SUPPLEMENTAL FIGURES AND TABLES
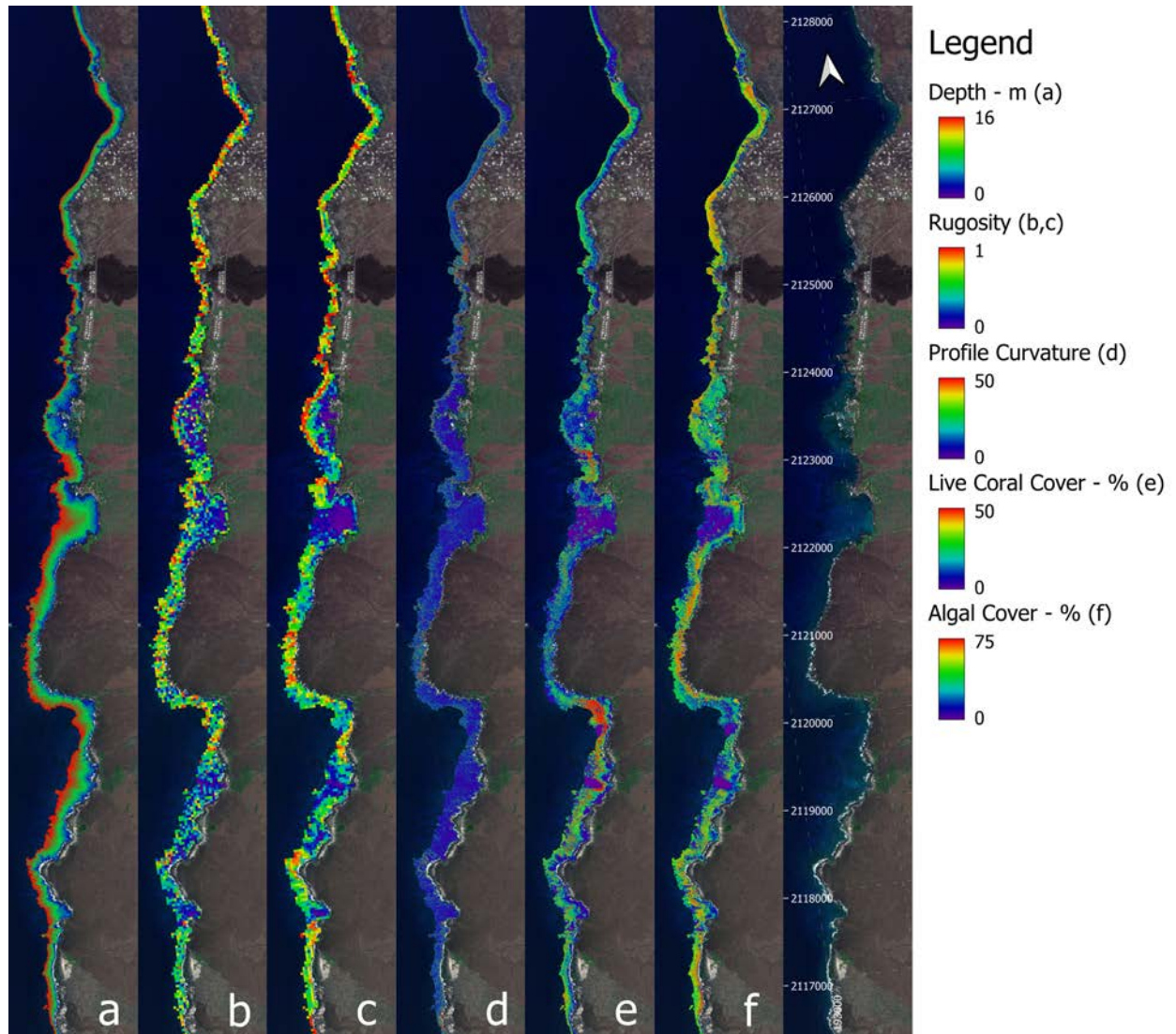

Figure S1. Example section of input variable maps derived from GAO data: a) Depth, b) Fine-scale rugosity, c) Coarse-scale rugosity, d) Profile curvature, e) Live coral cover, and f) Algal cover
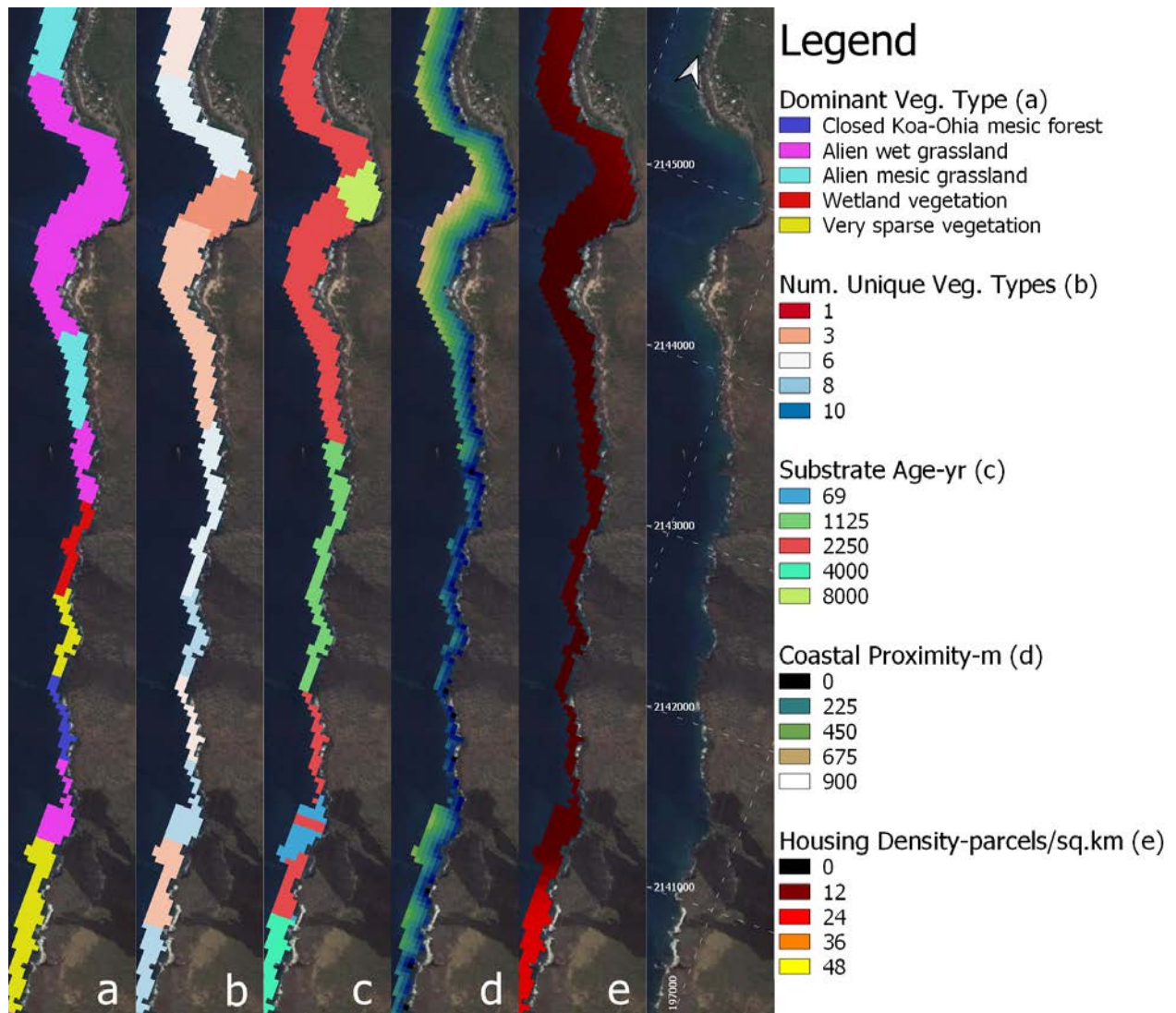
Figure S2. Example section of input variable maps derived from GIS data: a) Dominant vegetation type - 3 km, b) Number of vegetation types - 3 km, c) Substrate age, d) Coastal proximity, and e) Housing density - 3 km.
