**Supporting Information**

**Reef coral traits predict extinction risk**

Andreas Lauchstedt, John. M. Pandolfi, Sun W. Kim and Wolfgang Kiessling

**SI Data and Methods**

Modern reef coral occurrence data were downloaded on 4th of October 2018 from OBIS (<https://iobis.org>) and comprised 343,293 occurrences of 804 species in 140 genera. OBIS data were filtered to exclude unreliable sources following Kiessling et al. (2012). Pleistocene coral data were downloaded on 4th of October, 2017 from the Paleobiology Database (PBDB) (https://paleobiodb.org) and covered 3184 late Pleistocene coral occurrences. PBDB data were filtered to keep only last interglacial (LIG) occurrences. Data labelled “last” or “latest Pleistocene”, “MIS5e” or directly dated occurrences within the time interval 120-130ka were included in the analyses. Coral trait data were downloaded on 4th of October 2018 via an R script from the Coral Trait Database (https://coraltraits.org). Coral occurrences were gridded using the R package “icosa” ([Kocsis 2017](#_ENREF_3)). All further analyses were conducted in the R programming environment applying the functions and packages detailed below.

1. Spatial aggregation of data

Coral occurrences were aggregated into equal area hexagonal grid cells with mean edge lengths of 2.5° and 4.44° which equates to mean edge lengths of 278 km and 494 km and areas of 200,790 and 634,024 km2, respectively. For each grid cell, we categorized the extinction risk of occurrences of all extant and fossil species and calculated their percentages.

1. Treatment of data, imputation tables, correlation tests and artificial neural networks

From phylogenetic imputation we derived 10 data tables comprising information from 12 traits for 739 species. The proportion of imputed data is shown in Table S1. Geographic range was also estimated from species occurrences in OBIS, providing us with 13 potential variables.

As described in the main text, we only used traits with ≤ 5% imputed data in the Artificial Neural Networks and dropped near-duplicated traits for the model.

Table S1: Traits used for the phylogenetic imputation. Percentages show the proportion of missing data in the original dataset from the CTD. Cells are color coded in green (trait used for the analyses) and red (trait not used). All used traits are preservable in fossils.

|  |  |
| --- | --- |
| Trait | Proportion of imputed data |
| Growth form by Veron | 5.2 % |
| Water depth | 62 % |
| Maximum water depth | 3.8 % |
| *Symbiodinium* clade | 59.8 % |
| Mode of larval development | 97.5 % |
| Colony maximum diameter | 60.1 % |
| Typical growth form | 1.1 % |
| Corallite width maximum | 4.5 % |
| Range size | 4.6 % |
| *Symbiodinium* in propagules | 85.9 % |
| Life history strategy | 81.2 % |
| Growth rate | 84.3 % |
| Geographic range (great-circle distance) | 0% |

Figure S1 shows how many NA entries were imputed per species. In total 95% of the species show a data content of at least 50%.

We tested for correlations between individual traits and the IUCN threat status using Spearman rank-order correlation tests between IUCN categories and predictor variables (Table S2). This procedure avoided the problem that the IUCN categories are ordinal with unknown distances between categories. We tested for multicollinearity to avoid inflation of the parameter variances by the existence of correlations among the predictor variables. The variance inflation factors (vif-function in the ‘car’ package ([Fox et al. 2012](#_ENREF_2)) were all below 1.5, indicating no multicollinearity among the four traits.

Figure S1: Proportion of missing data per species.

Table S2: Spearman rank-order correlation tests between 13 species traits, converted to binary values, and the IUCN threat status (transformed to values between “critically endangered, CR” = 5, “endangered, EN” = 4, “vulnerable, VU” = 3, “near threatened, NT” = 2 to “least concern, LC” = 1). Transformation to binary values of predictor variables was done to make correlation coefficients comparable. Values below the correlation coefficient indicate the deviation from the mean, resulting from ten different datasets derived by phylogenetic imputation. The rationale indicates why a negative correlation between trait status and extinction risk is expected. Values below trait names indicate the number of significant correlations out of 10 imputation runs. No imputation was performed on geographic range as it was calculated from OBIS occurrences. Traits used for the ANN model are highlighted in gray. Rationale follows the reasoning of the Coral Trait Database, [van Woesik et al. (2012](#_ENREF_4)) and [Carpenter et al. (2008](#_ENREF_1)).

