyone rubra | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Red abalone | Haliotis rufescens | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Pink abalone | Haliotis corrugata | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Green abalone | Haliotis fulgens | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Kellett<U+0092>s whelk | Kelletia kelletii | NA | NA | X | NA | X | NA | NA |
| Invertebrates | Giant keyhole limpet | Megathura crenulata | NA | NA | X | NA | X | NA | NA |
| Invertebrates | California brown sea hare | Aplysia california | NA | NA | X | NA | NA | NA | NA |
| Invertebrates | Rock scallop | Crassedoma giganteum | NA | NA | X | NA | X | NA | NA |
| Invertebrates | California spiny lobster | Panulirus interruptus | NA | NA | X | NA | NA | NA | NA |
| Invertebrates | Tunicates | NA | NA | NA | NA | X | NA | NA | NA |
| Invertebrates | Stalked tunicate | Styela montereyensis | X | NA | NA | NA | NA | NA | NA |
| Invertebrates | Miscellaneous invertebrates | NA | NA | NA | NA | X | NA | NA | NA |
| Fish | Bluebanded goby | Lythrypnus dalli | X | NA | NA | NA | NA | NA | NA |
| Fish | Blackeye goby | Coryphopterus nicholsii | X | NA | NA | NA | NA | NA | NA |
| Fish | Island kelpfish | Alloclinus holderi | X | NA | NA | NA | NA | NA | NA |
| Fish | Blacksmith | Chromis punctipinnis | NA | NA | NA | NA | NA | X | X |
| Fish | Seorita | Oxyjulis californica | NA | NA | NA | NA | NA | X | X |
| Fish | Blue rockfish | Sebastes mystinus | NA | NA | NA | NA | NA | X | X |
| Fish | Olive rockfish | Sebastes serranoides | NA | NA | NA | NA | NA | X | X |
| Fish | Kelp rockfish | Sebastes atrovirens | NA | NA | NA | NA | NA | X | X |
| Fish | Kelp bass | Paralabrax clathratus | NA | NA | NA | NA | NA | X | X |
| Fish | California sheephead | Semicossyphus pulcher | NA | NA | NA | NA | NA | X | X |
| Fish | Black surfperch | Embiotoca jacksoni | NA | NA | NA | NA | NA | X | X |
| Fish | Striped surfperch | Embiotoca lateralis | NA | NA | NA | NA | NA | X | X |
| Fish | Pile perch | Damalichthys vacca | NA | NA | NA | NA | NA | X | X |
| Fish | Garibaldi | Hypsypops rubicundus | NA | NA | NA | NA | NA | X | X |
| Fish | Opaleye | Girella nigricans | NA | NA | NA | NA | NA | NA | X |
| Fish | Rock Wrasse | Halichoeres semicinctus | NA | NA | NA | NA | NA | X | X |
| Substrate | Bare | NA | NA | NA | NA | X | NA | NA | NA |
| Substrate | Rock | NA | NA | NA | NA | X | NA | NA | NA |
| Substrate | Cobble | NA | NA | NA | NA | X | NA | NA | NA |
| Substrate | Sand | NA | NA | NA | NA | X | NA | NA | NA |

Table XX. Protocol survey area and minimum number of divers.

|  |  |  |
| --- | --- | --- |
| Protocol | Area Surveyed | Divers |
| 1 m Quadrats | 24 sqm | 2 |
| 5 m Quadrats | 200 sqm | 2 |
| Band Transects | 720 sqm | 2 |
| RPCs | 600 points | 1 |
| NHSF | up to 2000 sqm | 1 |
| RDFC | 2000 sqm | 3 |
| FSF | up to 2000 sqm | 1 |

## Data Analyses and Calculations

In these analyses, we began by analyzing the trends within benthic and fish communities at KFMP MPA reference sites since 2005, using community diversity and community similarity, then used a machine-learning Random Forest model, and Indicator Species Analysis to identify the variables most strongly associated with MPA protection, and which species uniquely characterized each set of communities. We then compared the species identified by each approach, and analyzed biomass or density trends over time for these important species, and also compared biomass or density ratios (inside/outside MPAs) for targeted and non-targeted benthic and fish species.