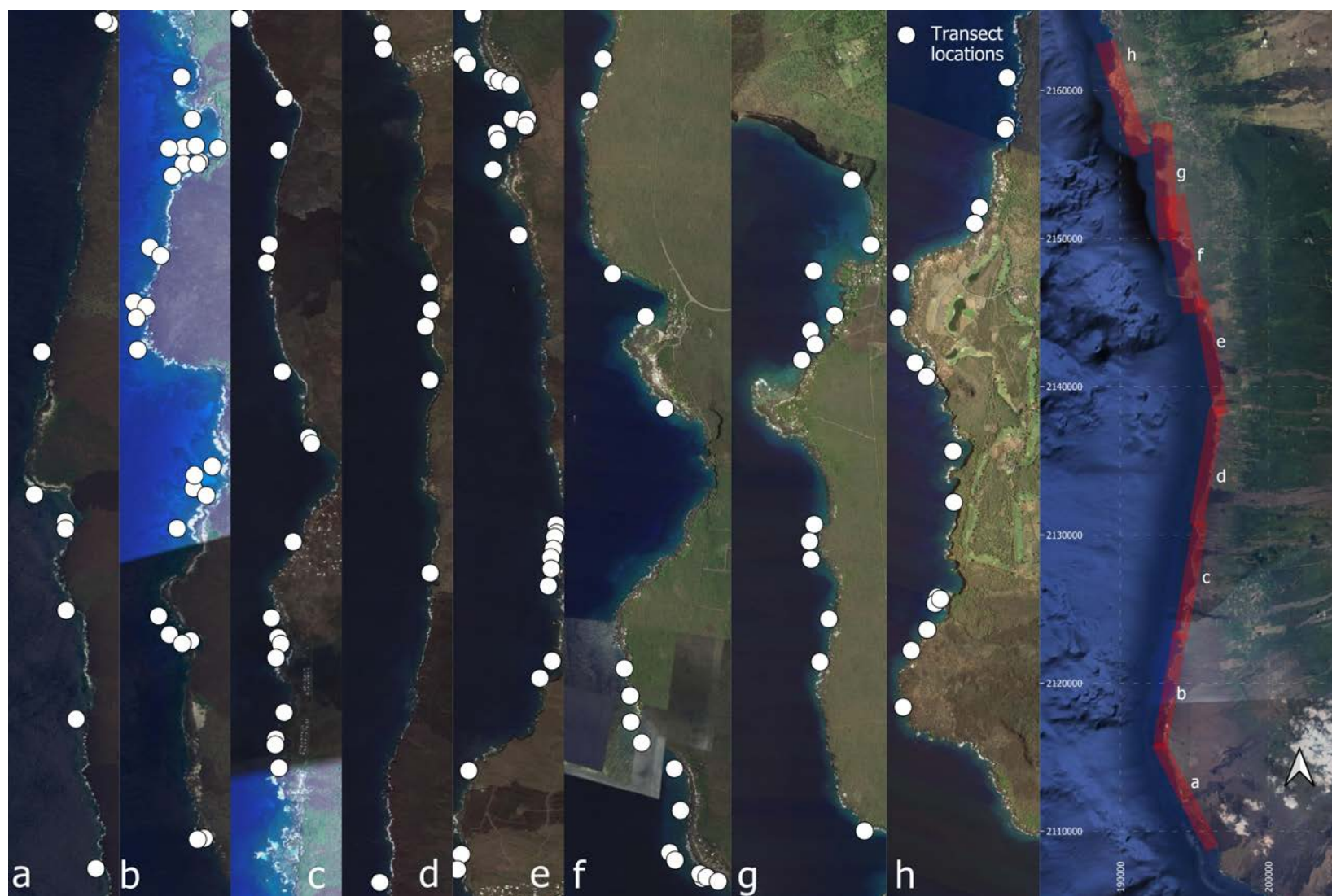
Figure S3. Location of 117 randomly selected field sites based on the 18-strata classification of the reefs of South Kona.
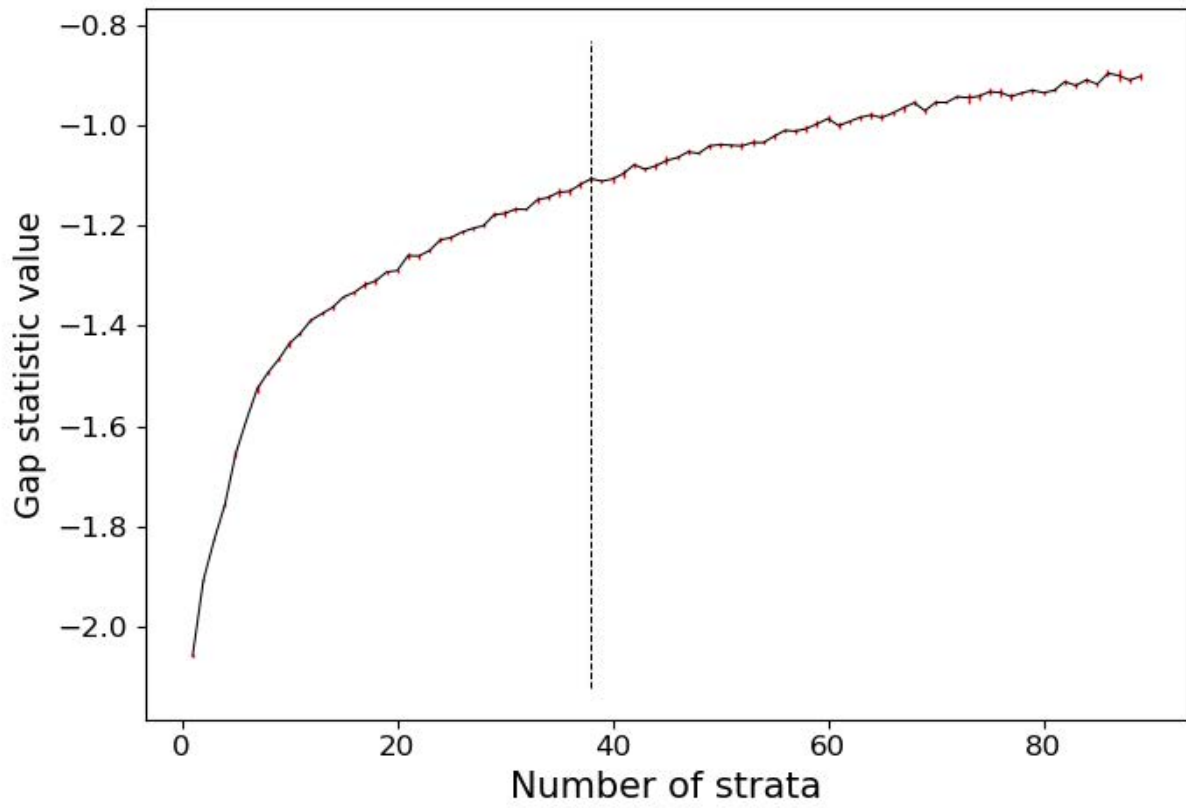
Figure S4. Gap statistic against number of strata (k from the k-Means clustering procedure) for the full input dataset from Step-1. Here, the standard methodology of taking the first k such that $Gap_k > Gap_{k+1} + StdDev(Gap_{k+1})$ gives an initial k of 38.
