The value of georeferenced collection records for predicting patterns of mosquito species richness and endemism in the Neotropics

DESMOND H. FOLEY[1], ANNA L. WEITZMAN[2], SCOTT E. MILLER[2], MICHAEL E. FARAN[3], LEOPOLDO M. RUEDA[1] and RICHARD C. WILKERSON[1]   [1]Division of Entomology, Walter Reed Army Institute of Research, Silver Spring, Maryland, U.S.A., [2]National Museum of Natural History, Smithsonian Institution, Washington DC, U.S.A., [3]Department of Psychiatry, Tripler Army Medical Center, Honolulu, Hawaii, U.S.A.

Abstract. 1. Determining large-scale distribution patterns for mosquitoes could advance knowledge of global mosquito biogeography and inform decisions about where mosquito inventory needs are greatest.

2. Over 43 000 georeferenced records are presented of identified and vouchered mosquitoes from collections undertaken between 1899 and 1982, from 1853 locations in 42 countries throughout the Neotropics. Of 492 species in the data set, 23% were only recorded from one location, and *Anopheles albimanus* Wiedemann is the most common species.

3. A linear log–log species–area relationship was found for mosquito species number and country area. Chile had the lowest relative density of species and Trinidad-Tobago the highest, followed by Panama and French Guiana.

4. The potential distribution of species was predicted using an Ecological Niche Modelling (ENM) approach. *Anopheles* species had the largest predicted species ranges, whereas species of *Deinocerites* and *Wyeomyia* had the smallest.

5. Species richness was estimated for 1° grids and by summing predicted presence of species from ENM. These methods both showed areas of high species richness in French Guiana, Panama, Trinidad-Tobago, and Colombia. Potential hotspots in endemicity included unsampled areas in Panama, French Guiana, Colombia, Belize, Venezuela, and Brazil.

6. Argentina, The Bahamas, Bermuda, Bolivia, Cuba, and Peru were the most underrepresented countries in the database compared with known country species occurrence data. Analysis of species accumulation curves suggested patchiness in the distribution of data points, which may affect estimates of species richness.

7. The data set is a first step towards the development of a global-scale repository of georeferenced mosquito collection records.

Key words. BIOCLIM, biogeography, collections, database, distribution, mosquito, Neotropics, species–area relationship, species endemism, species richness.

Introduction

With the advent of powerful tools enabling the use of museum specimen data for addressing questions in global biogeography, climate change, vector control, and conservation, there is an

Correspondence: Desmond H. Foley, Smithsonian Institution MSC MRC534, 4210 Silver Hill Road, Suitland MD 20746, U.S.A. E-mail: foleydes@si.edu

increasing need for continental and global scale databases of insect specimen data (Turner *et al.*, 2003; Elith *et al.*, 2006). Although, continental scale databases have been compiled for butterflies (e.g. Soberon *et al.*, 2000; Kerr *et al.*, 2001), insects lag far behind plants and vertebrates in the availability of such data.

For mosquitoes, georeferenced collections databases of large geographic scale exist but are usually not digitised or comprise unpublished survey and museum collection records. The paucity

of detailed data on the past and present distribution of vectors is a major limiting factor for global modelling of vector-borne diseases (Rogers & Randolph, 2003; Tatem *et al.*, 2006). There is also a challenge of choosing appropriate priority taxa that can be used to represent biodiversity in monitoring projects (Pereira & Cooper, 2006), and mosquitoes have been suggested as a priority group within the insects (e.g. Raven, 1980).

Country inventories or checklists of mosquito species are available, such as the Walter Reed Biosystematics Unit's (WRBU) Systematic Catalog of the Culicidae (SCC) (http://www.mosquitocatalog.org/main.asp). Recently, Geographical Information Systems (GIS) software was used to display these data as global maps to investigate geographic patterning of mosquito species numbers (Foley *et al.*, 2007). These authors described species–area relationships and a latitudinal diversity gradient for mosquito species. Higher resolution mosquito distribution data on a global scale, however, are needed.

An extensive database of georeferenced mosquito collection records for the Neotropical area is reported. The Neotropics have been proposed as the centre of origin of mosquitoes (e.g. Qu & Qian, 1989), including the Anophelinae (Harbach, 1998; Krzywinski *et al.*, 2001). A long history of anophelines in the New World is suggested from the finding of *Anopheles (Nyssorhynchus)* in Dominican amber 15–45 million years old (Zavortink & Poinar, 2000). Belkin (1962) regarded the Caribbean islands as a cradle of mosquito evolution; being climatically stable, geologically unstable and especially affected by changing sea levels. The long isolation of the New World, beginning with the separation of South America and Africa about 100 million years ago, appears to have encouraged the evolution of a unique mosquito fauna. For example, four subgenera of *Anopheles* (*Kerteszia*, *Lophopodomyia*, *Nyssorhynchus*, and *Stethomyia*) are confined to the Neotropics. Foley *et al.* (2007) showed that Brazil is a centre of diversity for mosquito species and French Guiana has one of the highest relative species densities anywhere in the world. These authors also showed that some Neotropical countries have been a focus for mosquito systematics studies.

