

# CHAPTER 3

## MACROECOLOGY MEETS MACROEVOLUTION: EVOLUTIONARY NICHE DYNAMICS IN THE SEAWEED *HALIMEDA*

Heroen Verbruggen<sup>1\*</sup>, Lennert Tyberghein<sup>1\*</sup>, Klaas Pauly<sup>1\*</sup>, Caroline Vlaeminck<sup>1</sup>,  
Katrien Van Nieuwenhuyze<sup>1</sup>, Wiebe H.C.F. Kooistra<sup>2</sup>, Frederik Leliaert<sup>1</sup>, Olivier De  
Clerck<sup>1</sup>

<sup>1</sup> Phycology Research Group and Center for Molecular Phylogenetics and Evolution, Ghent University, Krijgslaan 281, building S8 (WE11), 9000 Ghent, Belgium

<sup>2</sup> Stazione Zoologica 'Anton Dohrn', Villa Comunale, 80121 Naples, Italy

\* these authors contributed equally to this work

Adapted from: Verbruggen H, Tyberghein L, Pauly K, Vlaeminck C, Van Nieuwenhuyze K, Kooistra WHCF, Leliaert F & De Clerck O (2009) Macroecology meets macroevolution: evolutionary niche dynamics in the seaweed *Halimeda*. *Global Ecology and Biogeography* 18, 393-405.

# ABSTRACT

Because of their broad distribution in geographic and ecological dimensions, seaweeds (marine macroalgae) offer great potential as models for marine biogeographic inquiry and exploration of the interface between macroecology and macroevolution. This study aims to characterize evolutionary niche dynamics in the common green seaweed genus *Halimeda*, use the observed insights to gain understanding of the biogeographic history of the genus, and predict habitats that can be targeted for discovery of species of special biogeographic interest. The evolutionary history of the genus is characterized using molecular phylogenetics and relaxed molecular clock analysis. Niche modeling is carried out with Maximum Entropy techniques and uses macroecological data derived from global satellite imagery. Evolutionary niche dynamics are inferred through application of ancestral character state estimation. A nearly comprehensive molecular phylogeny of the genus was inferred from a six-locus dataset. Macroecological niche models showed that species' distribution ranges are considerably smaller than their potential ranges. We show strong phylogenetic signal in various macroecological niche features. The evolution of *Halimeda* is characterized by conservatism for tropical, nutrient-depleted habitats, yet one section of the genus managed to invade colder habitats multiple times independently. Niche models indicate that the restricted geographic ranges of *Halimeda* species are not due to habitat unsuitability, strengthening the case for dispersal limitation. Niche models identified hotspots of habitat suitability of Caribbean species in the East Pacific Ocean. We propose that these hotspots be targeted for discovery of new species separated from their Caribbean siblings since the Pliocene rise of the Central American Isthmus.



# INTRODUCTION

Various interacting features influence the distribution of a species. The niche of a species is commonly defined as the set of biotic and abiotic conditions in which it is able to persist and maintain stable population sizes (Hutchinson, 1957). Further distinction is made between a species' fundamental niche, which consists of the set of all conditions that allow for its long-term survival and the realized niche, which is a subset of the fundamental niche that a species actually occupies. Species' tolerances are determined by their morphological, reproductive and physiological traits, which are in turn susceptible to evolutionary forces. Hence, niche characteristics can be interpreted as evolutionary phenomena. Understanding niche evolution yields valuable insights into biogeography, biodiversity patterns and conservation biology (Wiens & Graham, 2005; Rissler *et al.*, 2006; Wiens *et al.*, 2007).

The niche concept provides a conceptual framework to predict geographical distributions of species. Niche models establish the macroecological preferences of a given species based on observed distribution records and a set of macroecological variables, and these preferences can subsequently be used to predict geographic areas with suitable habitat for the species (e.g., Guisan & Thuiller, 2005; Raxworthy *et al.*, 2007; Rissler & Apodaca, 2007). The availability of macroecological data, either in the form of remote sensing or interpolated measurement data, is increasing and has already provided many biological studies with environmental information (Kozak *et al.*, 2008). To date, most ecological niche modeling studies have focused on terrestrial organisms. A notable exception is the study by Graham *et al.* (2007), which used a synthetic oceanographic and ecophysiological model to identify known kelp populations and predict the existence of undiscovered kelp habitats in deep tropical waters.

Integration of niche models, macroecological data and phylogenetic information yields information on niche shifts and insights in the evolution of environmental preferences across phylogenetic trees. So far, evolutionary niche dynamics have been studied almost exclusively in terrestrial organisms (e.g., Graham *et al.*, 2004; Knouft *et al.*, 2006; Yesson & Culham, 2006) and little information is available on niche evolution of the organisms inhabiting the world's oceans. Seaweeds appear to be an excellent model system to study the evolutionary dynamics of the macroecological niche in coastal marine organisms. Individual seaweed specimens are fixed in one location, yielding a direct link to georeferenced macroecological data. As a whole, seaweeds occur in a wide range of coastal habitats and many seaweed genera or families have a worldwide distribution, resulting in sufficient variability in macroecological dimensions and biogeographic patterns. Evolutionary relationships between and within seaweed genera are being characterized in increasing detail as a result of molecular phylogenetic research during the past few decades. Finally, seaweeds are straightforward to collect and process, making them easy targets for this kind of research.

The marine green algal genus *Halimeda* is among the better studied seaweeds from a phylogenetic perspective and is therefore an obvious candidate for studies of niche evolution and biogeography. *Halimeda* consists of segmented, calcified thalli and abounds on and around coral reefs and in lagoons throughout the tropics and subtropics up to depths in excess of 150 m (Hillis-Colinvaux, 1980). *Halimeda* species are important primary producers and provide food and habitat for small animals and epiflora (Jensen *et al.*, 1985; Naim, 1988). After the algae reproduce, they die and their calcified segments are shed. *Halimeda* segments account for up to 90% of tropical beach sand and carbonate rock of tropical reefs (e.g., Drew, 1983; Freile *et al.*, 1995). The biogeography of *Halimeda* has been described in some detail. All but one species are restricted to a single ocean basin (Indo-Pacific or Atlantic), and biogeography has a strong phylogenetic imprint: each of the five

sections of the genus is separated into Atlantic and Indo-Pacific sublineages, suggestive of a strong vicariance event. Even though the species' distribution ranges and the historical biogeographic patterns have been identified, questions about what causes them remain (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005b). Are species restricted to one ocean basin because of habitat unsuitability in the other basin or should the limited distribution ranges be attributed to dispersal limitation? It is also not known with certainty which vicariance event may be responsible for the phylogenetic separation of Indo-Pacific and Atlantic lineages. So far, two geological events have been implied: the Miocene closure of the Tethys seaway in the Middle East and the Pliocene shoaling of the Central American Isthmus (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005b).

The first goal of the present study is to investigate evolutionary niche dynamics of the seaweed genus *Halimeda*, focusing on niche dimensions relevant to global geographic distributions rather than local distributional issues such as microhabitat preferences. The second goal is to investigate two aspects of the biogeography of the genus: why are species restricted to a single ocean basin and what caused the historical biogeographic splits. Our approach consists of a combination of molecular phylogenetics, niche modeling, optimization of models of macroecological trait evolution, and ancestral state estimation.



# MATERIAL AND METHODS

## SPECIES IDENTIFICATIONS

Species delimitation was based on a combination of DNA sequence data and morphological knowledge, with molecular data serving as the primary source of information used to define species boundaries and morphological species boundaries being assessed secondarily, using the species groups determined with DNA data. We used this approach because traditional morphological species definitions are often inaccurate in seaweeds due to morphological plasticity, convergence and cryptic speciation (e.g., Saunders & Lehmkuhl, 2005). The proposed approach has previously been applied to define species boundaries more accurately (Verbruggen *et al.*, 2005a).

The DNA datasets initiated by Kooistra *et al.* (2002) and Verbruggen (2005) were extended for this study using previously described protocols (Verbruggen, 2005), resulting in 264 UCP7 sequences, 337 ITS sequences and 106 *tufA* sequences belonging to a total of 444 specimens. These three datasets were subjected to Neighbor Joining analysis to detect species-level clusters. Using this approach, the sequenced specimens were attributed to 52 *Halimeda* species. If easily recognizable combinations of morphological features could be identified for species by studying the sequenced specimens, these features were used for identification of additional collections from various herbaria (BISH, GENT, L, PC, UPF, US) that were not suitable for sequencing. See the Index Herbariorum web site (<http://sweetgum.nybg.org/ih/>) for definitions of herbarium acronyms.

## PREPROCESSING OBSERVATION DATA

Recent collections had accurate coordinates recorded with a global positioning device. Older collections with detailed locality information were geocoded (latitude and longitude) using Google Earth (<http://earth.google.com>). Points that fell ashore when plotted on coarse resolution environmental grids were manually moved to the adjacent coastal waters using IDRISI Andes (<http://www.clarklabs.org/>). Data were examined for georeferencing errors by checking for geographical outliers with visual and overlay methods in ArcGIS (<http://www.esri.com>). Errors were identified by creating an overlay between the point locality layer and a maritime boundaries layer (exclusive economic zones and coastlines) provided by the Flanders Marine Institute (<http://www.vliz.be>). Any mismatch between these layers were indicative for a potential georeferencing error and outlying points were removed if their origin could not be confirmed.

## SPECIES PHYLOGENY

The evolutionary history underlying the 52 species of *Halimeda* included in the study was inferred from a multi-locus DNA dataset using Bayesian phylogenetic inference (Holder & Lewis, 2003). Bayesian phylogenetic inference techniques make explicit use of models of sequence evolution, an approach that has been shown to outperform methods that do not assume such models (Swofford *et al.*, 2001). Sequence data from four chloroplast loci (*rbcl*, *tufA*, UCP3, UCP7) and two nuclear markers (SSU nrDNA, ITS region) were obtained following previously described protocols (Famà *et al.*, 2002; Kooistra *et al.*, 2002; Provan *et al.*, 2004; Lam & Zechman, 2006) or from previously published studies (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005a; Verbruggen *et al.*, 2005b).



Individual loci were aligned by eye and ambiguous regions were removed. Data for a few loci were missing mainly for recently discovered species but the concatenated data matrix was 90% filled. The final alignment can be obtained from [www.phycoweb.net](http://www.phycoweb.net) and [www.treebase.org](http://www.treebase.org). All new sequences generated in this study have been submitted to Genbank (accession numbers FJ624485-FJ624863).

In order to identify a suitable model of sequence evolution for our dataset, we used model selection procedures based on the second order Akaike Information Criterion (AICc) (Sullivan, 2005). The phylogenetic analysis was carried out with the model of sequence evolution that yielded the lowest AICc score. This model contained 14 partitions: SSU nrDNA, the ITS region and 3 codon positions per protein-coding gene. The GTR+ $\Gamma_8$  substitution models yielded the best fit to the data for all partitions. Bayesian phylogenetic inference was carried out with MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003). Five runs of four incrementally heated chains were run for 10 million generations using default priors and chain temperature settings. Convergence of the MCMC runs was assessed with Tracer v.1.4 (Rambaut & Drummond, 2007). An appropriate burn-in was determined with the automated method proposed by Beiko *et al.* (2006) and a majority rule consensus tree was built from the post-burn-in trees. The tree was rooted at the point where root-to-tip path length variance was minimal.