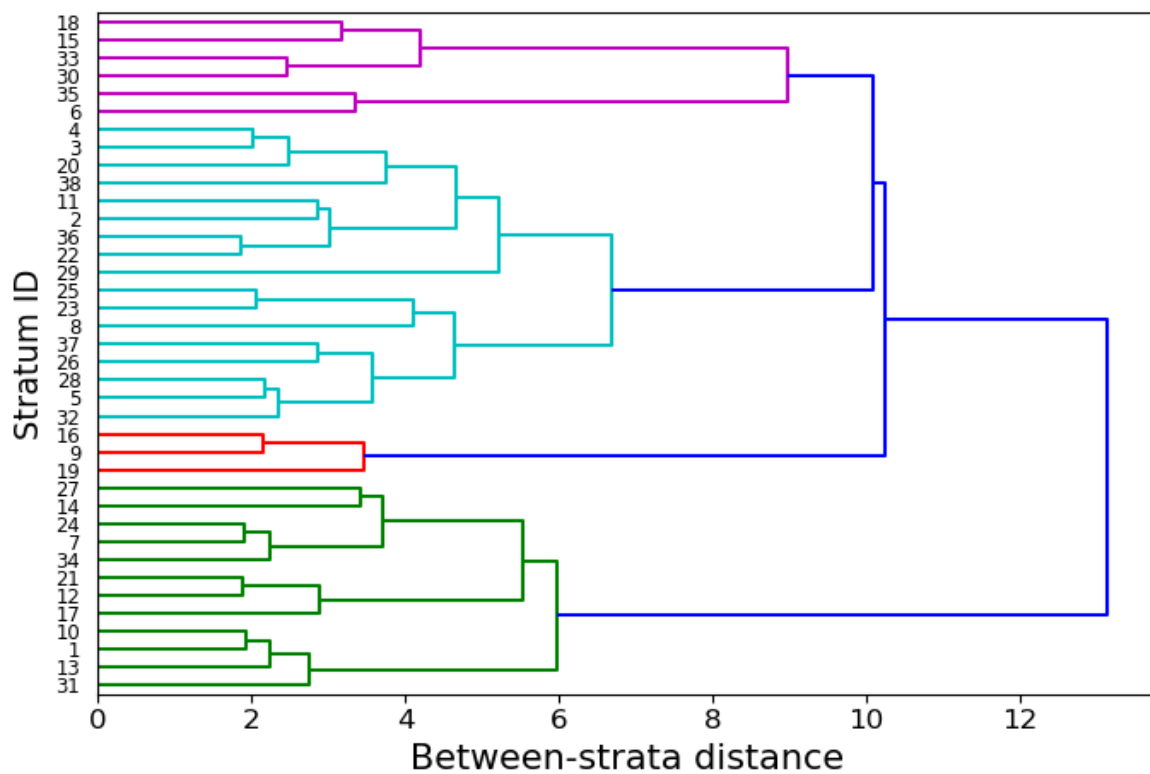
Figure S5. Dendrogram of the between-strata Euclidean distance within the high-dimensional full input variable suite from the first iteration of Step 1.
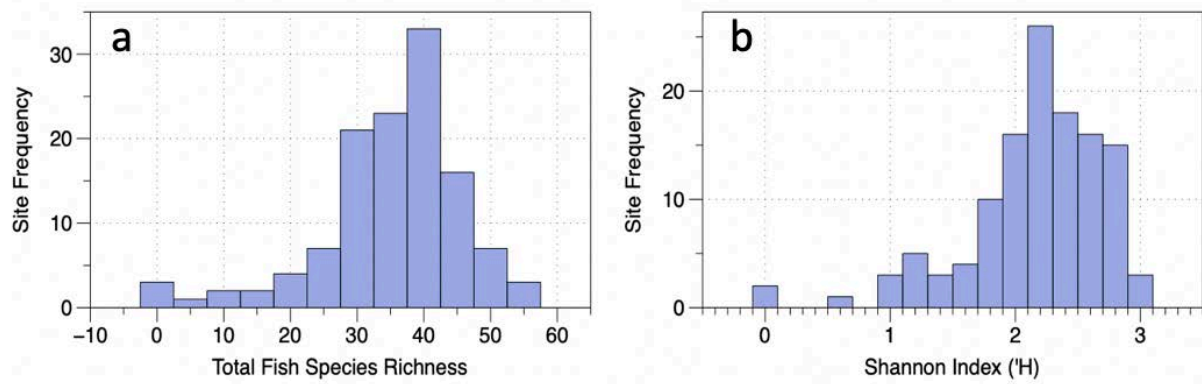
Figure S6. Histograms of the distribution of two measures of biodiversity: (a) species richness (number of species) and (b) Shannon Index for 117 field sites.
