ECOLOGICAL MODELLING 203 (2007) 375–386

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel

Analysis of environmental factors influencing the range of anopheline mosquitoes in northern Australia using a genetic algorithm and data mining methods

A.W. Sweeney[a,*], N.W. Beebe[a,**], R.D. Cooper[b]

[a] Institute for the Biotechnology of Infectious Diseases, University of Technology Sydney, PO Box 123, Broadway, NSW 2007, Australia
[b] Army Malaria Institute, Gallipoli Barracks, Enoggera, Qld 4052, Australia

ARTICLE INFO

Article history:
Received 20 September 2005
Received in revised form
2 December 2006
Accepted 11 December 2006
Published on line 16 January 2007

Keywords:
Ecological niche modelling
Genetic algorithms
Data mining
Mosquitoes
GARP
CART
KnowledgeSeeker

ABSTRACT

Environmental factors which influence the distributions of malaria vectors in northern Australia (*Anopheles farauti* ss, *A. farauti* 2 and *A. farauti* 3) were investigated by ecological niche modelling and data mining using an extensive data set of species presence and absence records obtained by systematic field surveys. Models were generated with GARP (the genetic algorithm for rule-set prediction) using geographical coverages of 41 climatic and topographic parameters for the north of the continent. Environmental variables associated with species records were identified with the ranking procedures of the decision tree software packages CART and KnowledgeSeeker. There was consistent agreement in the variables ranked by both methods. This permitted the selection of reduced sets of environmental variables to develop GARP models for the three target species with equivalent predictive accuracy to those which used all of the environmental information. The environmental parameters which define the realised distributions of *A. farauti* ss and *A. farauti* 3 were well described by this approach but the results were less satisfactory for *A. farauti* 2. Atmospheric moisture was shown to be a critical variable for each species which accords with many field and laboratory observations concerning the influence of humidity on adult mosquito survival.

1. Introduction

Members of the *Anopheles punctulatus* group are important malaria vectors in the southwest Pacific region. Understanding their geographical ranges is of considerable interest from both ecological and disease transmission viewpoints. Comprehensive surveys of the Australian species of the group, *Anopheles farauti* ss, *A. farauti* 2 and *A. farauti* 3, were undertaken in the formerly malarious areas of northern Australia between 1985 and 1994. The surveys, which included material from over 600 localities, yielded more than 300 locality records of the target species to provide a realistic indication of their realised distribution in the Northern Territory and Queensland. The results showed that these mosquitoes were distributed around the coast and 50–100 km inland north of 20°S latitude and east of 129°E longitude. In some areas the three species were found together but the overall patterns of occurrences for each species were different (Sweeney et al., 1990; Cooper et al., 1995, 1996). These apparent dissimilarities in realised distribution suggest that there may be differences in the ecological factors which influence the range of the individual species.

* *Corresponding author*. Tel.: +61 2 4385 8774; fax: +61 2 9514 4003.
** *Corresponding author*.
E-mail address: tony.sweeney@uts.edu.au (A.W. Sweeney).
0304-3800/$ – see front matter
doi:10.1016/j.ecolmodel.2006.12.003

Ecological niche modelling with GARP (the genetic algorithm for rule-set prediction) has proved to be particular useful for investigating species distributions by creating models, using point localities where species are known to occur and environmental data for the geographic region of interest, to predict species range (Stockwell and Peters, 1999). A comparative study of GARP predicted ranges for the five sibling species of *Anopheles quadrimaculatus* implied that *A. quadrimaculatus* ss was the only member of the group capable of transmitting malaria throughout the formerly malarious area of the United States (Levine et al., 2004a). Similarly, Levine et al. (2004b) used GARP to predict the potential range of the *Anopheles gambiae* complex in Africa. Models developed from African distribution data and projected to South America suggested that *A. gambiae* ss was the species introduced into Brazil in 1929.

For the present study we used GARP for modelling the potential distribution of the *A. farauti* sibling species in Australia. Inputs for model construction included species occurrence data from field surveys together with high resolution environmental information for northern Australia based on historical climate records. There are a number of procedures other than GARP which have been used for predictive modelling of species distributions including regression algorithms (such as generalised linear models and locally weighted regression), classification or decision tree analysis, environmental envelopes (including BIOCLIM which is part of the GARP algorithm), neural networks, and Bayesian statistics (reviewed by Guisan and Zimmerman (2000)). In addition to ecological modelling some of these analytical procedures can be applied to data mining, the process of statistical analysis to reveal previously unknown patterns from a set of data values. We selected two decision tree software packages, CART (classification and regression tree analysis, Breiman et al., 1984) and KnowledgeSeeker (Biggs et al., 1991), to search for significant environmental factors associated with species presence. GARP models were generated with reduced sets of environmental layers highlighted by these data mining techniques to determine whether the model outputs corresponded with high quality range predictions. The overall objective was to identify the key environmental factors responsible for defining the geographical ranges of the different vector species as such factors are of major epidemiological significance and of direct relevance for malaria control strategies.