The database reported here began with the Mosquito Information Management Project (MIMP), initiated in 1979 to develop a computer-based system for storing and retrieving data on mosquitoes (Faran *et al.*, 1984). MIMP was tasked with digitising the records of paper collection forms for vouchered specimens housed at the National History Museum, Smithsonian Institution. Notable among these sources was the Mosquitoes of Middle America (MOMA) collection from the University of California, Los Angeles. The MOMA records are a testament to the foresight and taxonomic skills of John Belkin, who saw the importance to future mosquito researchers of accessible collection details in a standardised format (Belkin & Heinemann, 1973; Zavortink, 1990). Belkin & Heinemann (1973, 1975a,b, 1976a,b,c), Heinemann & Belkin (1977a,b,c, 1978a,b,c, 1979), Heinemann *et al.* (1980), and Heinemann (1980) contain records for the MOMA collections, some of which date back to 1899. In the early 1980s, information pertaining to about 402000 specimens on over 15 500 paper collection forms, including details on the identification, collection location, and ecological characteristics of the collection site, were entered into a computer (Faran *et al.*, 1984) utilising the SELGEM software

(Creighton & Crockett, 1971). Although many paper records remained to be digitised, the MIMP project terminated in 1983, after which these digital records languished. In 1999 the Smithsonian Institution decided to convert all its electronic specimen data, including the mosquito database, into a single database. On 10 September 2001, the only digital copies of these records arrived at a company a couple of blocks from the World Trade Center in New York, where they were to be converted from magnetic tape to a modern digital format. Despite their proximity to *ground zero*, these records survived the terrorist attacks the following day, and were returned to the Smithsonian to be converted to a modern object-relational hybrid data structure. An analysis of a subset of this database is presented here to demonstrate the value of digitised georeferenced collection data, in this case, for questions about mosquito biogeography and survey history.

Materials and methods

Faran *et al.* (1984) describe the composition of the 78 categories and subcategories of information within the original MIMP database, which included locality description, collection code, collection date, latitude and longitude, species identification, a unique record identification number, collector, and date of collection. The database was divided into those records that had geographic coordinate data (i.e. degrees longitude and latitude) and those that did not. Entries were further divided into those of questionable taxonomy (i.e. species entries followed by group, complex or aff. = affinity), apparent identification failures (i.e. entries where species identification was not given or was followed by *uncertain* or ?), and those with unequivocal species identification. Those without geographic coordinates were divided into those that had Military Grid Reference System (MGRS) coordinates and those that did not. MGRS coordinates were converted to geographic coordinates for WGS-84 using the batch options in GEOTRANS V2.2.6 (US Army Topographic Engineering Center, Geospatial Information Division). The appropriate horizontal datum and ellipsoid were determined by inspection of maps housed at the WRBU that were originally used to arrive at MGRS coordinates. Where MGRS data entry errors were suspected, such as transposed letters and digits and incomplete coordinates, the Universal Transverse Mercator (UTM) zone number and designator were first confirmed by cross-checking against a world map of UTM grids. Error detection at this stage was helped by use of the electronic gazetteer EGAZ, and BIOLINK Map Assistant V2.1.309 (Shattuck, 1997), which located collection site names on a map. Many MGRS readings were to 1 km precision but for a number of Caribbean islands these could be increased to 100 m by re-georeferencing collection locations where these points were obvious on original maps. When geographic coordinates were already present in the database in degrees–minutes format, these were converted to decimal degrees and checked to ensure they had the correct sign (+ or −) for their hemisphere of origin.