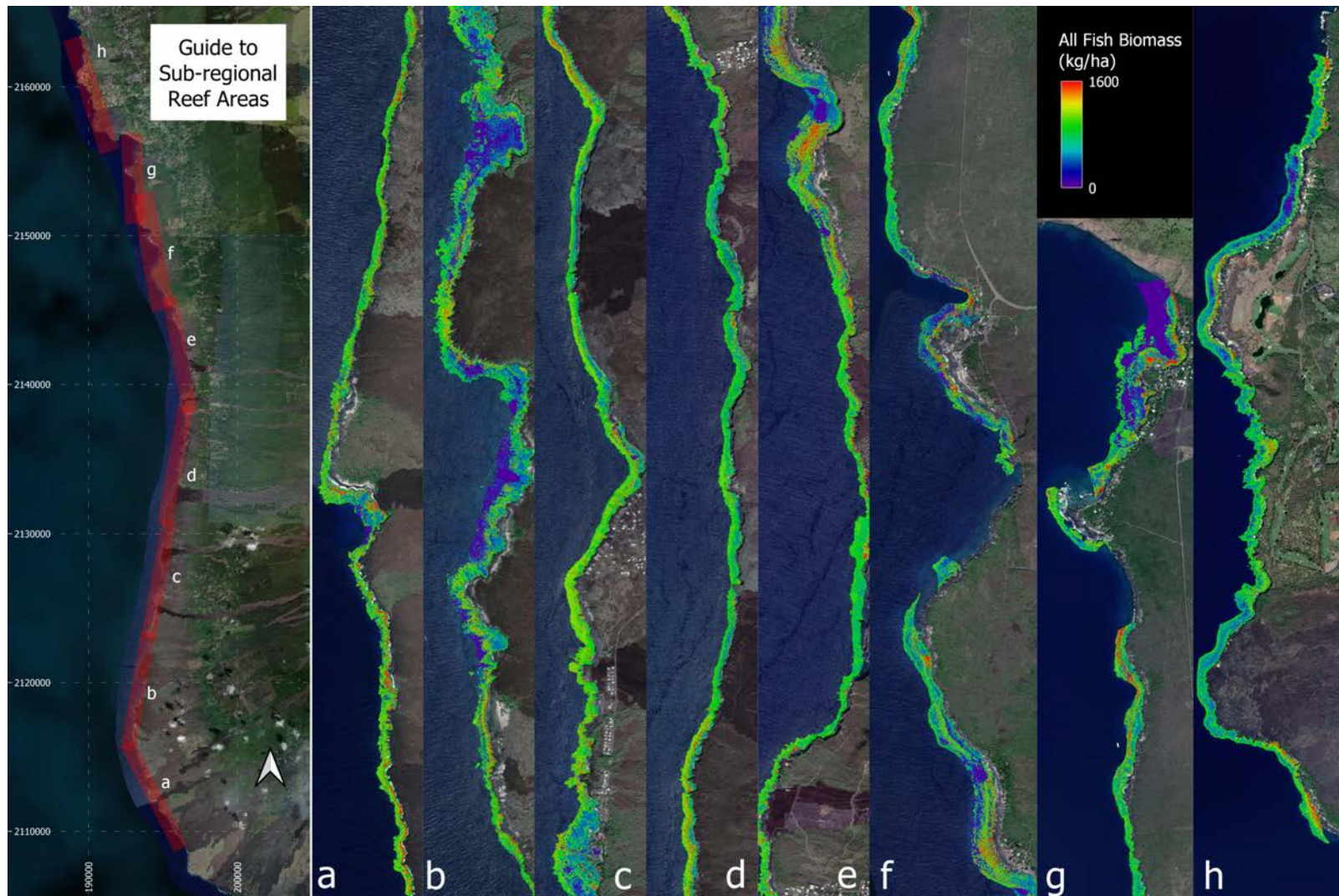
Figure S7. Map of **all fish biomass** for the South Kona reef ecosystem.

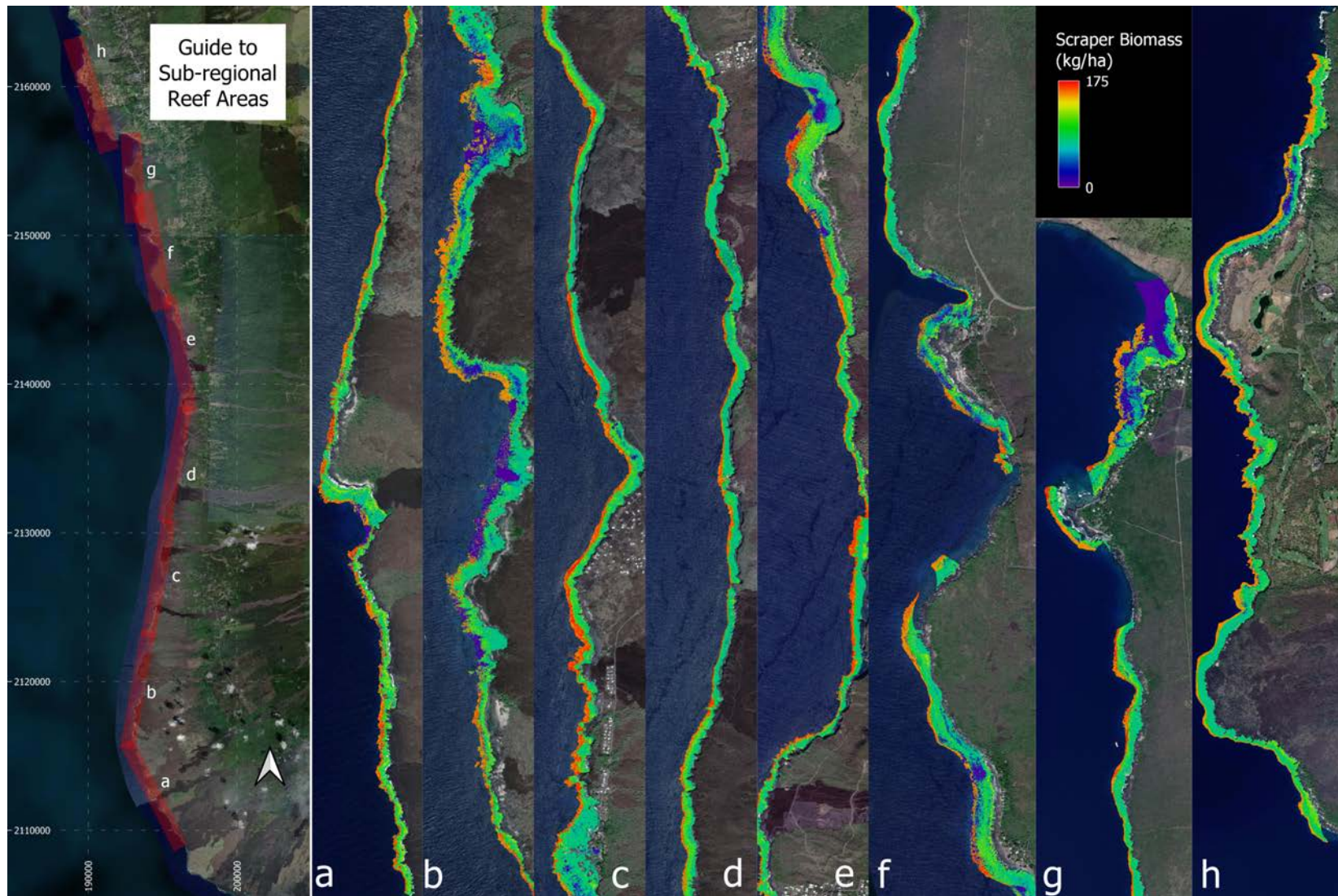**Figure S8**. Map of **scraper biomass** for the South Kona reef ecosystem.