|  |  |  |  |
| --- | --- | --- | --- |
| **Trait** | **Rationale for binary value** | **Mean correlation coefficient with IUCN status** | **Median p-value for correlation with IUCN status** |
| Geographic range | Compensates local extinction  large = 1  small = 0 | -0.295 | <<0.001 |
| Range size  10/10 | Compensates local extinction  large = 1  small = 0 | -0.272  +/- 0.009 | << 0.001 |
| Lowermost water  Depth 10/10 | > 20m = 1  < 20m = 0 | -0.264  +/- 0.009 | << 0.001 |
| Typical growth form  10/10 | massive = 1  branching = 0 | -0.172  +/- 0.0 | << 0.001 |
| Corallite width maximum 10/10 | Promotes energy storage  large = 1  small = 0 | -0.166  +/- 0.005 | << 0.001 |
| Growth rate  10/10 | Increases the demand for energy  slow = 1  fast = 0 | -0.147  +/- 0.023 | 0.001 |
| Growth form by Veron  10/10 | Influences thermal and light exposure  massive = 1  branching = 0 | -0.139  +/- 0.0053 | < 0.001 |
| Typical water depth  6/10 | Influences thermal and light exposure  > 20m = 1  < 20m = 0 | -0.124  +/- 0.028 | 0.005 |
| Mode of larval development  4/10 | Low investment and high numbers good in disturbed environment  spawner = 1  brooder = 0 | -0.115  +/- 0.024 | 0.011 |
| *Symbiodinium* clade  2/10 | Association with clade D *symbiodinium* gives thermal resistance  with clade D = 1  without clade D = 0 | -0.094  +/- 0.004 | 0.016 |
| Life history strategy  2/10 | Captures the various investments in growth, reproduction and survivorship  competitive = 1  weedy = 0 | 0.113  +/- 0.008 | 0.004 |
| Colony maximum diameter  7/10 | Energy availability  large = 1  small = 0 | 0.117  +/- 0.028 | 0.009 |
| Symbiodinium in propagules  6/10 | Promotes energy availability  yes = 1  no = 0 | 0.117  +/- 0.045 | 0.017 |



a

b

c

d

Figure S2: Trait-modeled percentage of reef coral species in respective IUCN Red List relative to total species in equal-area grid cells including maximum colony diameter. (a) Critically endangered (CR) (b) Critically endangered (CR) plus endangered (EN) species (c) Critically endangered (CR), endangered (EN) plus vulnerable (VU) (d) Critically endangered (CR), endangered (EN), vulnerable (VU) plus near threatened (NT) species.

To test the effect of adding a trait with many missing entries, we added maximum colony diameter (60% missing data) to the predictor variables. Results (Fig. S2) are graphically very similar to results using the more conservative assessment (Fig. 2) but the AUC performance is slightly better in both the binary (AUC=0.85) and the multiple case (AUC=0.93). Treated individually, coral maximum diameter yields 48.8% accuracy in predicting the IUCN status.

Figure S3 depicts the categorization of the two extinct Caribbean species *Orbicella nancyi* and *Pocillopora palmata* in one of the threatened categories (CR, EN, VU) after iteratively analyzing the ten imputation tables. Results indicate that *O. nancyi* would be placed in a non threatened category whereas *P. palmata* would be more likely to be threatened. *O. nancyi* was assigned the category LC and *P. palmata* VU in most of the cases.



Figure S3: Visualization of the threat status categorization of the two extinct Caribbean species *Orbicella nancyi* and *Pocillopora palamata.* The y-axis shows the number a species was assigned the same threat status (can be seen as percentages), either CR, EN or VU analyzing the 10 different imputation tables, the x-axis shows the number of iterative calculations performed on the 10 imputation tables. The two horizontal lines represent the median number of same results.

