Proceedings of the 11th International Coral Reef Symposium, Ft. Lauderdale, Florida, 7-11 July 2008
Session number 26

Biodiversity of reefs: inferring from sparse data

D.G. Fautin

Department of Ecology and Evolutionary Biology, and Natural History Museum, University of Kansas, Lawrence, Kansas 66045 USA

Abstract. Data on occurrences of a particular organism from publications and museum specimens can be used to infer occurrence of members of that species in places where sampling has not been done and at times in the past and future. Programs to make such inferences are based on knowledge of the habitat correlates of the species and determining where else in the world those habitat parameters occur. Precision of such an inference depends on 1) accurate and precise knowledge of the species' habitat requirements, 2) detailed spatially-explicit environmental data, and 3) comprehensive taxonomic and nomenclatural information. Such inferential tools can be important in understanding biogeographic consequences of climate change, in predicting where invasive species might persist, and in recognizing invasive species.

Key words: biogeographic modeling, habitat requirements, taxonomy.

The problem

Species range maps like those in many field guides (eg that to anemonefishes and their hosts by Fautin and Allen 1992, and that to marine life of southern Africa by Branch et al. 2005), are abstractions of the places where members of the species of concern are known to occur. Although the open Indian Ocean between Sumatra and Sri Lanka is depicted by Fautin and Allen (1992) as being within the range of the sea anemone *Heteractis aurora* (Fig. 1), one could not actually find the animal there because, as Fautin and Allen (1992) point out, these animals are confined to shallow water. Thus they occur along land, in strips so narrow they are not easily indicated on small-scale maps – so the range is represented as covering the entire sea bounded by known or inferred occurrences at the extremes. Local guides portray a smaller portion of the globe, and so can indicate ranges in strips along the coast, as Branch et al (2005) do, but the strips are continuous, and it is likely that members of most species occur in only some of the places indicated, occurrences dictated by the presence of appropriate habitat. Such abstractions over-represent the range of a species.

Maps indicating actual occurrences are typical in technical publications, such as Wallace (1999), and in the growing number of biodiversity websites such as that to sea anemones and relatives (Fautin 2008: eg Fig. 2). With few exceptions, it is likely that members of a species occur in more than the places indicated on the maps, which are where they have been observed or from which they have been collected. These concrete depictions under-represent the range of a species.

A solution

As has been known for many years (Allee et al. 1949), to a first approximation, an organism's range can be inferred from abstracting its habitat characteristics based on places the organism is known to occur, then assuming the organism occurs where its habitat does. Computerized methods make the four-step process of developing such range maps theoretically straight-forward. 1) Occurrence records of a species and 2) maps depicting the distribution of relevant environmental parameters are overlain. 3) An algorithm abstracts the values of the environmental parameters that coincide with occurrence records, then 4) maps the occurrence of the environmental parameters within the values known to be compatible with the existence of the organism.

The distribution of these values of environmental parameters is a first-cut inference of the occurrence of habitat suitable for the species and thus the potential range of the species. Modeled environmental data for the past and future allow inferences to be made about historical biogeography and how ranges may shift with climate change.

Many models have been developed to carry out such computations, differing in assumptions and underlying algorithms. Some are referred to as "niche models" but few actually consider biological features other than by proxy with habitat. Sixteen such modeling programs were tested against one another

Figure 1: Shaded area is range of the sea anemone *Heteractis aurora* (Quoy and Gaimard, 1833) as given by Fautin and Allen (1992).