The branch lengths of the obtained consensus phylogram are proportional to the estimated amount of molecular evolution occurring on the branches. In order to model character evolution, in our case evolutionary niche dynamics, branch lengths should be proportional to evolutionary time rather than amounts of molecular evolution. To obtain a chronogram (i.e. a phylogram with branch lengths proportional to evolutionary time), penalized likelihood rate smoothing (Sanderson, 2002) was carried out on the consensus tree with r8s (Sanderson, 2003), using both the additive and the log-additive penalty settings. The root of the phylogeny was assigned an age of 147 my, following the molecular clock result from Verbruggen *et al.* (2009). We refer to the latter paper for details regarding the dating of the phylogeny.

## MACROECOLOGICAL DATA

Macroecological variables were selected to represent the major environmental dimensions assumed to influence seaweed distribution at a global scale and subject to data availability (Lüning, 1990) (Table 1). The base macroecological data included geophysical, biotic and climate variables derived from level-3 reprocessed satellite data (Aqua-MODIS and SeaWiFS) available at OceanColor Web (<http://oceancolor.gsfc.nasa.gov>).

We downloaded grids representing monthly averages at a 5 arcmin ( $\approx 9.2$  km) spatial resolution. These geometrically corrected images are two-dimensional arrays with an Equidistant Cylindrical (Platte Carre) projection of the globe. Yearly minimum, maximum and average values were calculated from the monthly averages with MATLAB (<http://www.mathworks.com/>). To achieve this, average monthly images were generated by averaging images of the same month across years (e.g. average SST of July from 2003–2007). Subsequently, yearly minimum and maximum images were composed by selecting the minimum and maximum pixels from these monthly averages. Finally, yearly average images were created by taking the mean value for every grid cell of the monthly averages. All images were cropped to the latitudinal range 50°N–40°S, which includes the highest latitudes at which *Halimeda* can be found.

**Table 1. Geophysical parameters included in the macroecological dataset.**

Macroecological parameter	Units	Original spatial resolution	Date	Source	Derived parameters
Sea Surface Temperature (SST)	°C	2.5 arcmin	2003-2007	Aqua-MODIS (NASA)	max, min, average (day & night)
Diffuse Attenuation (DA)	m <sup>-1</sup>	2.5 arcmin	2003-2007	Aqua-MODIS (NASA)	max, min, average
Calcite concentration (Ca)	moles/m <sup>3</sup>	2.5 arcmin	2006	Aqua-MODIS (NASA)	average
Chlorophyll A (chlA)	mg/m <sup>3</sup>	5 arcmin	1998-2007	SeaWiFS (NASA)	max, min, average
Photosynthetically Active Radiation (PAR)	Einstein/m <sup>2</sup> /day	5 arcmin	1998-2007	SeaWiFS (NASA)	max, min, average

## EVOLUTIONARY ANALYSIS OF NICHE CHARACTERISTICS

The evolutionary dynamics of niche features were studied by inferring their patterns of change along the chronogram in a maximum likelihood (ML) framework. The macroecological niche features included in our study are continuous variables and we inferred their evolution with common models of continuous trait evolution. Brownian motion models, also known as constant-variance random walk models, assume that traits vary naturally along a continuous scale and that variation is accumulated proportionally to evolutionary time, as measured by the branch lengths in a chronogram (Martins & Hansen, 1997; Pagel, 1999). Two branch length scaling parameters ( $\lambda$  and  $\kappa$ ) were used to extend this model and better describe the mode and tempo of trait evolution (Pagel, 1999).  $\lambda$  transformations measure the amount of phylogenetic signal present in a continuous character. The transformation consists of multiplying all internal branch lengths of the tree by  $\lambda$ , leaving tip branches their original length. When the ML estimate of  $\lambda$  is close to one, the internal branches retain their original length, indicating strong phylogenetic signal in the trait. If  $\lambda$  approaches zero, the evolution of the trait is virtually independent of phylogeny.  $\kappa$  transformations measure the degree of punctuational versus gradual evolution of characters on a phylogeny, by raising all branch lengths to the power  $\kappa$ . If the ML estimate of  $\kappa$  is close to zero, all branch lengths approach unity, and path lengths become proportional to the number of lineage splitting events, suggesting that the evolution of the trait approximates punctuated evolution associated with speciation events. If the ML estimate of  $\kappa$  is close to one, branch lengths remain unchanged, indicating that the amount of change in the character is proportional to evolutionary time. In other words,  $\kappa$  values close to one indicate gradual evolution.

In order to fit the models above and infer changes of the macroecological niche along the species phylogeny, a species  $\times$  variables matrix had to be constructed. To achieve this, the values of the macroecological data layers were extracted for each sample locality. For each species, the minimum, maximum and average of each macroecological parameter were stored in the species  $\times$  variables matrix. To reduce the influence of geographical sampling bias on the average values, they



were calculated by weighted averaging. The Euclidean distance from the sample location to the centre of gravity for the species in question was used as the sample weight. The centre of gravity for the species was determined by averaging the three-dimensional Cartesian coordinates of all sample locations for that species.

The models of continuous trait evolution listed above were optimized along the phylogenetic tree for the minimum, average and maximum values of a selection of niche variables using the ML optimization of the GEIGER package (Harmon *et al.*, 2008). Ancestral character values for macroecological niche features were estimated by ML inference (Schluter *et al.*, 1997) with the APE package (Paradis *et al.*, 2004). Resulting ancestral state values were plotted on the phylogeny with TreeGradients v1.03 (Verbruggen, 2008).

## NICHE MODELING PROCEDURE

The macroecological niches of species were modeled with Maxent, a presence-only niche modeling technique based on the maximum entropy principle (Phillips *et al.*, 2006). We used a presence-only technique because only specimen collection data are available and absence data cannot be reliably obtained for seaweed species on a global scale. Maxent has shown remarkably good performance in a comparative study of presence-only niche modeling techniques (Elith *et al.*, 2006). It estimates the probability distribution of maximum entropy (i.e. that is most spread out, or closest to uniform) of each macroecological variable across the study area. This distribution is calculated with the constraint that the expected value of each macroecological variable under the estimated distribution matches the empirical average generated from macroecological values associated with species occurrence data. The model output consists of a spatially explicit probability surface that represents an ecological niche (habitat suitability) translated from macroecological space into geographical space. The output grid is in the logistic format, where each pixel value represents the estimated probability that the species can be present at that pixel (Phillips & Dudik, 2008).

To avoid using redundant and correlated macroecological layers for niche modeling, an unstandardized principal components analysis was performed on the original variables in IDRISI Andes. The first, second and third principle component grids, which together accounted for 98.82% of the overall variance in the original variables, were exported for subsequent use in Maxent.

Global species' niches were modeled for all *Halimeda* species for which more than ten distribution records were available, while excluding species with distribution records suffering from high spatial autocorrelation. Additionally, a single regional model was generated using pooled occurrence data of six Caribbean species (*H. goreauii*, *H. simulans*, *H. incrassata*, *H. monile*, *H. discoidea.atl* and *H. tuna.car*).

The Maxent algorithm was run with default parameters (convergence threshold =  $10^{-5}$ , maximum iterations = 500, regularization multiplier = 1, maximum number of background points = 10000, and use of linear, quadratic, product and hinge features). Models were created using 80% of the localities for model training and 20% for model testing.

Statistical evaluation of the models was based on threshold-independent receiver operating characteristic (ROC) analysis (Phillips *et al.*, 2006). For presence-only modeling, the ROC curve is a plot of sensitivity (proportion of correctly predicted presences) against the fractional area predicted present. The area under the ROC curve (AUC) is subsequently compared to the area under the null expectations line connecting the origin and (1,1), thus providing a measure of predictive model performance. An AUC approximating 1 would mean optimal discrimination of suitable versus unsuitable sites, whereas an AUC between 0 and 0.5 is indicative of predictions no better than random. Additionally, we use a modified AUC based on partial ROC curves as proposed by Peterson



*et al.* (2008). This approach accounts for a user-defined maximum acceptable omission error, which we set at 0.1, and takes only the range of acceptable models in terms of omission error into account. The partial AUC is then rationed to the partial area under the null expectations line. Hence, the AUC ratio equals one for models performing no better than random, and increases with improving model accuracy. All partial AUC calculations were performed in the R statistical computing environment (R Development Core Team, 2008).

# RESULTS

## SPECIES DELIMITATION AND PHYLOGENY

Neighbor Joining analysis of the UCP7, ITS and *tufA* sequence alignments pointed out 52 clusters with low sequence divergence within clusters and relatively high divergence between clusters, as is typically found at the species boundary (Hebert *et al.*, 2004; Verbruggen *et al.*, 2005a). Not all clusters corresponded to described, named species. The undescribed clusters represent cryptic or pseudo-cryptic species (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005a; Verbruggen *et al.*, 2005b). The clusters inferred from DNA data formed the basis of the species definitions used in the remainder of the paper. After addition of morphologically identified herbarium specimens, the database consisted of 1080 samples from 538 unique localities. Analysis of the concatenated alignment of *rbcl*, *tufA*, UCP3, UCP7, 18S and ITS sequences (4965 characters) yielded a well-resolved species phylogeny in which five lineages, corresponding to the genus' five sections, could be recognized (Figure 1).