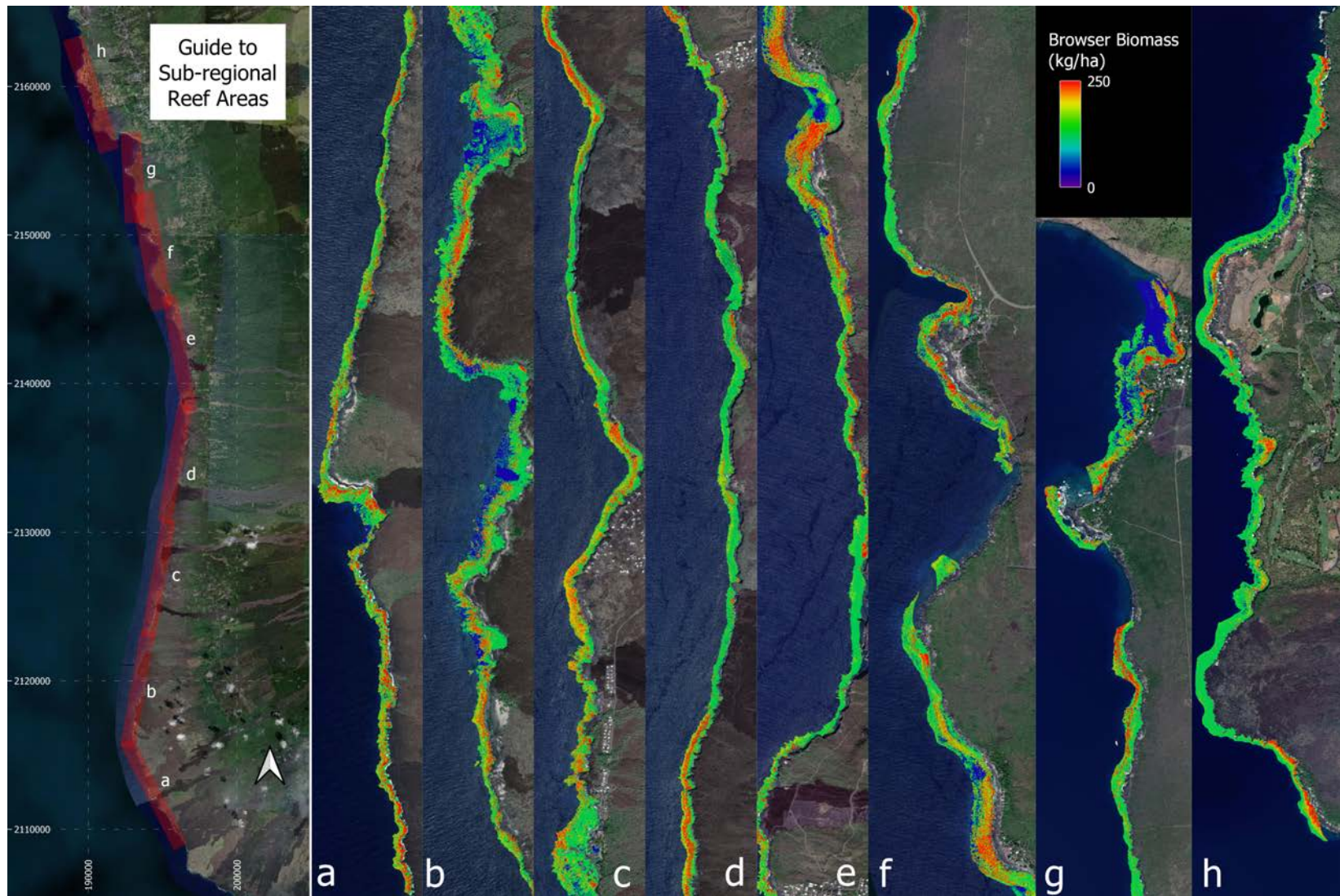
Figure S9. Map of **browser biomass** for the South Kona reef ecosystem.
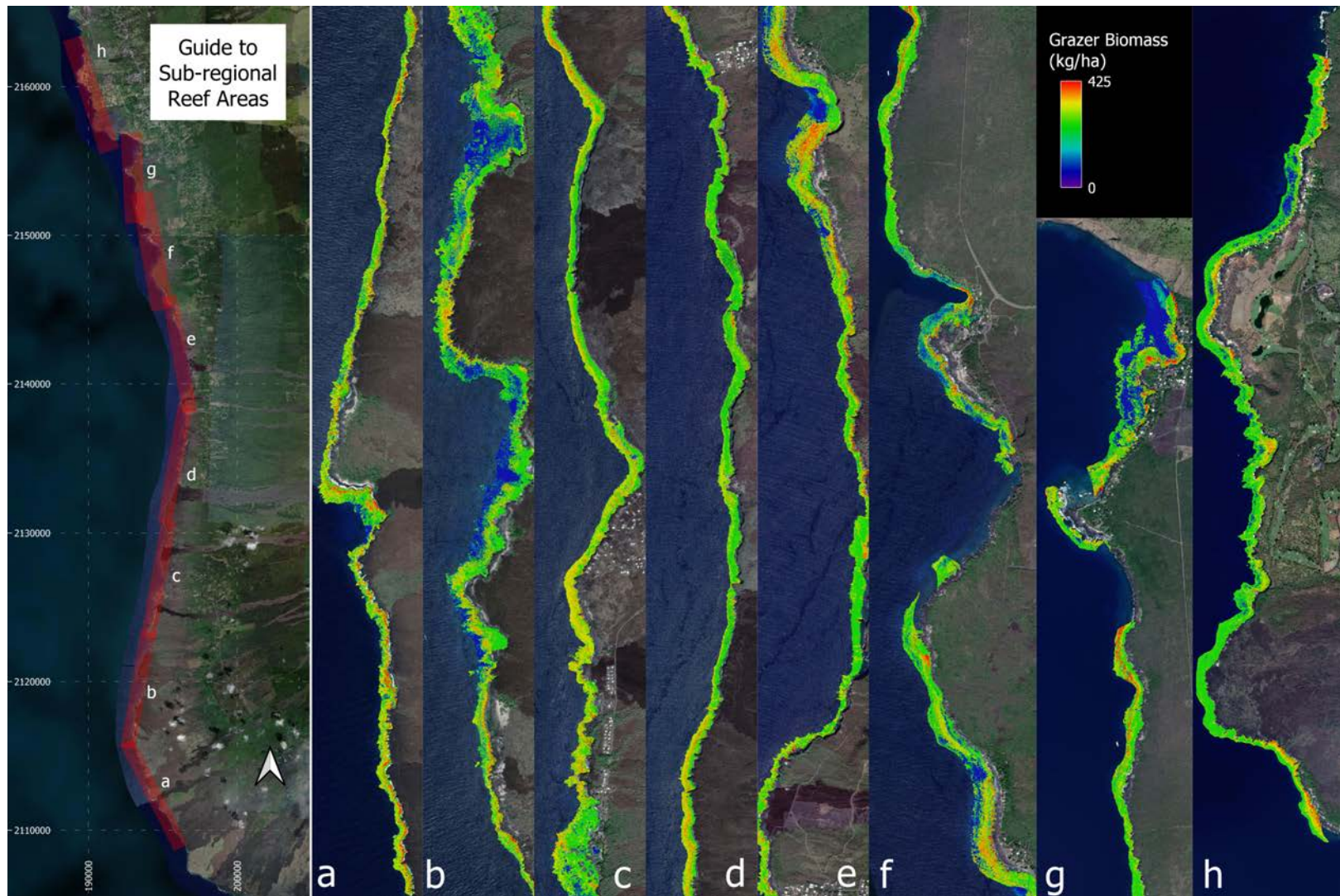
Figure S10. Map of **grazer biomass** for the South Kona reef ecosystem.