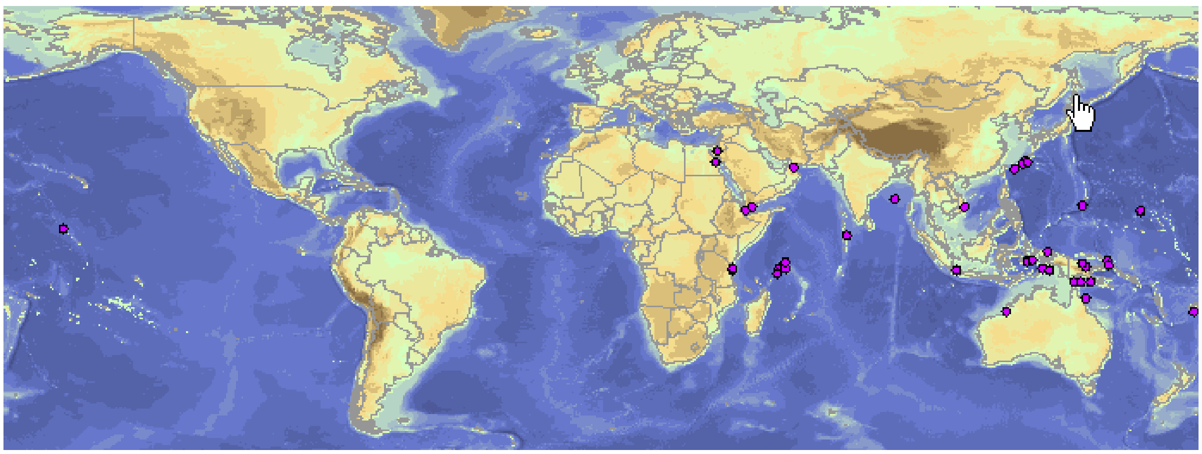

Figure 2: Published occurrences of the sea anemone *Heteractis aurora* (Quoy and Gaimard, 1833) as shown in Fautin (2008).

(Elith et al. 2006), but all were for terrestrial species. One of those programs had been used to model the distribution of some marine fishes, but did poorly for a variety of reasons (Wiley et a.l 2003).

Guinotte et al (2006) and Fautin and Buddemeier (2008) demonstrated the use of KGSMapper, a modeling program that handles marine data well. KGSMapper is associated with 42 spatially-explicit environmental parameters relevant to marine species. Grid size of the associated environmental data is half a degree -- coarse for some purposes – but the coverage is worldwide, which makes it appropriate for many uses. Further, KGSMapper is scale-independent, as demonstrated at "Scleractinian Corals and other Hexacorallians of the Northwestern Hawai'ian Islands" (http://hercules.kgs.ku.edu/ hexacoral/hawaii/biodata) in which it is associated with finer-scale environmental data.

Some considerations

Presumably, precision of the environmental data and knowledge of a species' habitat requirements correlate directly with reliability of a range inference. Accuracy of the inference of a species' range should also increase with improved knowledge of the known distribution of the species. One way to maximize information of a species' occurrence is to include all records for the species regardless of the scientific name that was used for it.

Although ideally a single scientific name refers to a single species, many species have been given more than one name (such names are synonyms; International Commission of Zoological Nomenclature 1999). The sea anemone species for which the valid name is *Heteractis aurora* (Quoy and Gaimard, 1833) has been referred to in publication by 11 names (Fautin 2008): it was described as a new species three times, has been placed in nine genera, and has been misidentified.

The points on distribution maps in "Hexacorallians of the World" (Fautin 2008), which depict occurrence records from the published literature, are color-coded by the name used for the species. Because synonyms are a matter of taxonomy – they depend on a scientist's opinion, and scientists may legitimately differ on whether two names (or more) refer to a single species – this convention allows a user to distinguish among the records by name, disregarding those that the user considers do not refer to the species in question. In fact, such a depiction can be informative in deciding whether a particular synonymy is justified.

Figure 3 is the KGSMapper output of suitable habitat for *Heteractis aurora* based on 75 occurrence records and the environmental parameters mean depth, mean and minimum surface seawater temperature, and maximum and minimum salinity. These parameters were selected because they are relevant to the biology of these anemones. The animals are zooxanthellate, so occur only in shallow water, and are confined to warm, fully saline seas. They are also known only from the Indo-Pacific. Within that natural range, the model output depicts the animal's potential range; this can be used to infer occurrence of members of the species in places where sampling has not been done, which can be useful, for example, in planning fieldwork, or inferring areas into which the animals might expand as climate changes. Outside that natural range (in the Caribbean, for example), the model output can be used to infer where the species might unnaturally occur, as, for example, where it might invade and persist. On the website "Hexacorallians of the World," the darker the reddish color, the more likely the habitat is to be suitable (see Guinotte et al. 2006).