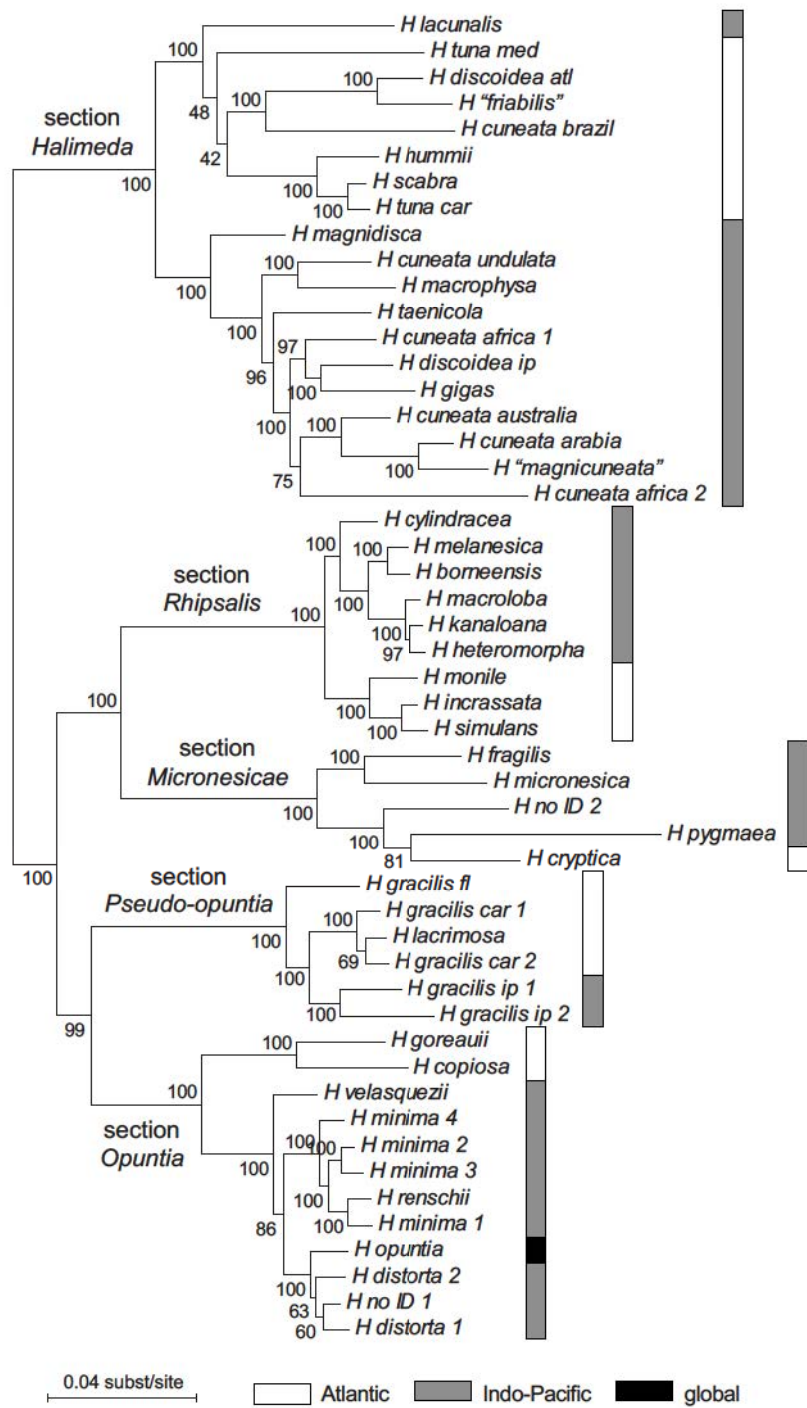


Figure 1. Phylogenetic tree of 52 *Halimeda* species inferred from six molecular loci using Bayesian techniques, rooted at the point where root-to-tip path length variance is minimal. Numbers at nodes indicate statistical support (Bayesian posterior probabilities, in percentages).

## EVOLUTION OF NICHE CHARACTERISTICS

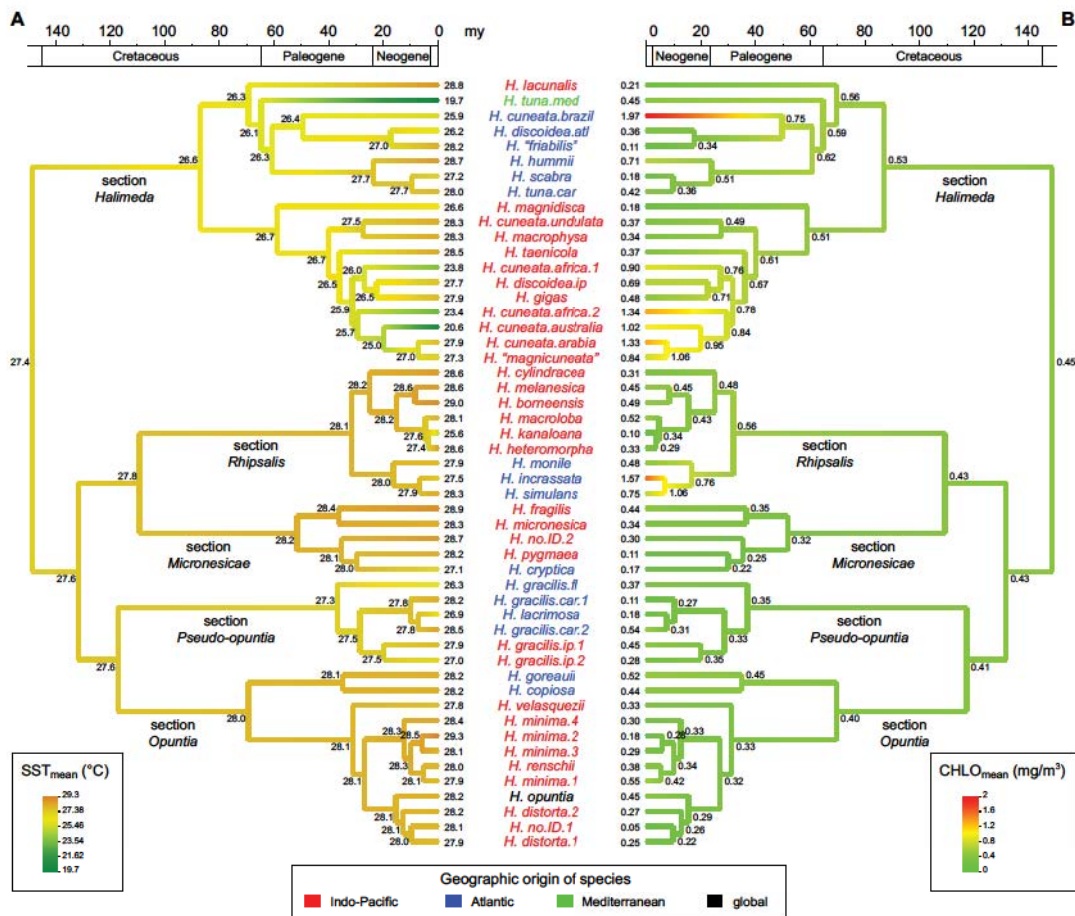
A few niche features contained considerable amounts of phylogenetic signal, as indicated by the high ML estimates of lambda values using Pagel's lambda branch length modifier (Table 2). A general observation was that average trait values contained more phylogenetic signal than minimum and maximum trait values (e.g., average temperature, not minimum or maximum temperature). High kappa values for the average trait values indicate that change of these traits is proportional to evolutionary time; in other words, change is gradual (Table 2). Some traits that also contained phylogenetic signal were not included in the table because of significant correlation with the listed variables. This is the case for photosynthetically active radiation, which is correlated with sea surface temperatures, and diffuse attenuation, which is correlated with chlorophyll values (caused by phytoplankton).

Figure 2 illustrates the estimated evolutionary patterns of average annual temperature and chlorophyll values. Estimated ancestral trait values are shown at the internal nodes and visualized using a color gradient. An average annual temperature of 27.4°C (95% confidence interval: 25.6–29.2) is inferred at the basal split (Figure 2A), indicating a tropical origin for the genus.

**Table 2.** Optimum values the branch length scaling parameters  $\lambda$  and  $\kappa$  used to test the mode and tempo of evolution of niche features. The niche traits are sea surface temperature (SST) and chlorophyll A values. The high optimal  $\lambda$  values inferred for average trait values indicate strong phylogenetic signal in these traits whereas the low  $\lambda$  values obtained for the minimum and maximum traits suggest a lack of phylogenetic signal. The relatively high optimum values for  $\kappa$  for average SST values suggest that evolution of this niche feature was more or less gradual (proportional to time). The lower value for average chlorophyll A suggests that there is a non-negligible punctuated component to the evolution of nutrient preferences. The first two columns used the tree smoothed with the additive penalty; the last two columns used the tree smoothed with the log-additive penalty. The  $\kappa$  parameter was not optimized when there was poor phylogenetic signal in the data (low  $\lambda$ ).

Trait	Optimal $\lambda$	Optimal $\kappa$	Optimal $\lambda$	Optimal $\kappa$
max SST	0.07262		0.05448	
average SST	0.90087	0.83616	0.15159	0.66764
min SST	0.09371		0.05308	
min chlA	0.10672		0.06804	
average chlA	0.78528	0.47084	0.80792	0.37894
max chlA	0.01364		0.00000	





**Figure 2.** Inferred evolutionary history of niche features in *Halimeda*. Ancestral values for (A) mean sea surface temperature and (B) mean chlorophyll concentration are plotted along the phylogeny. Numbers plotted at nodes indicate the inferred ancestral values. These values were obtained using a maximum likelihood approach as described in the text. Values are also drawn along a color gradient to allow rapid visual assessment of evolutionary patterns. Green indicates low values, red stands for high values, and yellowish colors indicate intermediate values. The geographic origin of species is indicated with colored taxon names.

The tree clearly shows that evolution along the SST niche dimension is not homogeneous throughout the tree. Whereas the sections *Rhipsalis*, *Micronesicae*, *Pseudo-opuntia* and *Opuntia* barely deviate from typical tropical temperatures, evolution along the temperature axis has been common in section *Halimeda*. More specifically, the lineages leading to *H. tuna.med*, *H. cuneata.africa.1*, *H. cuneata.africa.2* and *H. cuneata.australia* have evolved a preference for colder water. Chlorophyll values were mapped onto the phylogeny as a proxy for nutrient preferences (figure 2B). Deviations from the average (low) nutrient preference values are present in *Halimeda* section *Halimeda* (*H. cuneata.brazil*, *H. cuneata.africa.1*, *H. cuneata.africa.2*, *H. cuneata.australia*, *H. cuneata.arabia* & *H. magnicuneata*) and in *Halimeda* section *Rhipsalis* (*H. incrassata*).