For the benthic community, we compiled data from each benthic survey type described above - 1m quadrats, 5m quadrats, and band transects, and eliminated redundant species between survey types, using only the “best” survey for each species. We then corrected for sampling areas between survey types by calculating estimated total site abundances, with Roving Diver Fish Counts, however, occur over the entirety of the site, so we included these counts directly. Hereafter, this dataset of “site-level” counts for each benthic and fish species, at each site, for each year, is referred to as “community data”. Since Random Point Contact surveys (RPCs) are frequency, and not count data, we excluded substrate types (rock, sand, etc) and calculated percent cover for all species and taxonomic categories.

To test for differences in diversity between islands and MPA protection status, we calculated Shannon Index scores from our community dataset (see “Equations Used” below) for each site and year combination, using the “vegan” package in R (Oksanen et al. 2019). Similarly, we calculated Simpson’s index (similar to Shannon Index, commonly used for percent cover data) from the RPCs data to analyse algal diversity. To test whether MPA reserve status, island, their interaction, or seawater temperature anomaly had a significant effect on community diversity over time, we used a pair of generalized linear mixed models with Shannon or Simpson Indicies as the independent variables, Reserve Status, Island, Oceanic Nino Index (ONI), and the interaction of Reserve Status and Island as fixed effects, and Survey Year as a random effect, using the “lme4” package in R (Bates et al. 2015). From our community dataset, we then conducted non-metric multidimensional scaling (nMDS) to visualize the structure and grouping of site communities, and used analysis of similarity (ANOSIM) to test for significant differences in groupings between islands. To conduct these analyses, we used the “vegan” package in R (Oksanen et al. 2019).

Random Forest analysis is a machine-learning approach based on classification, or “decision” trees, which operate by repeatedly splitting data into two groups that are as homogeneous as possible, selecting the predictor which best discriminates amongst the data at each split, or “node”, and combines the results from this hierarchal process into a “tree”. To measure which variables result in splits that are homogoneous as possible at each node, or minimize “impurity”, the model uses the Gini index, which reflects the proportion of responses in each level of a categorical variable - if the observations fall into few categories, the Gini index is low. Because even a single classification tree will likely overfit the data, The Random Forest method uses an aggregation of many individual classification trees, created using bootstrapped resamplings and the process described above, assigns a class for each observation based on which leaf it fell in, and after repeating this process 1000-5000 times, calculates the proportion of times each observation is assigned to a given class, and makes a final assignment for each observation based on the majority proportion (Lefcheck, 2015). Since the model is derived from repeated bootstrap resampling, the set of data not selected in the bootstrap, known as the “out of bag” sample, is used to validate the tree by tallying the number of misclassifications, resulting in an out-of-bag error rate, which can be interpreted similarly to measure of model “fit”. Random Forest analyses return predictor variables ranked by their Gini Index (measure of impurity), as well as the decrease in accuracy, which measures how much worse the model performs without that variable included. Since Random Forests are a “black box” type of model, their process is often visualized using Partial dependence plots, which display the range of the predictor variable along the x axis, and the line converges either at the top or bottom of the plot, depending on the level of the response variable. These curves can be interpreted to show the values at which a predictor variable becomes associated with levels of the response variable. We used a random forest analysis on our fish and benthic community data, using biomass, or counts for species which did not have computed biomasses, as well as the RPC data measuring percent benthic cover. Since the analysis automatically accounts for the interactions between predictor variables based on the heirarchal nature of the classification tree, we also included Shannon and Simpson’s indicies as measures of diversity, Island, and survey year. This process allowed us to determine the relative importance of each variable for accurately predicting reserve status based on the data, the corollary of which indicates the species most correlated with reserve status.