Figure 4 depicts the habitat suitable for *Heteractis aurora* based on the same environmental parameters as for Figure 3, but uses only those occurrence records that referred to the species by one of its synonyms, *Radianthus koseirensis*. Presumably both the smaller number of records and the restriction of them to the western portion of the species' range leads to the conclusion that the species is likely to be

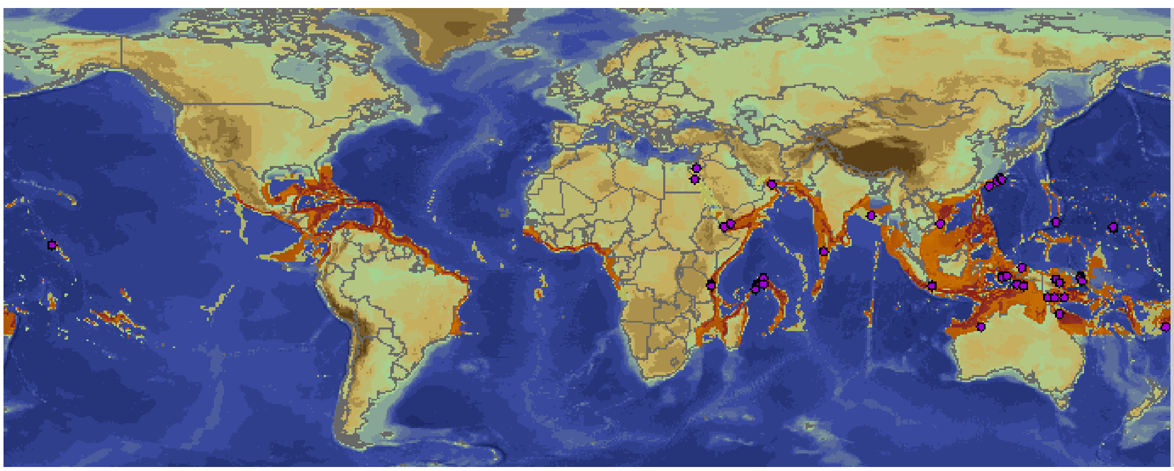

Figure 3: Suitable habitat for *Heteractis aurora* (Quoy and Gaimard 1833) inferred by KGSMapper from 75 published occurrence records (closed circles), and the environmental parameters mean depth, mean and minimum surface seawater temperature, and maximum and minimum monthly salinity. The darker (the reddish color), the more likely the habitat is to be suitable (Guinotte et al. 2006).

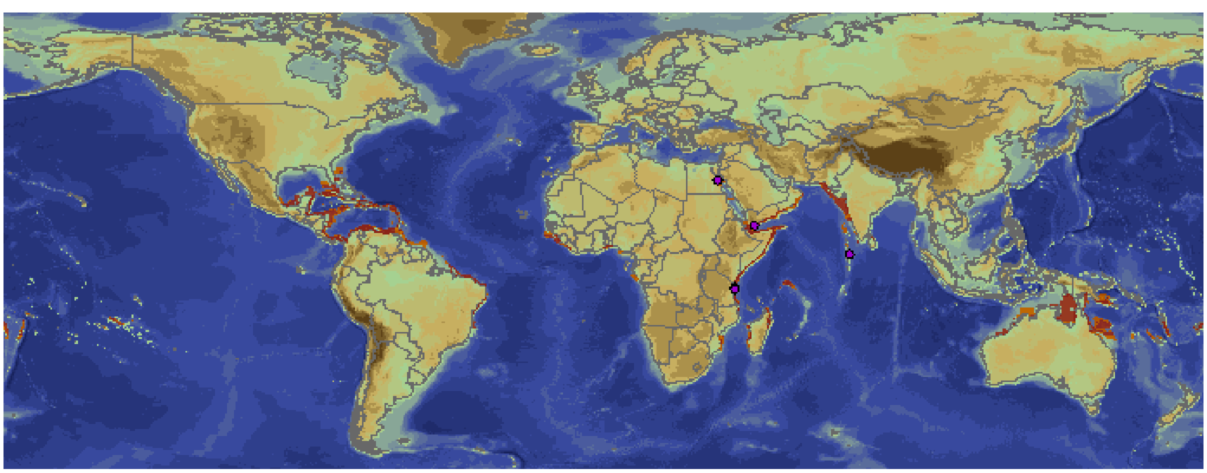

Figure 4: Suitable habitat for *Heteractis aurora* (Quoy and Gaimard 1833) inferred by KGSMapper only from published occurrence records (closed circles) that used the synonymous name *Radianthus koseirensis*. The darker (the reddish color), the more likely the habitat is to be suitable (Guinotte et al. 2006).