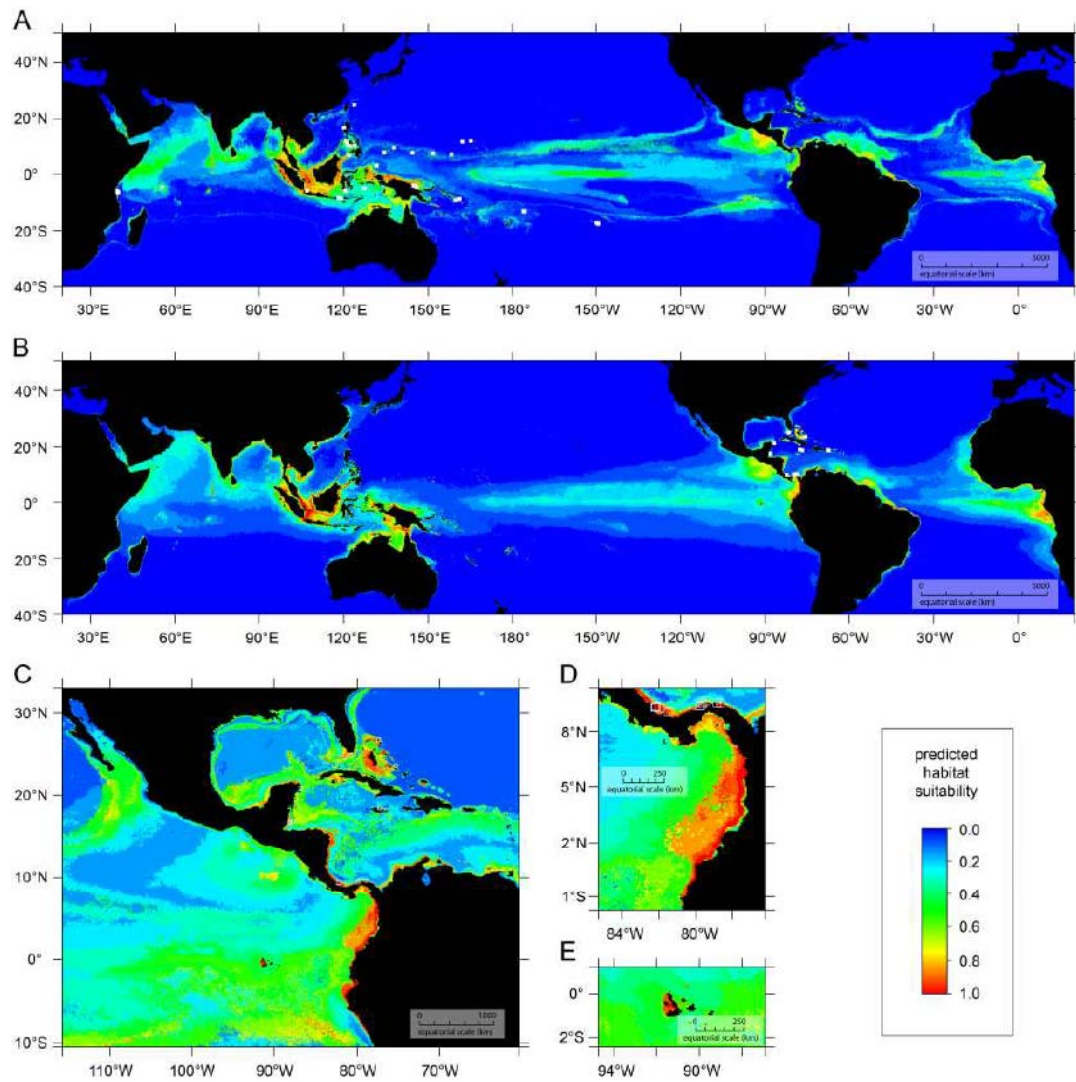
## NICHE MODELS AT THE GLOBAL SCALE

Niche models indicating the areas where macroecological conditions are suitable for species to occur were generated for all species (figure 3 and figure S1). The average AUC across all models with 20% test localities was 0.917 (SD = 0.046) for the training data and 0.906 (SD = 0.054) for the test data. The corresponding average AUC ratios were 1.576 (SD = 0.209) for the training data and 1.615 (SD = 0.234) for the test data. The high AUC values and ratios indicate that the most essential macroecological variables determining species distributions were accounted for in the dataset. The high scores for the test data indicate adequate model performance rather than overfitting of the model on the training data. The predicted distributions are clearly broader than the species known distributions. For example, the distribution model of the exclusively Caribbean species *H. simulans* (figure 3A) predicts habitat suitability in parts of the Indo-Pacific basin. Similarly, the model of the Indo-Pacific species *H. borneensis* (figure 3B) predicts habitat suitability in parts of the Atlantic Ocean. In general, there was a stronger tendency of predicting Atlantic species into the Indo-Pacific than vice versa.

## NICHE MODEL AT THE REGIONAL SCALE

The model predicting suitable habitat for a suite of six Caribbean species is shown in figure 3C (AUC ratio = 1.783). Potentially suitable habitats of these Caribbean species in the East Pacific are mainly predicted along the southern coast of Panama, the western coast of Colombia and in the Galapagos Islands (figure 3C–E).





**Figure 3. Predictive ecological niche models of *Halimeda* species inferred from environmental data and species occurrence records. (A) Niche model of the exclusively Indo-Pacific species *H. borneensis* indicating habitat suitability in some Atlantic regions. (B) Niche model of the exclusively Caribbean species *H. simulans* predicting habitat suitability in several Indo-Pacific regions. (C) Pooled niche model of six Caribbean *Halimeda* species predicting habitat suitability along parts of the Pacific coast of Central America. (D) Detailed view of the areas along the Pacific coastlines of Panama, Colombia and Ecuador predicted by the model from panel C. (E) Detailed view of the Galapagos archipelago as predicted by the model from panel C. Predicted habitat suitability is indicated with colors along a gradient, warmer colors indicating areas with better predicted conditions. White squares indicate specimen localities used for model training. All maps are equidistant cylindrical projections.**

# DISCUSSION

The obtained results invite discussion about several issues related to the macroecological niche of seaweeds, how it evolves and how it relates to patterns of biogeography.

## MODELING SEAWEED DISTRIBUTIONS

### Niche modeling vs. previous approaches

Our niche models indicate areas where the macroecological conditions are likely to be suitable for various *Halimeda* species to establish populations. They reflect the marked tropical nature of most species and show that many species occupy only part of the potentially suitable habitat (see below). Previous knowledge about the macroecological niche of seaweeds mainly stemmed from comparing distribution ranges with isotherms (isotherm fitting), studying survival and growth under various culture conditions, or the combination of both (e.g., van den Hoek, 1982). These approaches and the niche modeling approach presented here differ from each other in a number of aspects. Whereas the fundamental niche is investigated with *in vitro* studies of survival and growth, the realized niche is central in modeling techniques and isotherm fitting. A fundamental difference between niche modeling and both other approaches is that the former yields probabilistic output whereas the latter usually propose hard thresholds. The ease with which a niche modeling study can be carried out has benefits as well as drawbacks. The advantage is obvious when targeting species that are difficult to grow in culture. A disadvantage of niche modeling is that the choice of a specific niche modeling algorithm and the parameter settings may influence niche predictions and predictive model performance (Elith *et al.*, 2006; Peterson *et al.*, 2008). The maximum entropy method with ROC modifications appeared to be the most suitable option for our goals.

### Taxonomic caveat

An additional concern about the application of niche models in seaweed research is the ease with which heterogeneous distribution records can be used to generate models. As mentioned earlier, morphological species delimitation is troublesome in algae and, as a consequence, published species occurrence records based on morphological identifications are not always meaningful. We have taken great caution to avoid identification errors through DNA-guided species delimitation. All methods share the drawback of being sensitive to specimen sampling. In this respect, the absolute number of samples is likely to be of inferior importance compared to the spread of samples across relevant macroecological dimensions (Pearson, 2007).



## MACROEVOLUTION OF THE MACROECOLOGICAL NICHE

### Historical perspective

Evolutionary processes are influenced by environmental variation in space and time (Kozak *et al.*, 2008). Many studies taking a niche modeling approach to study environmental variation in a phylogenetic framework have shown strong heritability of macroecological preferences (e.g., Martinez-Meyer & Peterson, 2006; Yesson & Culham, 2006). To our knowledge, these studies have all focused on terrestrial organisms. The evolutionary dynamics of the niche of seaweeds have hardly been studied in the past. Breeman *et al.* (2002) investigated the evolution of temperature responses in the seaweed genus *Cladophora*. Their approach consisted of measuring cold tolerance, heat tolerance, and growth of various culture strains at different temperature regimes. The response variables (tissue damage and growth rates) were interpreted along a phylogenetic tree, leading to the conclusion that the two main lineages of the *Cladophora vagabunda* complex had divergent cold tolerances. Although the experimental data of this study differs from ours as discussed above, the approach taken to infer niche dynamics in both studies is not fundamentally different. However, thanks to the advances in models describing the evolution of continuous characters that have taken place since the publication of Breeman *et al.* (2002) and their implementation in user-friendly packages for the R statistical computing environment (Paradis *et al.*, 2004; Harmon *et al.*, 2008), much more detailed inferences can now be made.

### Niche conservatism

Our study shows that the macroecological niche in the seaweed genus *Halimeda* has a strong phylogenetic imprint and that niches appear to change gradually with time. The results clearly indicate the phylogenetic heritability of macroecological preferences: four out of five sections (*Rhipsalis*, *Micronesica*, *Opuntia* and *Pseudo-opuntia*) demonstrate conserved preference for high temperatures and low nutrient levels, confirming the association of these sections with tropical coral reefs and shallow lagoons (figure 2). Adaptation to colder and more nutrient-rich water only occurred in section *Halimeda*. Remarkably, the transition into colder water seems to have taken place four times independently (in *H. tuna.med*, *H. cuneata.africa.1*, *H. cuneata.africa.2* and *H. cuneata.australia*). The species *H. tuna.med* is the only one inhabiting the Mediterranean Sea and can maintain populations at sites with yearly sea surface temperature minima around 10°C. The species *H. cuneata.africa.1* and *H. cuneata.africa.2* occur in SE Africa. *H. cuneata.australia* is found along the shores of SW Australia. Chlorophyll values, used as a proxy for the trophic status of the surface water (Duan *et al.*, 2007), are above average for certain species in section *Halimeda*, often in the subtropical species. It is known that nutrient levels increase with latitude in the latitudinal range studied here (Sasai *et al.*, 2007). *Halimeda cuneata.brazil* occurred in waters with high average chlorophyll values due to an overall high concentration along the Brazilian coast. The high average chlorophyll value of waters in which *H. incrassata* was recorded is largely due to an outlier observation in Florida.

### Sources of uncertainty

Our study of evolutionary niche dynamics involves several subsequent analyses, hence a discussion of the potential sources of uncertainty affecting the final result is in place. The first source of uncertainty is in the species phylogeny. A lack of support of phylogenetic relationships will have direct repercussions on the accuracy of downstream analyses. In our study, the use of a multi-locus



alignment yielded very high statistical support for the great majority of branches in the tree. Therefore we have used the tree resulting from the Bayesian analysis (figure 1) in subsequent analyses as if it were known without uncertainty. Second, inferences of trait evolution also depend on branch lengths, which are affected by two potential sources of uncertainty: branch length estimation error in the phylogenetic analysis and error from the rate smoothing process that transforms the phylogram into a chronogram. Especially rate smoothing can lead to variation in branch lengths if different settings are used. We followed the recommendations in the manual of the r8s program. Third, the values used as character states of the terminal taxon influence the results. We used distance-weighted averages as fixed character states for the terminal taxa whereas in reality there is variation around the average. Taking this variation into account is expected to broaden confidence intervals on inferred ancestral states (Martins & Hansen, 1997). A fourth source of error could result from the inability of Brownian motion models to capture the complexity of historical forces affecting niche evolution, a source of error inherent to using simple models to describe a more complex reality. The last element of uncertainty lies in the ancestral character estimation, which infers values for ancestral taxa based on values of recent taxa. These analyses, however, report the 95% confidence intervals around the inferred value. If a character evolves fast, this will be reflected in broader confidence intervals on ancestral character states (Martins, 1999). We have not attempted to quantify the accumulation of uncertainty throughout our sequence of analyses due to practical limitations but the reader should be aware of the assumptions that were made.

## Paleobiological perspective

Despite the relatively high levels of uncertainty usually associated with ancestral state estimation of continuous characters (Schluter *et al.*, 1997), the observed conservatism for environmental preferences yields a relatively narrow 95% confidence interval for the average sea surface temperature characterizing the habitat of the most recent common ancestor of extant *Halimeda* species (25.6–29.2°C). The ML estimate of 27.4°C appears to be in agreement with the tropical Tethyan origin of *Halimeda* that was previously derived from the fossil record. The earliest known fossil that is considered to belong to the genus is *Halimeda soltanensis* from the Upper Permian ( $\pm 250$ – $270$  my) of Djebel Tebaga in South Tunisia (Poncet, 1989), which was at that time located at a low latitude along the western shore of the Tethys Ocean (Smith *et al.*, 1994). A more diverse assemblage of species with a markedly tropical distribution had evolved by the Upper Cretaceous ( $\pm 100$ – $65$  my) (Dragastan & Herbig, 2007). The invasion of *Halimeda* into higher latitudes has not been documented in the fossil record. Our chronogram suggests that the invasion occurred during late Paleogene and Neogene times, a period characterized by global cooling (Zachos *et al.*, 2001). This finding confirms earlier hypotheses that at least parts of the warm-temperate seaweed floras originated from tropical ancestry during this period of globally decreasing temperatures (van den Hoek, 1984; Lüning, 1990).