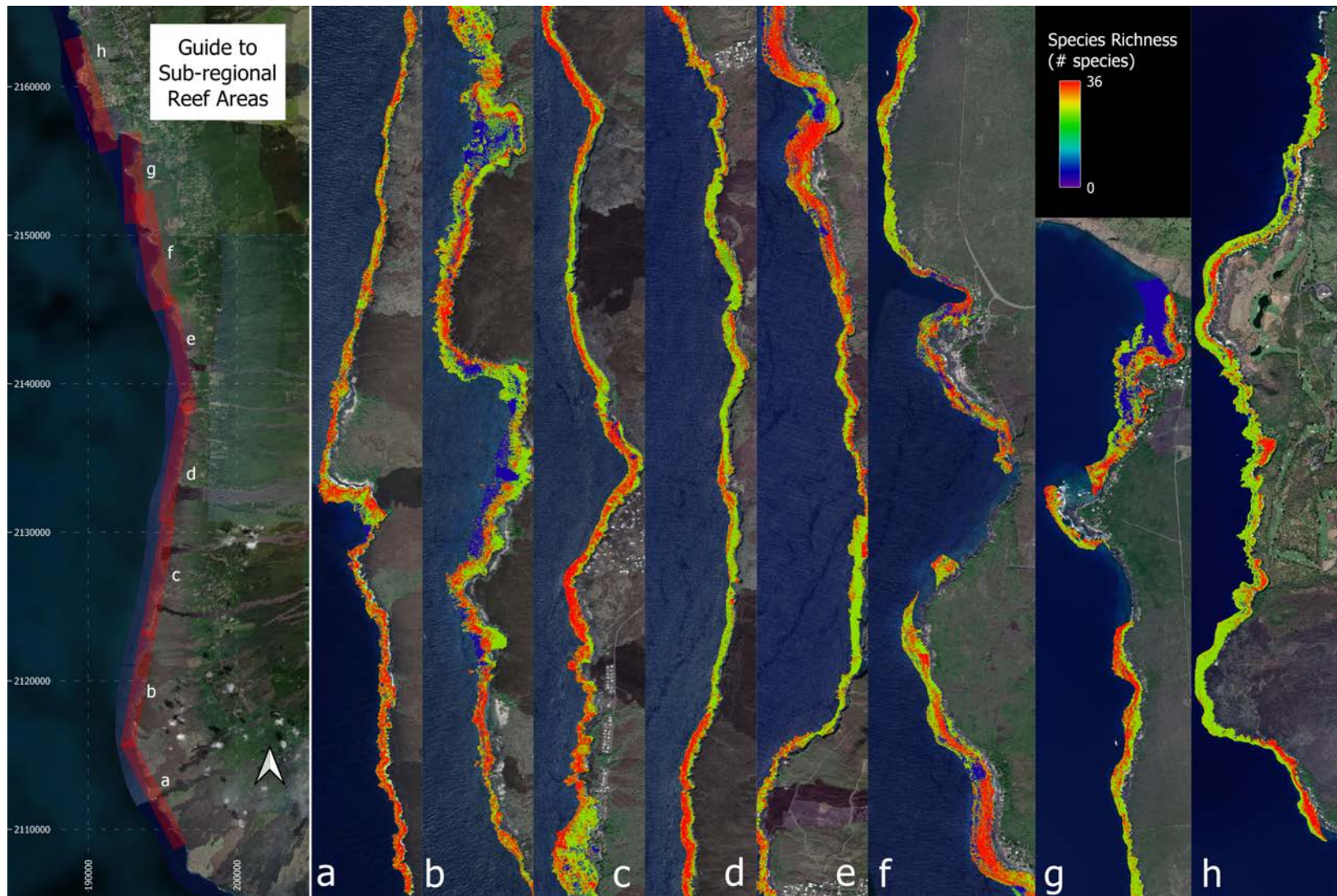
Figure S11. Map of **all fish species richness** for the South Kona reef ecosystem.
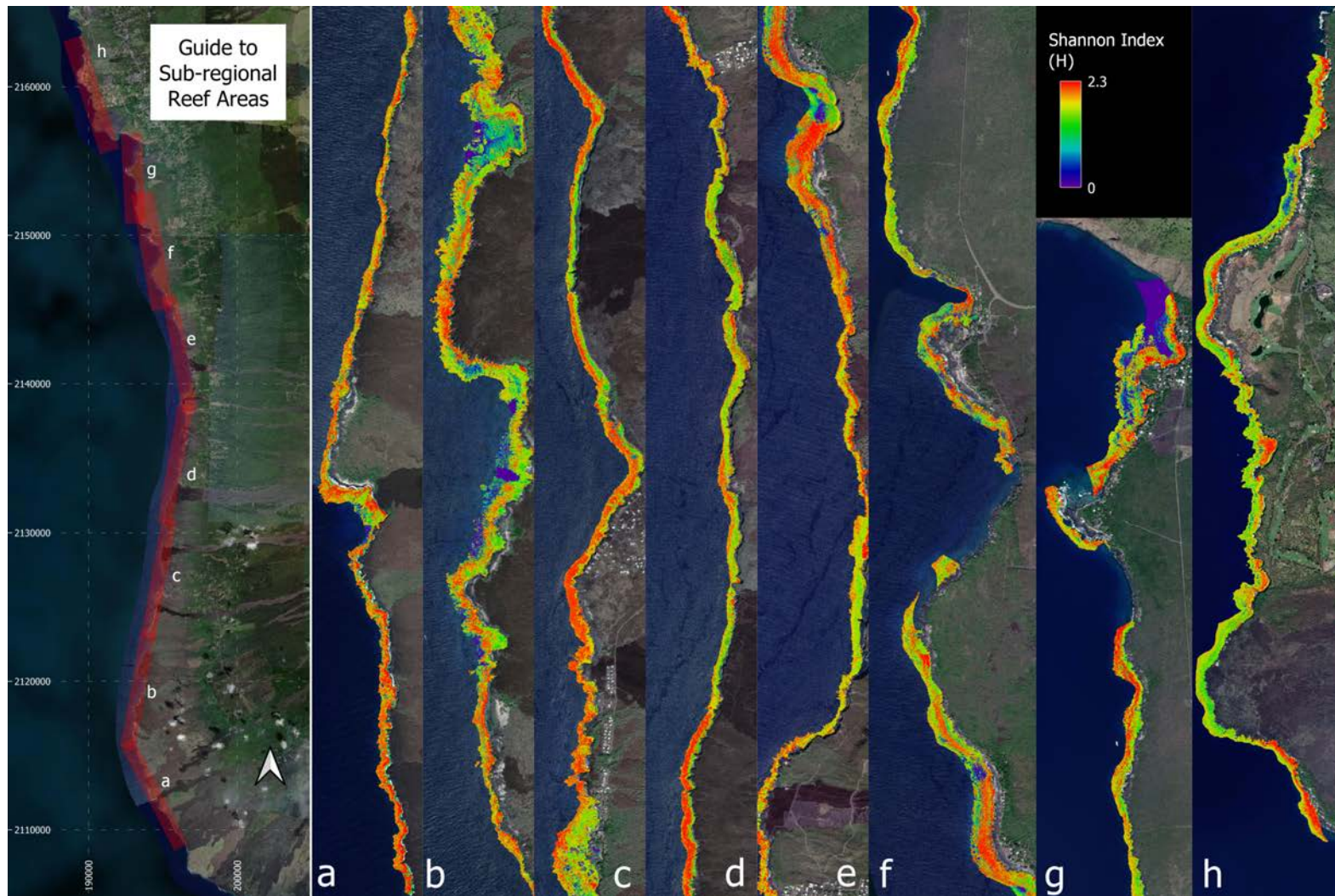
Figure S12. Final map of **Shannon diversity index** for the South Kona reef ecosystem.