Following our Random Forest analysis, we also used Indicator Species Analyses (ISA), on the inside-MPA and outside-MPA data, respectively, to identify the species significantly associated with each island and MPA protection grouping, for each year. ISA assigns strength of association and significance values by comparing species occurrence to a permuted association level likely to occur by chance (De Caceres and Legendre, 2009). One of the advantages of ISA is that it accounts for both the abundance and frequency of species, is calculated separately for each species in the assemblage, and can be applied to many experimental designs, including a priori classifications like levels of a categorical variable (Bakker 2008, Dufrêne & Legendre 1997; McGeoch & Chown 1998). We then compiled the results from ISA from each year, and assembled paired heatmaps of inside- and outside-MPA data for each island, where a given species in a given year displays as red if it was not significant, or if significant displays as shades of green indicating the strength of association statistic. Thus, these heatmaps identify and visualize the trends in species importance over time associated with changes in community diversity and similiarity. Rather than the Random Forest model approach, where in order to identify the most important variables associated with Reserve Status, all data were tested as predictor variables in a model with Reserve Status as the response variable, the ISA allows for nested factors and returns series of indicator species for inside and outside MPA communities at each island. These two approaches combined provide better insight into which species are important in terms of MPA effects – the random forest, machine learning approach allows the widest range of data to be used and provides insight into the relative strength of association with Reserve Status between many variables, while the ISA returns results for fish and benthic species specific to each combination of Island and Reserve Status, allowing for more direct insight into the changing community structures at each Reserve Status and Island, and how these trends compare to those shown in community diversity and similarity over time. We conducted ISA tests using the “indicspecies” package in R (De Caceres and Legendre, 2009).

Following our community-level analyses, we calculated biomass (see “Equations Used” below) for several species identified as significant from Random Forest and ISA, as well as during our literature review (eg. kelp declines, urchin proliferation, SSWD-associated star declines, and abalone declines, spiny lobster fishery effects), as well as some “iconic” species (eg Garibaldi, California Sheephead). To test whether MPA reserve status, island, their interaction, or seawater temperature anomaly had a significant effect on these species’ biomasses over time, we used a series of generalized linear mixed models with Biomass as the independent variable, Reserve Status, Island, Oceanic Nino Index (ONI), and the interaction of Reserve Status and Island as fixed effects, and Survey Year as a random effect, using the “lme4” package in R (Bates et al. 2015). We also used bootstrapping to derive 95% confidence intervals for fish and invertebrate biomass ratios between Reserve statuses (ie inside biomass/outside biomass), and compared these ratios between targeted and non-targeted fishery species. The bootstrapping process randomly resamples the dataset with replacement, calculates a user-defined statistic (such as the biomass ratio), iterates through this process many times, and derives a 95% confidence interval using the derived distribution’s 2.5th and 97.5th percentiles (Percentile Method, Cite\_A\_Source\_Here…Rdocumentation?). Since bootstrapping is a non-parametric process, it is capable of estimating 95% confidence intervals without requiring data transformations, and provides a more accurate estimate of the data’s underlying distribution than lumped averages based on small sample sizes (n=4, one inside/outside ratio per island per year). For our bootstrapping process, we calculated biomass ratios for each bootstrap subsample (equal numbers of inside and outside MPA sites were sampled in each bootstrap iteration) for 1000 iterations using the “boot” package in R (Citation for Boot), and assessed significance by comparing whether our confidence intervals contained 1 (equal biomasses inside and outside reserves). Since dividing by or with zero results in undefined values, we removed any species in a given year where biomass was zero for all inside or outside- MPA sites.

### Equations Used

Where

* is the total number of species in the community (species richness)
* is the proportion of the total # of individuals in the community of species

Where

* is the total number at a given size
* is the total number measured
* is the biomass in grams for an individual at a given size
* is the density in #/ for a given site, species, and year

Where

* is the regression coefficient relating stipe density to foliar standing crop (after Rassweiler et al. 2018)
* is the total stipe count at a given size
* is the total stipe count
* is the density in #/ for a given site, species, and year