## GLOBAL BIOGEOGRAPHY

### Dispersal limitation

*Halimeda* species have previously been shown to be geographically restricted to either the Atlantic Ocean or the Indo-Pacific basin (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005a; Verbruggen *et al.*, 2005b). One could ask whether the absence of Atlantic species in the Indo-Pacific (and vice versa) is a consequence of dispersal limitation or if habitat differences may be responsible for the limited



distributions. The niche model of the Indo-Pacific species *H. borneensis* clearly indicates that some parts of the Caribbean Sea would be suitable habitat (figure 3A) and the niche model of the Atlantic species *H. simulans* suggests that it could survive in large parts of the Indo-Pacific tropics (figure 3B). Similar patterns were observed for other species (figure S1). So, unless *Halimeda* species are limited by habitat differences between the Atlantic and Indo-Pacific basins that are not represented in our macroecological data, it can be concluded that dispersal limitation is the most likely explanation for the strong separation of Atlantic and Indo-Pacific species. Dispersal limitation of benthic tropical marine organisms between oceans is not uncommon (Lessios *et al.*, 2001; Teske *et al.*, 2007) and can be explained by the North–South orientation of the African and American continents, prohibiting marine dispersal between the Atlantic and Indo-Pacific basins through tropical waters. *Halimeda opuntia* is the only species that occurs in both ocean basins. It is part of a clade of Indo-Pacific species, indicating that it originated in the Indo-Pacific basin and subsequently dispersed to the Atlantic Ocean and spread throughout its tropical regions. It was previously suggested that *H. opuntia* was introduced in the Atlantic Ocean by early interoceanic shipping (Kooistra & Verbruggen, 2005). If this scenario is correct, our models' prediction that parts of the tropical Atlantic Ocean form suitable habitat for Indo-Pacific species and the conclusion of dispersal limitation between ocean basins would be confirmed.

## Vicariance patterns

Geographic distribution patterns show a clear phylogenetic signal: each section separates largely into an Atlantic and an Indo-Pacific lineage (figure. 1), confirming previous observations (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005b). This pattern indicates ancient lineage splitting through vicariance and subsequent diversification within the Atlantic and Indo-Pacific basins. A number of geological events are commonly invoked to explain sister relationships between strictly Atlantic and strictly Indo-Pacific lineages. The first is the spreading of the Atlantic Ocean, which started during the Jurassic ( $\pm 170$ – $160$  my) (Smith *et al.*, 1994). The second is the collision of the African and Eurasian plates in the Middle East during the Miocene ( $\pm 15$ – $12$  my) (Rögl & Steininger, 1984). The third event is the closure of the Central American Seaway in the Pliocene ( $\pm 3$  my) (Coates & Obando, 1996). Different events have been hypothesized to be at the basis of the geographic splits in *Halimeda* but results have remained inconclusive (Kooistra *et al.*, 2002; Verbruggen *et al.*, 2005b). Our chronogram suggests that the splits between Atlantic and Indo-Pacific lineages originated at various times during the Paleogene (65–25 my). In other words, the time frame of initial divergence does not correspond closely with either one of the geological events. During the Paleogene, however, an important oceanographic event that may have limited dispersal between the Atlantic and Indo-Pacific ocean basins took place: the circum-equatorial current that homogenized the tropical marine biome during the Cretaceous was deflected to south of Africa (Lawver & Gahagan, 2003). This result suggests that geological barriers may not be the initial cause of divergence between populations but instead act as barrier reinforcements after divergence has been initiated by oceanographic events. A similar conclusion was reached in molecular and paleontological studies of species across the Central American Isthmus (e.g., Collins *et al.*, 1996; Knowlton & Weigt, 1998). The generality of this pattern requires additional study. At least for some organisms, divergence times between Atlantic and Indo-Pacific lineages obtained with a molecular clock match more closely with the timing of the collision of the African and Eurasian plates in the Middle East (e.g., Teske *et al.*, 2007).

## REGIONAL BIOGEOGRAPHY OF TROPICAL AMERICA

As an alternative to the molecular clock, one would also be able to infer which geological events were involved in species partitioning between Atlantic and Indo-Pacific through a thorough study of eastern Pacific *Halimeda* species. The Caribbean and East Pacific formed a single tropical marine biota that was separated by the shoaling of the Central American Isthmus during the Pliocene, resulting in the formation of many trans-isthmian sister species (Knowlton & Weigt, 1998). The emergence of a land bridge has been dated at approximately three million years (Coates & Obando, 1996). The presence of trans-isthmian species pairs with a distribution limited to the tropical Americas (i.e., not in the wider Indo-Pacific) can be taken as evidence for vicariance across the Central American Isthmus.

Only *Halimeda discoidea* has been reported from the East Pacific and, curiously, molecular analyses have shown these populations not to be related to the Caribbean species *H. discoidea.atl* as one may expect but to the Indo-Pacific species *H. discoidea.ip* (Verbruggen *et al.*, 2005b). So either *Halimeda* does not have trans-isthmian species pairs in the tropical Americas or they have not been discovered yet. The seaweed flora of the tropical East Pacific Ocean has not been studied in great detail in the past and recent inventories have shown lots of new discoveries (Wysor, 2004). We aimed to facilitate the discovery of trans-isthmian sister pairs by identifying geographic regions in the East Pacific Ocean that are hotspots of habitat suitability for Caribbean species. The niche model of pooled distribution data of six Caribbean species predicted parts of the East Pacific Ocean as suitable habitat (figure 3C) and identified three hotspots of habitat suitability: the western Galapagos Islands (figure 3E), the West coast of Colombia and parts of the South coast of Panama (figure 3D). We suggest that these areas should be targeted in future research expeditions aiming to discover trans-isthmian species pairs. The utility of ecological niche models to guide discovery has already been documented. Unexplored deep-water kelp forests were recently found in the Galapagos archipelago based on predictions of a synthetic oceanographic and ecophysiological model (Graham *et al.*, 2007). Similarly, expeditions directed by niche models of chameleons led to the discovery of additional populations of known species and several species new to science (Raxworthy *et al.*, 2003). It should be noted that the niche model presented here predicts habitat suitability only as a function of the macroecological variables included in the dataset. It is beyond doubt that factors not included in our dataset (e.g., micro-habitat characteristics, tidal amplitudes, grazing pressure and other biotic interactions) affect the actual distribution of species. If such data were available, it could be used to create a more specific model and would likely result in smaller hotspots, allowing even more targeted expeditions.



## *ACKNOWLEDGEMENTS*

We thank W. Willems for providing the R script to calculate partial AUC values and for discussion of techniques. We are grateful to M. Accioly, K. Arano, M. Bandeira-Pedrosa, C. Battelli, B. Brooks, K. Clifton, M. Coffroth, P. Colinvaux, R. Collin, E. Coppejans, O. Dargent, Y. de Jong, G. De Smedt, E. Demeulenare, R. Díaz, E. Drew, S. Fredericq, C. Galanza, S. Guimaraes, F. Gurgel, O. Gussmann, R. Haroun, I. Hendriks, J. Hernandez, L. Hillis, J. Huisman, M. Kaufmann, L. Kirkendale, L. Liao, D. Littler, M. Littler, G. Llewellyn, P. Marshall, J. Maté, A. Maypo, A. N'Yeurt, D. Olandesca, C. Ortuno, K. Page, F. Parrish, C. Payri, G. Procaccini, W. Prud'homme van Reine, L. Raymundo, T. Schils, E. Tronchin, M. Van Veghel, P. Vroom, S. Williams, S. Wilson, B. Wysor and J. Zuccarello for providing specimens. Funding was provided by the Research Foundation – Flanders (research grant G.0142.05 and post-doctoral fellowships to HV and FL) and Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT, fellowship LT).

## *AUTHOR CONTRIBUTIONS*

LT, HV and KP designed the general concept of the study. HV conducted the phylogenetic analyses. LT carried out all other analyses and wrote an initial draft of the manuscript. KP supervised the GIS analyses and niche modeling. HV finalized the manuscript. Other authors have carried out technical work and commented on the manuscript.

# REFERENCES

- Beiko R G, Keith J M, Harlow T J & Ragan MA (2006) Searching for convergence in phylogenetic Markov chain Monte Carlo. *Systematic Biology*, **55**, 553-565.
- Breeman AM, Oh YS, Hwang MS & van den Hoek C (2002) Evolution of temperature responses in the *Cladophora vagabunda* complex and the *C. albida/sericea* complex (Chlorophyta). *European Journal of Phycology*, **37**, 45-58.
- Coates AG & Obando JA (1996) The geologic evidence of the Central American Isthmus. *Evolution and environment in tropical America* (ed. by B.C. Jackson, A.F. Budd & A.G. Coates), pp. 21-56. University of Chicago Press, Chicago.
- Collins LS, Budd AF & Coates AG (1996) Earliest evolution associated with closure of the Tropical American Seaway. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 6069-6072.
- Dragastan ON & Herbig HG (2007) *Halimeda* (green siphonous algae) from the Paleogene of (Morocco) – Taxonomy, phylogeny and paleoenvironment. *Micropaleontology*, **53**, 1-72.
- Drew EA (1983) *Halimeda* biomass, growth rates and sediment generation on reefs in the central Great Barrier Reef province. *Coral Reefs*, **2**, 101-110.
- Duan H, Zhang Y, Zhang B, Song K. & Wang Z (2007) Assessment of chlorophyll- a concentration and trophic state for lake Chagan using Landsat TM and field spectral data. *Environmental Monitoring and Assessment*, **129**, 295-308.
- Elith, J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS & Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.
- Famà P, Wysor B, Kooistra W & Zuccarello GC (2002) Molecular phylogeny of the genus *Caulerpa* (Caulerpales, Chlorophyta) inferred from chloroplast *tufA* gene. *Journal of Phycology*, **38**, 1040-1050.
- Freile D, Milliman JD & Hillis L (1995) Leeward bank margin *Halimeda* meadows and draperies and their sedimentary importance on the western Great Bahama bank slope. *Coral Reefs*, **14**, 27-33.
- Graham CH, Ron SR, Santos JC, Schneider CJ & Moritz C (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1781-1793.
- Graham MH, Kinlan BP, Druehl LD, Garske LE & Banks S (2007) Deep-water kelp refugia as potential hotspots of tropical marine diversity and productivity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 16576-16580.
- Guisan A & Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993-1009.
- Harmon LJ, Weir JT, Brock CD, Glor RE & Challenger W (2008) GEIGER: investigating evolutionary radiations. *Bioinformatics*, **24**, 129-131.
- Hebert PDN, Penton EH, Burns JM, Janzen DH & Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 14812-14817.
- Hillis-Colinvaux L (1980) Ecology and taxonomy of *Halimeda*: primary producer of coral reefs. *Advances in Marine Biology*, **17**, 1-327.
- Holder M & Lewis PO (2003) Phylogeny estimation: Traditional and Bayesian approaches. *Nature Reviews Genetics*, **4**, 275-284.
- Hutchinson GE (1957) Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology*, **22**, 415-427.
- Jensen PR, Gibson RA, Littler MM & Littler DS (1985) Photosynthesis and calcification in four deep-water *Halimeda* species (Chlorophyceae, Caulerpales). *Deep Sea Research Part I: Oceanographic Research Papers*, **32**, 451-464.
- Knouft JH, Losos JB, Glor RE & Kolbe JJ (2006) Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. *Ecology*, **87**, S29-S38.
- Knowlton N & Weigt LA (1998) New dates and new rates for divergence across the Isthmus of Panama. *Proceedings of the Royal Society of London Series B - Biological Sciences*, **265**, 2257-2263.
- Kooistra W, Coppejans EGG & Payri C (2002) Molecular systematics, historical ecology, and phylogeography of *Halimeda* (Bryopsidales). *Molecular Phylogenetics and Evolution*, **24**, 121-138.