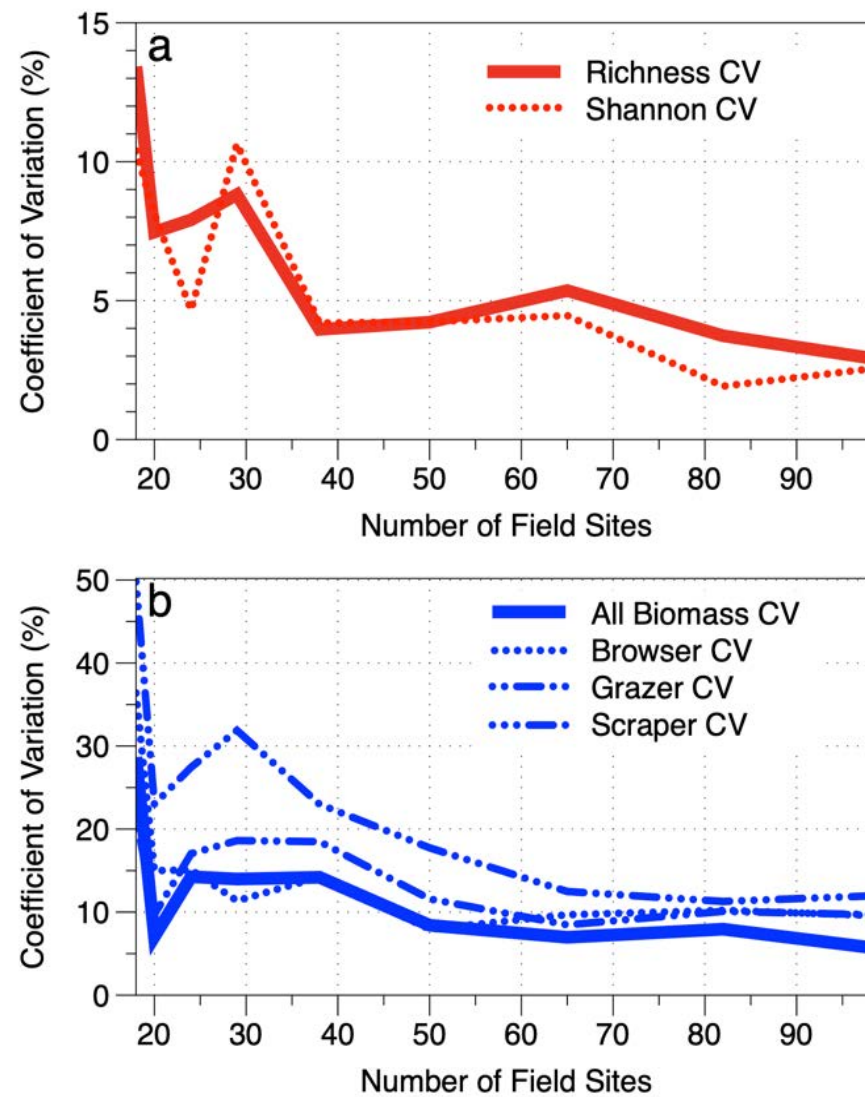
Figure S13. Change in regional mapped-based estimates of fish biomass and diversity as indicated in coefficient of variation (CV).
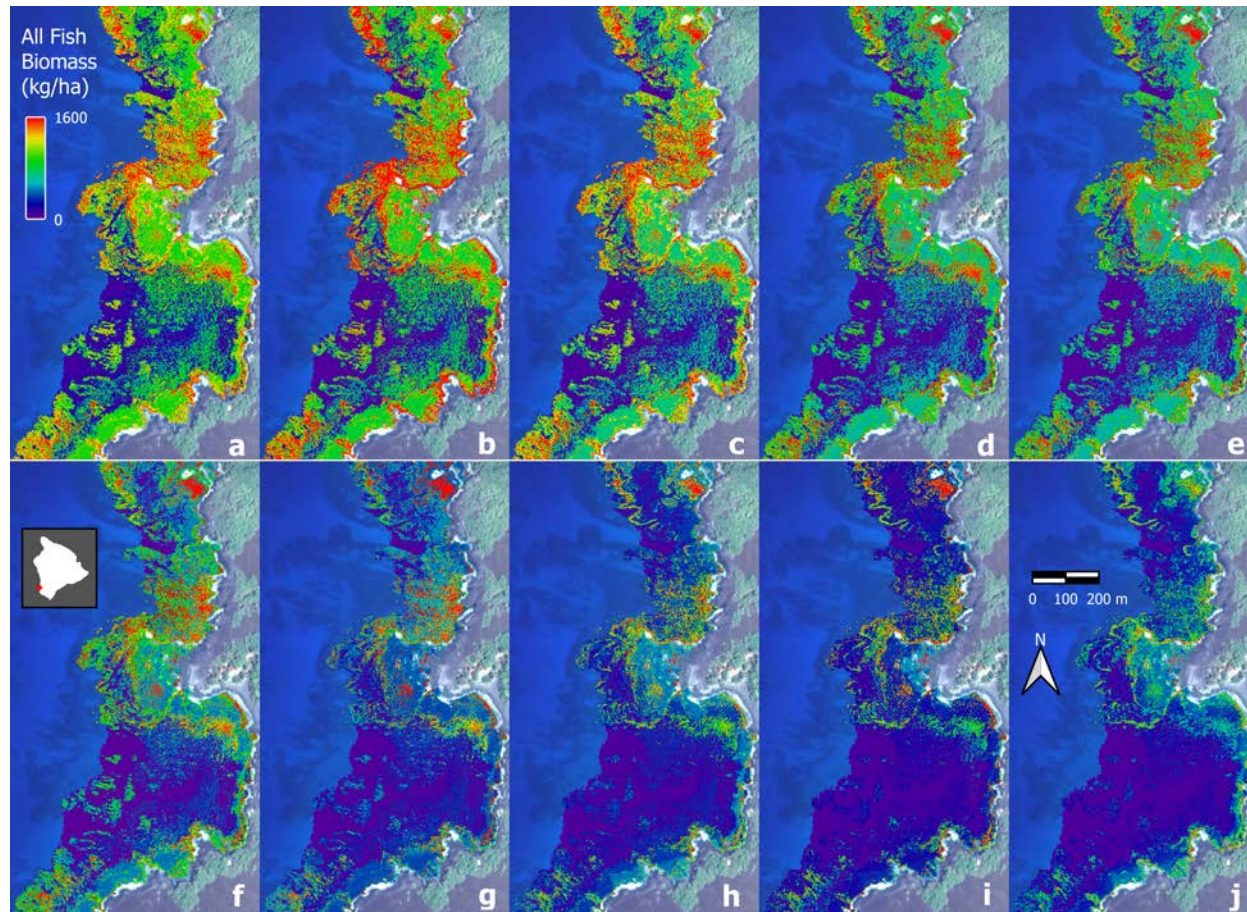
Figure S14. Loss in map detail resulting from reducing the number of field sites used in the upscaling process for all fish biomass, with the same example Honomalino Bay zoom. The maps shown are the result of using (a) all 117, (b) 99, (c) 82, (d) 65, (e) 50, (f) 38, (g) 29, (h) 24, (i) 20 and (j) 18 field sites during the upscaling process.