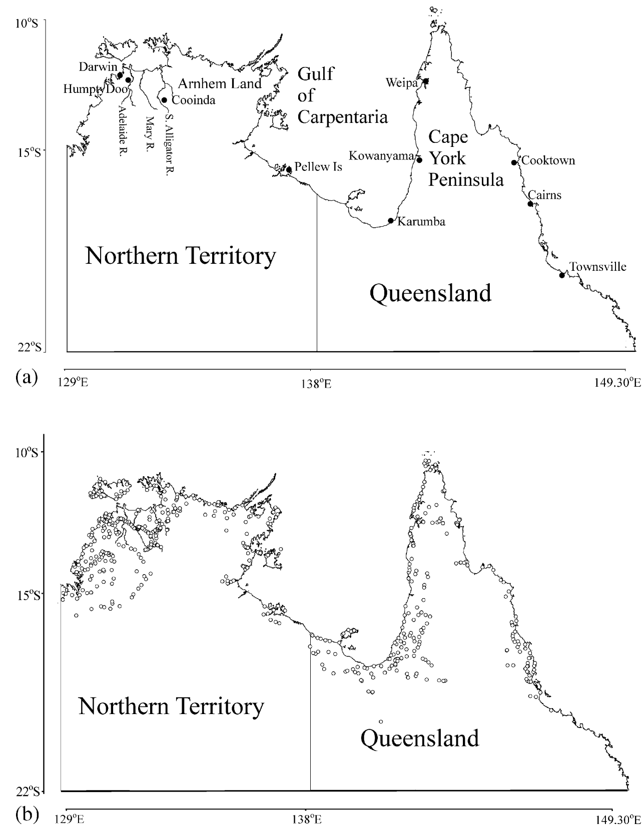

Fig. 1 – Survey area of northern Australia. (a) Place names and localities mentioned in text. (b) Localities of 619 collection sites (designated ○) at which anopheline mosquitoes were collected during field surveys. These included both record and no-record sites for *A. farauti* sl.

2. Materials and methods

2.1. *Mosquito surveys*

The area surveyed included the coast of Queensland and the Northern Territory and up to 300 km inland between latitudes 10–19°S and longitudes 128–146°E (Fig. 1a). A different sector of this region was covered progressively each year between 1985 and 1991: the Queensland coast from Townsville to Cooktown in 1985, Cape York Peninsula in 1986, the southern coast of the Gulf of Carpentaria in 1987 and the Northern Territory in 1988, 1989, and 1990. The 1991 survey included the Torres Strait Islands and areas of far north Queensland not adequately covered in previous surveys. In 1994 a final survey was made in the Northern Territory to provide additional material in the floodplains of the Adelaide and Mary Rivers. Each yearly survey was made for a month during March–May which normally coincides with the end of the north Australian wet season when larval sites are expected to be plentiful and adult densities are usually high. In order to adequately sample the major land cover classes in this thinly populated part of the world the surveys relied on Australian Army helicopters to access remote areas inaccessible by roads and tracks. Adult mosquitoes were caught overnight in battery-powered CO_2 baited light traps (Rohe and Fall, 1979). Each locality was plotted on 1:100,000 Australian Army Topographical maps which show grid lines at 1000 m intervals. The latitude and longitude of individual collection sites were plotted to a tenth of a grid square resulting in a spatial accuracy to the nearest 100 m. Specimens were morphologically identified to species level in the field and the collections from each locality were then stored separately in liquid nitrogen for later identification of sibling species in the laboratory using either allozyyme electrophoresis (Mahon, 1984) or DNA probes (Cooper et al., 1991). Anopheline larvae, collected from a wide range of water bodies, were reared in a field laboratory to the adult stage and then processed in the same way as adult specimens. The survey data set comprises collection records at 619 localities at which members of the *A. farauti* group were either found or not found (Fig. 1b). There were over 26,000 anopheline mosquitoes