far less widespread than is seen in Figure 3. When rendered in black and white, the difference in model output is easiest to perceive in Southeast Asia, where the shaded area of inferred occurrence is more widespread and darker in Figure 3 than Figure 4.

Plotting occurrences using a species name rather than a species concept (by including synonyms) can be not only incomplete but may be misleading. That is because more than one species may have been given the same scientific name (such names are termed homonyms; International Commission of Zoological Nomenclature 1999). Homonyms are not a matter of opinion, but are objectively verifiable, and thus fall within the province of nomenclature. The anemone species to which the name *Heteractis aurora* properly belongs occurs strictly in tropical waters of the Indo-Pacific. That species was first placed in the genus *Actinia* (to which nearly all sea anemones were initially assigned). What is clearly a different species in Britain was also given the name *Actinia aurora*. Applying a model to infer distribution from records that include occurrences of the British species would obviously give a meaningless result.

Conclusions

Use of modeling programs to infer the range of a species from known occurrences and features of the environment benefits from

- accurate and precise knowledge of the species' habitat and natural history (to select environmental parameters relevant to the organism's life that control its distribution)
- knowledge of biogeography (where the organism would be expected to occur, so invasions can be distinguished from occurrences in places where the species might be expected but has not previously been recorded)
- consideration of taxonomy and nomenclature (what other names have been applied to it – and which applications of a name do not refer to the species in question)

Acknowledgments

With gratitude for grants from US National Science Foundation (DEB 99-78106, OCE 00-03970, EF-0531779), the US Geological Survey (in the National Biological Information Infrastructure -- NBII), and the Alfred P. Sloan Foundation, and to individuals too numerous to mention other than Adorian Ardelean, Jeremy Bartley, Asif Iqbal, and Suman Kansakar.

References

Allee WC, Emerson AE, Park O, Park T, Schmidt KP (1949) Principles of animal ecology. WB Saunders Company, p 837

Branch GM, Griffiths CL, Branch ML, Beckley LE (2005) Two oceans: a guide to the marine life of southern Africa, 2nd edition. David Philip, p 359

Elith J, Graham CH, Anderson R P, Dudík 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 J McC, Peterson AT, Phillips SJ, Richardson KS, Scachetti-Pereira R, Schapire RE, Soberón J., Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151

Fautin DG (2008) Hexacorallians of the World. http://geoportal.kgs.ku.edu/hexacoral/anemone2/index.cfm

Fautin DG, Allen GR (1992) Field guide to anemonefishes and their host sea anemones. Western Australian Museum, p 160

Fautin DG, Buddemeier RW (2008) Inferring distributions of organisms from sparse occurrence data: KGSMapper. Globec International Newsletter 14(1):48-49

Guinotte JM, Bartley JD, Iqbal A, Fautin DG, Buddemeier RW (2006) Modeling habitat distribution from organism occurrences and environmental data: a case study using anemonefishes and their sea anemone hosts. Marine Ecology Progress Series 316: 269-283

International Commission on Zoological Nomenclature (1999) International code of zoological nomenclature, 4th edition. International Trust for Zoological Nomenclature, p 289

Quoy JRC, Gaimard P (1833) Voyage de découvertes de l'Astrolabe exécuté par ordre du roi, pendant les années 1826-1827-1828-1829, sous le commandement de M. J. Dumont d'Urville, vol 4 (zoophytes). J Tastu, p 390

Wallace CC (1999) Staghorn corals of the world: a revision of the coral genus *Acropora*. CSIRO, Collingwood, Australia, p 421

Wiley EO, McNyset KM, Peterson AT, Robins CR, Stewart AM (2003) Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography 16:120-127