Figure S15. Number of region-wide field sites required to maintain statistical stability of mapped (pixel-level) estimates of all-fish biomass. See Figure 11 in the main text for a zoom image example of this map.

Table S1. Description of the input data layers considered in the clustering and stratification process.

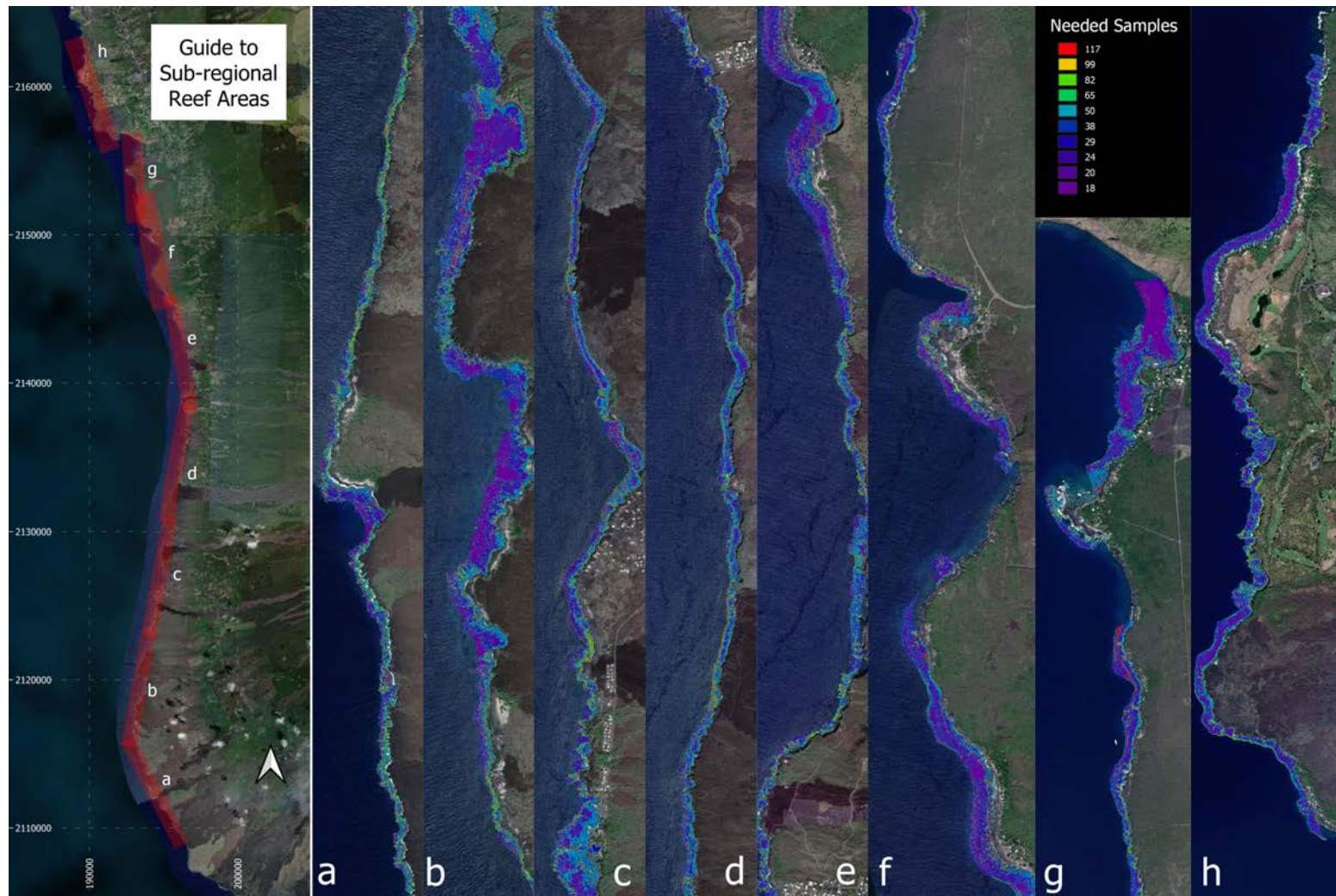| Name	Description	Source
Depth	GAO spectrometer data (2 m resolution)	GAO (Asner et al. 2020a)
Fine-scale rugosity	Planar rugosity from GAO depth (9 m moving window on 2 m grid)	GAO (Asner et al. 2021)
Coarse-scale rugosity	Planar rugosity on resampled GAO depth maps (54 m moving window on 6m grid)	GAO (Asner et al. 2021)
Profile curvature	Profile curvature on GAO depth maps (6m moving window on 2 m grid)	GAO (Asner et al. 2020a)
Live coral cover	GAO spectrometer data (2 m)	GAO (Asner et al. 2020b)
Algal cover	GAO spectrometer data (2 m)	GAO (Asner et al. 2020b)
Distance from shore	Distance from nearest point of coast (30 m)	HI State Coastal Outline (http://geoportal.hawaii.gov)
Northing	UTM northing coordinate (30 m)	None
Land-based substrate age	Midpoint of age range from the *Geologic Map of Hawaii*.	USGS (Sherrod et al. 2007)
Housing density - 1km	Number of tax parcels within 1 km radius of nearest point on shore.	HI State Tax Map (http://geoportal.hawaii.gov)
Housing density - 3km	Number of tax parcels within 3 km radius of nearest point on shore.	HI State Tax Map (http://geoportal.hawaii.gov)
Dominant vegetation type - 1km	Dominant vegetation type within 1 km radius of nearest point on shore	HIGAP (Gon et al. 2006)
Number of veg. types - 1km	Number of vegetation types within 1 km radius of nearest	HIGAP (Gon et al. 2006)

	point on shore	
Dominant vegetation type - 3km	Dominant vegetation type within 3 km radius of nearest point on shore	HIGAP (Gon et al. 2006)
Number of veg. types - 3km	Number of vegetation types within 3 km radius of nearest point on shore	HIGAP (Gon et al. 2006)

Table S2. Optimal Random Forest Machine Learning (RFML) parameters and fit statistics for each of the biomass and diversity response variables upscaled to output maps. Parameters allowed to vary in the grid search algorithm were Trees: the total number of trees used in the RFML model, Depth: the maximum branching depth allowed for trees in the model, and Samples: the minimum number of samples needed add an end node (leaf) for a split to be allowed.

Response variable	Trees	Depth	Samples	RMSE	R^2
Total Biomass (kg ha^{-1})	50	3	10	740.3	0.37
Browser Biomass (kg ha^{-1})	50	3	10	173.3	0.32
Grazer Biomass (kg ha^{-1})	100	3	2	298.6	0.32
Scraper Biomass (kg ha^{-1})	75	3	2	213.2	0.38
Diversity - Richness (count)	250	3	5	11.4	0.54
Diversity - Shannon (unitless)	100	3	5	0.6	0.58