- Kooistra W & Verbruggen H (2005) Genetic patterns in the calcified tropical seaweeds *Halimeda opuntia*, *H. distorta*, *H. hederacea* and *H. minima* (Bryopsidales, Chlorophyta) provide insights in species boundaries and inter-oceanic dispersal. *Journal of Phycology*, **41**, 177-187.
- Kozak KH, Graham CH & Wiens JJ (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution*, **23**, 141-148.
- Lam DW & Zechman FW (2006) Phylogenetic analyses of the Bryopsidales (Ulvophyceae, Chlorophyta) based on Rubisco large subunit gene sequences. *Journal of Phycology*, **42**, 669-678.
- Lawver LA & Gahagan LM (2003) Evolution of Cenozoic seaways in the circum-Antarctic region. *Palaeogeography Palaeoclimatology Palaeoecology*, **198**, 11-37.
- Lessios HA, Kessing BD & Pearse JS (2001) Population structure and speciation in tropical seas: Global phylogeography of the sea urchin *Diadema*. *Evolution*, **55**, 955-975.
- Lüning K (1990) *Seaweeds: their environment, biogeography, and ecophysiology*. Wiley & Sons, New York.
- Martinez-Meyer E & Peterson AT (2006) Conservatism of ecological niche characteristics in North American plant species over the Pleistocene-to-Recent transition. *Journal of Biogeography*, **33**, 1779-1789.
- Martins EP (1999) Estimation of ancestral states of continuous characters: A computer simulation study. *Systematic Biology*, **48**, 642-650.
- Martins EP & Hansen TF (1997) Phylogenies and the comparative method: A general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist*, **149**, 646-667.
- Naim O (1988) Distributional patterns of mobile fauna associated with *Halimeda* on the Tiahura coral-reef complex (Moorea, French Polynesia). *Coral Reefs*, **6**, 237-250.
- Pagel M (1999) Inferring the historical patterns of biological evolution. *Nature*, **401**, 877-884.
- Paradis E, Claude J & Strimmer K (2004) APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289-290.
- Pearson RG (2007) Species' distribution modeling for conservation educators and practitioners - synthesis. In: American Museum of Natural History.
- Peterson AT, Papes M & Soberón J (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling*, **213**, 63-72.
- Phillips SJ, Anderson RP & Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231-259.
- Phillips SJ & Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161-175.
- Poncet J (1989) Présence du genre *Halimeda* Lamouroux, 1812 (algue verte calcaire) dans le Permien supérieur du sud Tunisien. *Revue de Micropaléontologie*, **32**, 40-44.
- Provan J, Murphy S & Maggs CA (2004) Universal plastid primers for Chlorophyta and Rhodophyta. *European Journal of Phycology*, **39**, 43-50.
- R Development Core Team (2008) R: A language and environment for statistical computing. <http://www.r-project.org>
- Rambaut A & Drummond AJ (2007) Tracer. <http://beast.bio.ed.ac.uk/tracer>
- Raxworthy CJ, Ingram CM, Rabibisoa N & Pearson RG (2007) Applications of ecological niche modeling for species delimitation: A review and empirical evaluation using day geckos (*Phelsuma*) from Madagascar. *Systematic Biology*, **56**, 907 - 923.
- Raxworthy CJ, Martinez-Meyer E, Horning N, Nussbaum RA, Schneider GE, Ortega-Huerta MA & Peterson AT (2003) Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, **426**, 837-841.
- Rissler LJ & Apodaca JJ (2007) Adding more ecology into species delimitation: Ecological niche models and phylogeography help define cryptic species in the black salamander (*Aneides flavipunctatus*). *Systematic Biology*, **56**, 924 - 942.
- Rissler LJ, Hijmans RJ, Graham CH, Moritz C & Wake DB (2006) Phylogeographic lineages and species comparisons in conservation analyses: A case study of California herpetofauna. *American Naturalist*, **167**, 655-666.
- Rögl F & Steininger FF (1984) Neogene Paratethys, Mediterranean and Indo-Pacific seaways. *Fossils and climate* (ed. by P. Brenchley), pp. 171-200. John Wiley & Sons.
- Ronquist F & Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572-1574.
- Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution*, **19**, 101-109.
- Sanderson MJ (2003) r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, **19**, 301-302.



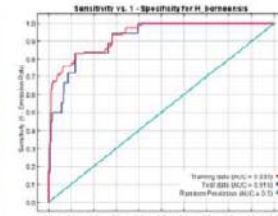
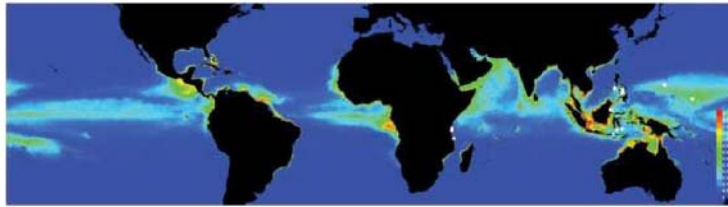
- Sasai Y, Sasaoka K, Sakaki H & Ishida A (2007) Seasonal and intra-seasonal variability of chlorophyll-a in the North Pacific: model and satellite data. *Journal of the Earth Simulator*, **8**, 3-11.
- Saunders GW & Lehmkuhl KV (2005) Molecular divergence and morphological diversity among four cryptic species of *Plocamium* (Plocamiales, Florideophyceae) in northern Europe. *European Journal of Phycology*, **40**, 293-312.
- Schluter D, Price T, Mooers AO & Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. *Evolution*, **51**, 1699-1711.
- Smith AG, Smith DG & Funnell BM (1994) *Atlas of Mesozoic and Cenozoic coastlines*. Cambridge University Press, Cambridge.
- Sullivan J (2005) Maximum-likelihood methods for phylogeny estimation. *Molecular Evolution: Producing the Biochemical Data, Part B. Methods in Enzymology*, **395**, 757-779.
- Swofford DL, Waddell PJ, Huelsenbeck JP, Foster PG, Lewis PO & Rogers JS (2001) Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology*, **50**, 525-539.
- Teske P, Hamilton H, Matthee C & Barker N (2007) Signatures of seaway closures and founder dispersal in the phylogeny of a circumglobally distributed seahorse lineage. *BMC Evolutionary Biology*, **7**, 138.
- van den Hoek C (1982) The distribution of benthic marine algae in relation to the temperature regulation of their life histories. *Biological Journal of the Linnean Society*, **18**, 81-144.
- van den Hoek C (1984) World-wide latitudinal and longitudinal seaweed distribution patterns and their possible causes, as illustrated by the distribution of rhodophyten genera. *Helgolander Meeresuntersuchungen*, **38**, 227-257.
- Verbruggen H (2005) Resegmenting *Halimeda*. Molecular and morphometric studies of species boundaries within a green algal genus. In *Phycology Research Group*. Ghent University, Ghent, p. 213.
- Verbruggen H (2008) TreeGradients. <http://www.phycoweb.net>
- Verbruggen H, Ashworth M, LoDuca ST, Vlaeminck C, Cocquyt E, Sauvage T, Zechman FW, Littler DS, Littler MM, Leliaert F & De Clerck O (2009) A multi-locus time-calibrated phylogeny of the siphonous green algae. *Molecular Phylogenetics and Evolution*, in press.
- Verbruggen H, De Clerck O, Kooistra W & Coppejans E (2005a) Molecular and morphometric data pinpoint species boundaries in *Halimeda* section *Rhipsalis* (Bryopsidales, Chlorophyta). *Journal of Phycology*, **41**, 606-621.
- Verbruggen H, De Clerck O, Schils T, Kooistra W & Coppejans E (2005b) Evolution and phylogeography of *Halimeda* section *Halimeda*. *Molecular Phylogenetics and Evolution*, **37**, 789-803.
- Wiens JJ & Graham CH (2005) Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annual Review of Ecology Evolution and Systematics*, **36**, 519-539.
- Wiens JJ, Parra-Olea G, Garcia-Paris M & Wake DB (2007) Phylogenetic history underlies elevational biodiversity patterns in tropical salamanders. *Proceedings of the Royal Society B-Biological Sciences*, **274**, 919-928.
- Wysor B (2004) An annotated list of marine Chlorophyta from the Pacific Coast of the Republic of Panama with a comparison to Caribbean Panama species. *Nova Hedwigia*, **78**, 209-241.
- Yesson C & Culham A (2006) Phyloclimatic modeling: combining phylogenetics and bioclimatic modeling. *Systematic Biology*, **55**, 785-802.
- Zachos J, Pagani M, Sloan L, Thomas E & Billups K (2001) Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, **292**, 686-693.