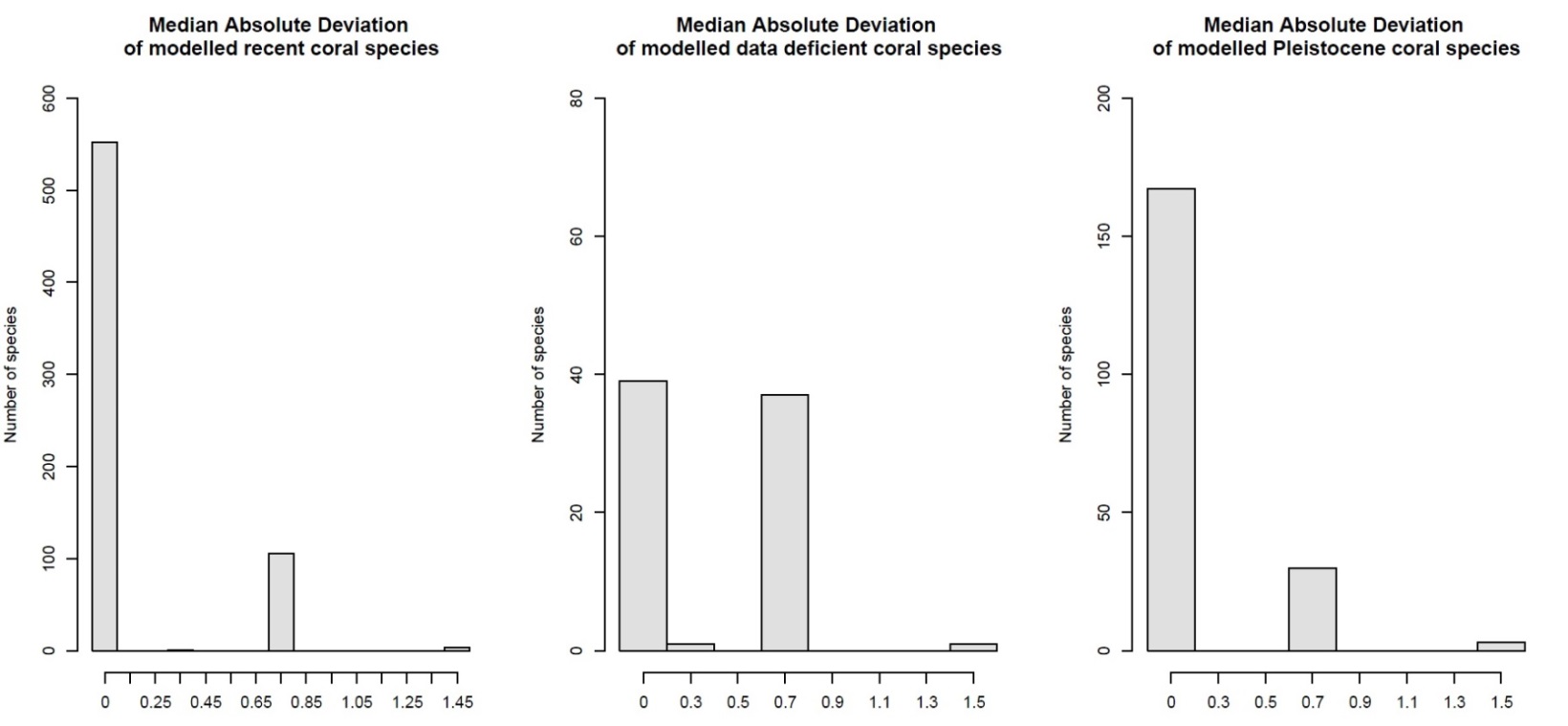


Figure S4: Median absolute deviation of the modeled threat status of recent and LIG corals.

To provide an estimate of uncertainties, we computed the median absolute deviation (MAD) among imputations (Fig. S4). MAD was adjusted by a factor ensuring asymptotically normal consistency (1.48). The bulk of coral species in the modeled recent (S4 left) shows highly consistent results after iterating the procedure over the ten imputation tables, as does the modeled Pleistocene (S4 right). Only the modeled statuses of the data deficient species (S4 middle) show a greater MAD.

References

Carpenter KE, Abrar M, Aeby G, Aronson RB, Banks S, Bruckner A, Chiriboga A, Cortés J, Delbeek JC, DeVantier L (2008) One-third of reef-building corals face elevated extinction risk from climate change and local impacts. Science 321:560-563

Fox J, Weisberg S, Adler D, Bates D, Baud-Bovy G, Ellison S, Firth D, Friendly M, Gorjanc G, Graves S (2012) Package ‘car’. Vienna: R Foundation for Statistical Computing

Kocsis AT (2017) The R package'icosa'for coarse resolution global triangular and penta-hexagonal gridding. EGU General Assembly Conference Abstracts 19:17057

van Woesik R, Franklin EC, O'Leary J, McClanahan TR, Klaus JS, Budd AF (2012) Hosts of the Plio-Pleistocene past reflect modern-day coral vulnerability. Proceedings of the Royal Society of London B: Biological Sciences:rspb20112621