# APPENDIX

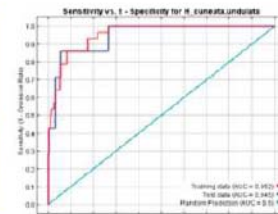
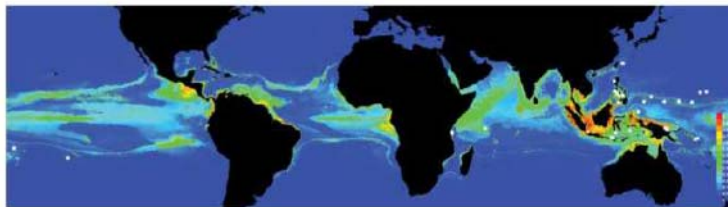
Figure S1. Model output for *Halimeda* species

*Halimeda borneensis*



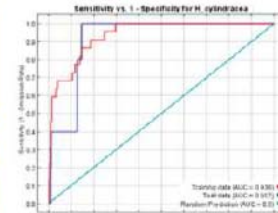
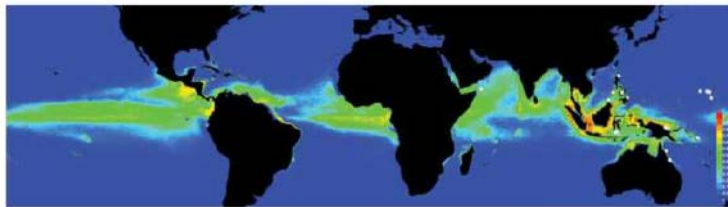
# Training samples = 76    # Test samples = 18    AUC (training) = 0.933    AUC (test) = 0.916    AUC ratio (training) = 1.540    AUC ratio (test) = 1.537

*Halimeda cuneata.undulata*



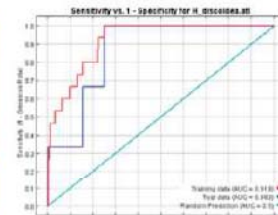
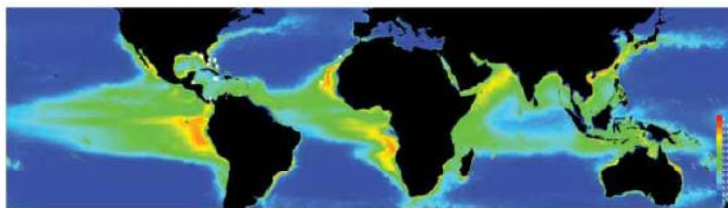
# Training samples = 28    # Test samples = 7    AUC (training) = 0.952    AUC (test) = 0.945    AUC ratio (training) = 1.691    AUC ratio (test) = 1.576

*Halimeda cylindracea*



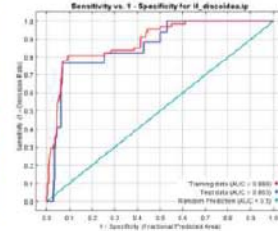
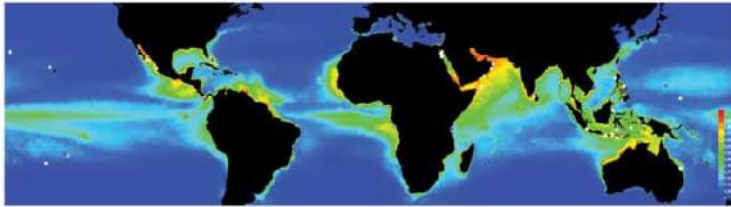
# Training samples = 22    # Test samples = 5    AUC (training) = 0.936    AUC (test) = 0.917    AUC ratio (training) = 1.652    AUC ratio (test) = 1.737

*Halimeda discoidea.atl*



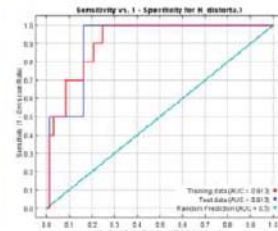
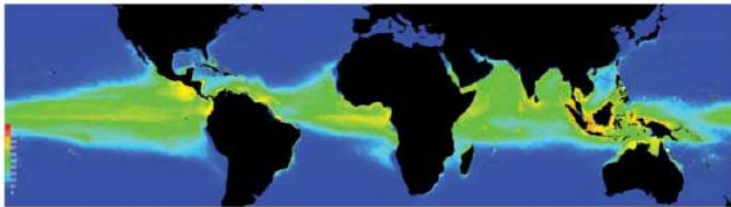
# Training samples = 12    # Test samples = 3    AUC (training) = 0.919    AUC (test) = 0.863    AUC ratio (training) = 1.554    AUC ratio (test) = 1.698

*Halimeda discoidea*.ip



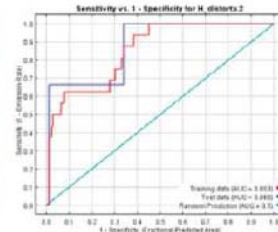
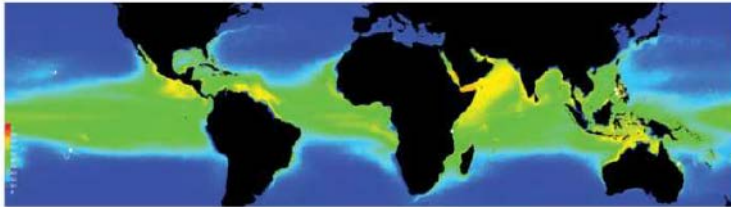
# Training samples = 68    # Test samples = 17    AUC (training) = 0.888    AUC (test) = 0.863    AUC ratio (training) = 1.392    AUC ratio (test) = 1.312

*Halimeda distorta*.1



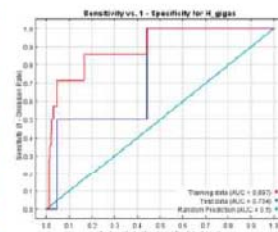
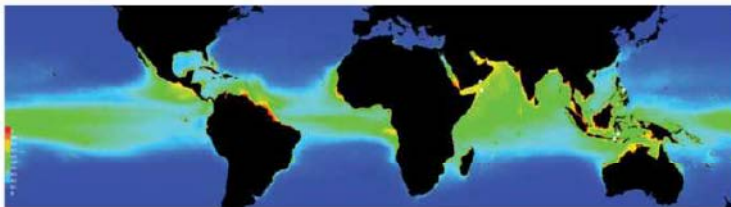
# Training samples = 10    # Test samples = 2    AUC (training) = 0.913    AUC (test) = 0.913    AUC ratio (training) = 1.652    AUC ratio (test) = 1.541

*Halimeda distorta*.2



# Training samples = 16    # Test samples = 3    AUC (training) = 0.853    AUC (test) = 0.880    AUC ratio (training) = 1.423    AUC ratio (test) = 1.854

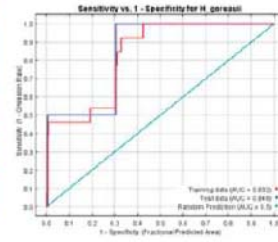
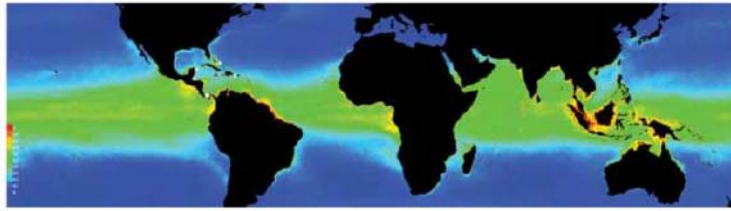
*Halimeda gigas*



# Training samples = 14    # Test samples = 2    AUC (training) = 0.897    AUC (test) = 0.754    AUC ratio (training) = 1.384    AUC ratio (test) = 1.708

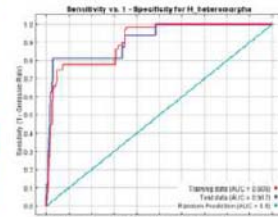
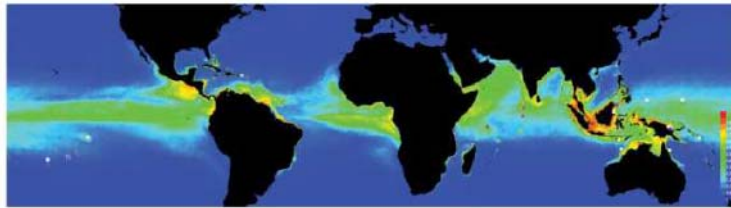


*Halimeda goreauii*



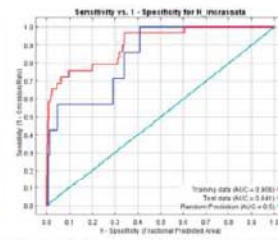
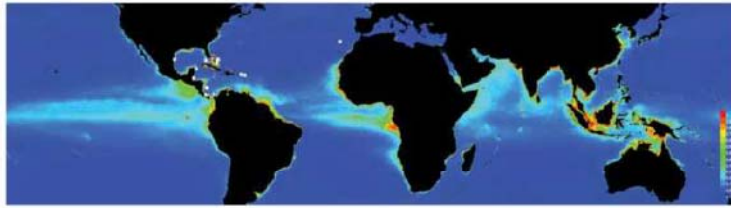
# Training samples = 13    # Test samples = 2    AUC (training) = 0.832    AUC (test) = 0.846    AUC ratio (training) = 1.488    AUC ratio (test) = 1.991

*Halimeda heteromorpha*



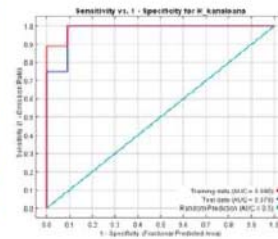
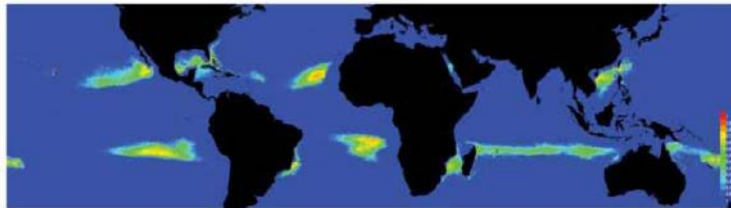
# Training samples = 67    # Test samples = 16    AUC (training) = 0.908    AUC (test) = 0.917    AUC ratio (training) = 1.470    AUC ratio (test) = 1.458

*Halimeda incrassata*



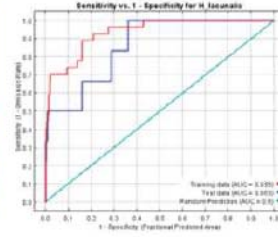
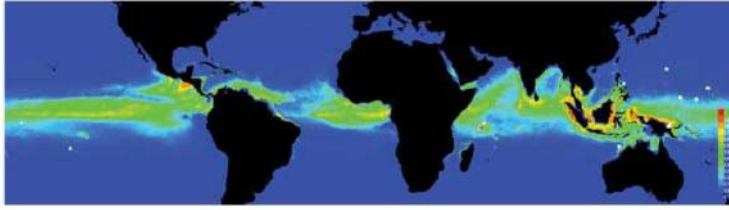
# Training samples = 29    # Test samples = 7    AUC (training) = 0.906    AUC (test) = 0.841    AUC ratio (training) = 1.464    AUC ratio (test) = 1.411

*Halimeda kanoana*



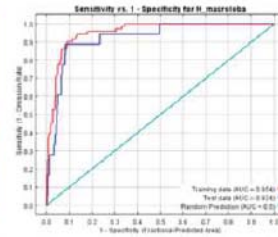
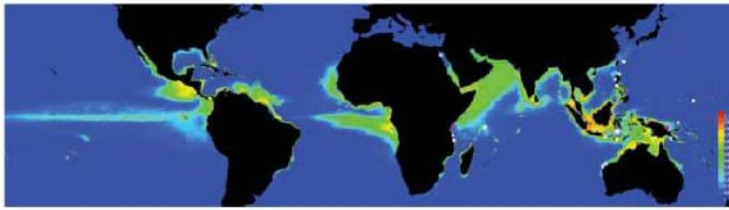
# Training samples = 18    # Test samples = 4    AUC (training) = 0.990    AUC (test) = 0.978    AUC ratio (training) = 0.940    AUC ratio (test) = 0.940

*Halimeda lacunalis*



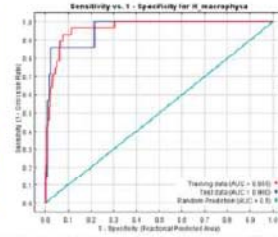
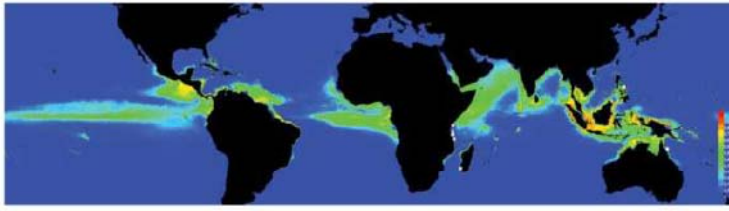
# Training samples = 27   # Test samples = 6   AUC (training) = 0.935   AUC (test) = 0.853   AUC ratio (training) = 1.569   AUC ratio (test) = 1.381

*Halimeda macroloba*



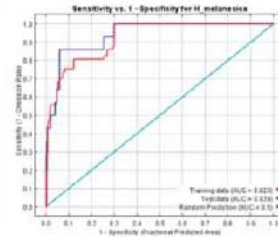
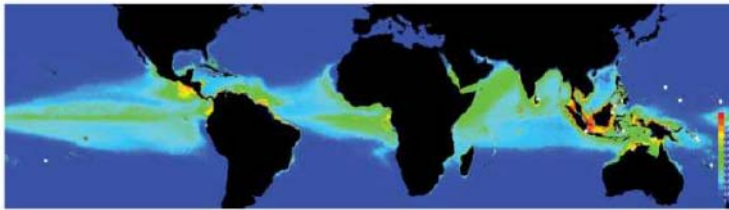
# Training samples = 72   # Test samples = 18   AUC (training) = 0.954   AUC (test) = 0.924   AUC ratio (training) = 1.806   AUC ratio (test) = 1.582

*Halimeda macrophysa*



# Training samples = 28   # Test samples = 7   AUC (training) = 0.965   AUC (test) = 0.950   AUC ratio (training) = 1.837   AUC ratio (test) = 1.640

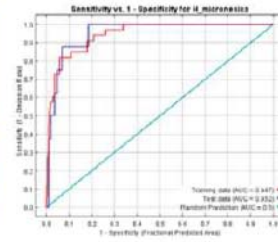
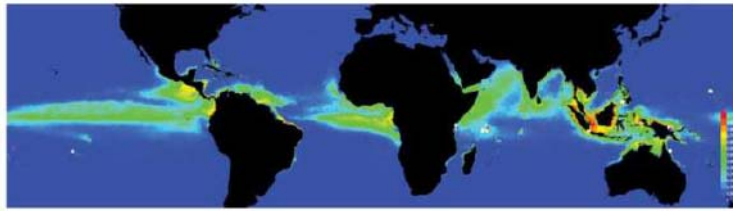
*Halimeda melanesica*



# Training samples = 57   # Test samples = 14   AUC (training) = 0.923   AUC (test) = 0.939   AUC ratio (training) = 1.535   AUC ratio (test) = 1.563

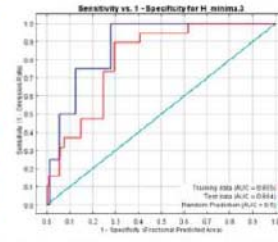
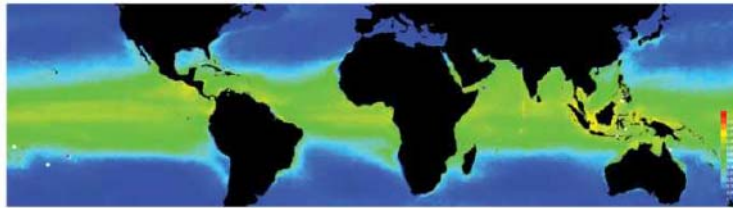


*Halimeda micronesica*



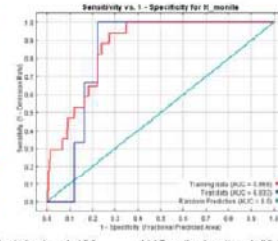
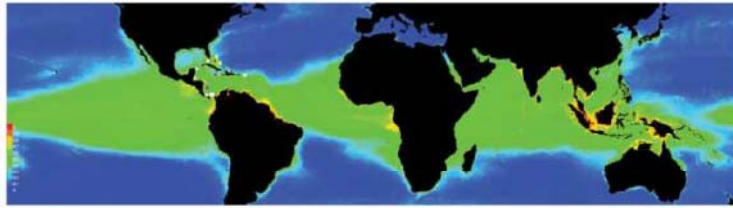
# Training samples = 33 # Test samples = 8 AUC (training) = 0.947 AUC (test) = 0.852 AUC ratio (training) = 1.675 AUC ratio (test) = 1.683

*Halimeda minima.3*



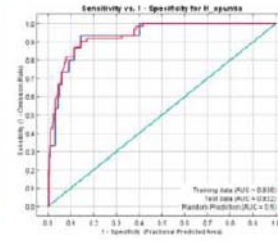
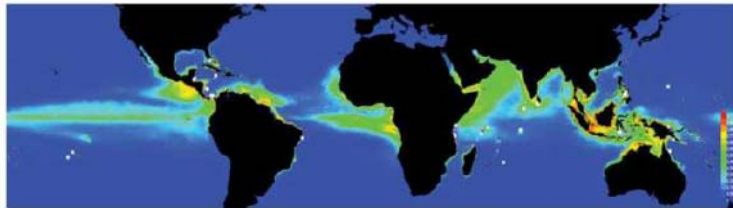
# Training samples = 19 # Test samples = 4 AUC (training) = 0.805 AUC (test) = 0.884 AUC ratio (training) = 1.388 AUC ratio (test) = 1.555

*Halimeda monile*



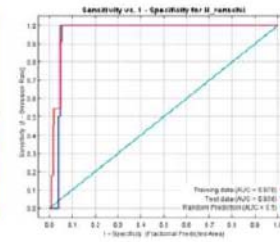
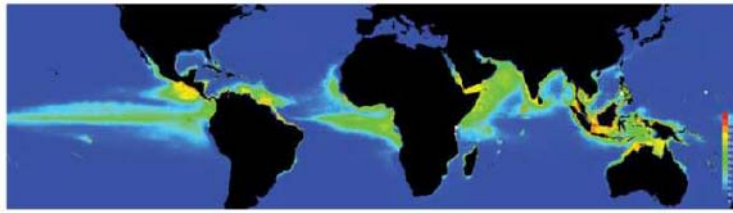
# Training samples = 17 # Test samples = 3 AUC (training) = 0.866 AUC (test) = 0.832 AUC ratio (training) = 1.499 AUC ratio (test) = 1.531

*Halimeda opuntia*



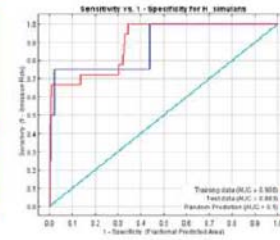
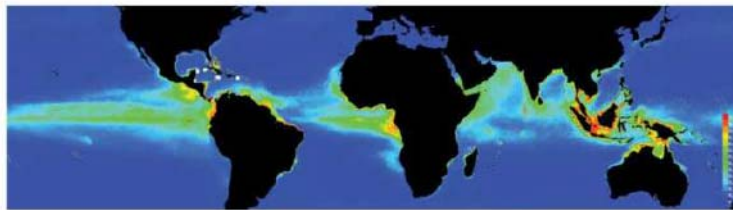
# Training samples = 61 # Test samples = 15 AUC (training) = 0.936 AUC (test) = 0.932 AUC ratio (training) = 1.707 AUC ratio (test) = 1.711

*Halimeda renschii*



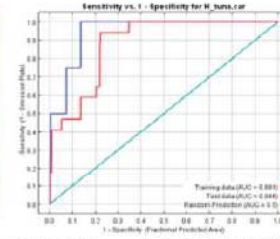
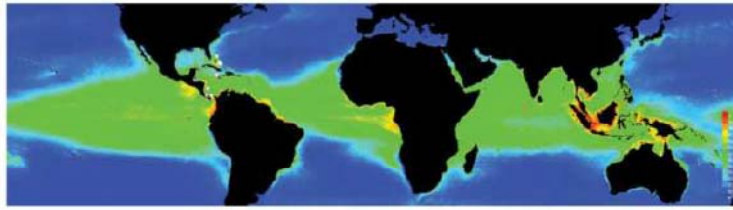
# Training samples = 11 # Test samples = 2 AUC (training) = 0.970 AUC (test) = 0.956 AUC ratio (training) = 1.902 AUC ratio (test) = 1.907

*Halimeda simulans*



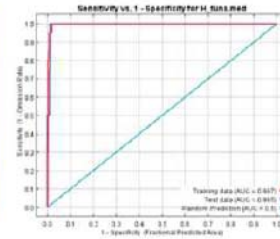
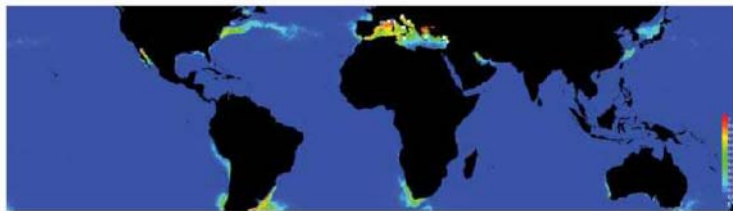
# Training samples = 18 # Test samples = 4 AUC (training) = 0.900 AUC (test) = 0.683 AUC ratio (training) = 1.495 AUC ratio (test) = 1.376

*Halimeda tuna.car*



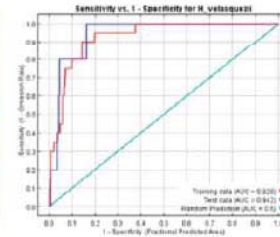
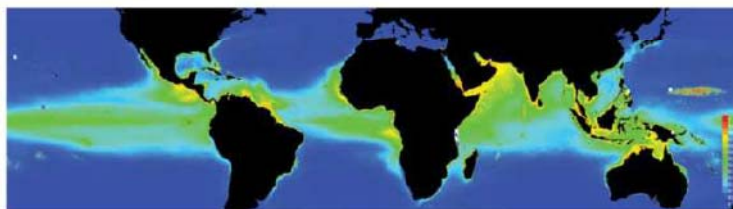
# Training samples = 17 # Test samples = 4 AUC (training) = 0.881 AUC (test) = 0.946 AUC ratio (training) = 1.621 AUC ratio (test) = 1.757

*Halimeda tuna.med*



# Training samples = 20 # Test samples = 4 AUC (training) = 0.997 AUC (test) = 0.995 AUC ratio (training) = 1.984 AUC ratio (test) = 1.977

*Halimeda velasquezii*



# Training samples = 20 # Test samples = 5 AUC (training) = 0.926 AUC (test) = 0.942 AUC ratio (training) = 1.719 AUC ratio (test) = 1.721