Specimens with unequivocal identifications and geocodes were filtered in Microsoft Excel for unique locations, and these point data were converted to shape files for mapping in DIVA-GIS 5.3 (http://www.diva-gis.org/). Further data cleaning was undertaken

by the *check coordinates* option of DIVA-GIS, a *point-in-polygon* method (Chapman, 2005a), which identifies points located outside all polygons (i.e. fell in the ocean), and points that did not match relations for the country names (i.e. fell in another country). Locations so identified ($n = 273$) were rechecked and corrected by consulting original collection cards and maps housed at the WRBU or through the Alexandria Digital Library online Gazetteer (http://middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp). Data were imported into ARCVIEW GIS 3.3 for graphical display. Generic, subgeneric, and species names were updated to follow the SCC on the WRBU website (http://www.mosquitocatalog.org/main.asp, accessed 23 May 2006). Mosquito species composition by country was obtained from the WRBU website (accessed 16 June 2006). Land area of countries was determined via the Central Intelligence Agency (CIA) factbook website (http://www.cia.gov/cia/publications/factbook/geos/mh.html, accessed 19 June 2006). Species number and country area data were $\log_{10}$ transformed and linear regression performed using MINITAB version 14.20 (Minitab, State College, PA, U.S.A.).

The MIMP database is composed of mosquito collections, some of which were undertaken from the same location at different dates. Each collection records one or more individuals of the same species, but the number of specimens during a single collection was not meant to be an estimate of the abundance of that species at that location. It was decided to act conservatively by reducing the cleaned subset of the MIMP database to record only the presence/absence (occurrence or incidence) of each species in each collection.

Distribution modelling

The potential distribution of species was predicted using the BIOCLIM algorithm (Nix, 1986), an Ecological Niche Modelling (ENM) approach based on climate matching, in DIVA-GIS. BIOCLIM attempts to identify suitable and unsuitable areas or *niches* in which the organism is likely to occur, or could survive if it was introduced, based on the climatic features of the data point locations. The BIOCLIM model was implemented using the WORLDCLIM 2.5 arc minute resolution database of 19 bioclimatic variables, that is, annual mean temperature, mean monthly temperature range, isothermality, temperature seasonality, maximum temperature of the warmest month, minimum temperature of the coldest month, annual temperature range, mean temperature of the wettest quarter, mean temperature of the driest quarter, mean temperature of the coldest quarter, mean temperature of the warmest quarter, annual precipitation, wettest month precipitation, driest month precipitation, precipitation seasonality, wettest quarter mean precipitation, driest quarter mean precipitation, coldest quarter mean precipitation, warmest quarter mean precipitation. All areas that are within the climate envelope described by the data, cut off beyond the 0.025 percentile, were mapped as true (1) or false (0). The batch option was used to run BIOCLIM for each species.

Species richness

The number of different species (i.e. species richness) and the number of observations was calculated from point data and displayed as maps in DIVA-GIS for 1° grid cells. Potential species richness was also estimated by summing the incidence (presence/absence) of each species for each grid cell as determined by the BIOCLIM distribution model. This methodology is similar to that of Jarvis *et al.* (2003), who used principal components analysis applied to climate variables.

Endemism

A weighted endemism algorithm based on the BIOCLIM potential species distribution maps was used. This is a variant of the methodology used in the Biodiversity Analysis Tool (BAT) (http://www.deh.gov.au/biodiversity/abif/bat/index.html, accessed 19 June 2006). Whereas in BAT the range size of each species is estimated as the number of grid cells occupied by records for that species in the data, the present study used predicted distribution data from the output of ENM. First, the range size of each species was estimated from the number of positive 2.5 min grid cells from the distribution modelling. An endemism score for each species grid file was calculated in DIVA-GIS as the inverse of the range size multiplied by 100. Grid files for species so calculated were then summed. Species whose predicted distribution was greater than 100 000 grid cells were not included in the calculation because they would add only a negligible amount to the final score. Also, species with predicted distribution of one grid cell or less were excluded, as it could not be demonstrated that the modelling in these cases yielded a valid prediction. For ease of visualising high value grid cells, the summed endemism grid values were recalculated according to the maximum values of neighbouring grid cells (9×9) with the neighbourhood function of DIVA-GIS.

Species accumulation curves

A square was defined, in DIVA-GIS, with lower left and upper right corners at longitude/latitude: $-118°/-40°$ and $-31°/35°$, respectively. The projection was equal-area cylindrical, ensuring that grid cells are comparable with each other. This area represented 87 rows $\times$ 75 columns of $1° \times 1°$ grid cells (total = 6525 cells). One degree of either latitude or longitude at the equator is approximately 111 km. Of the 6525 grid cells in the area, 27% (1830) occurred over land and 325 of these (18%) had at least one species observation.

The program ESTIMATES (version 7.5.1, Colwell, 2005) was used to investigate under-sampling and spatial aggregation in the data. ESTIMATES calculates randomised species accumulation curves (also known as sample-based rarefaction curves) and computes a variety of species richness indicators. For an idealised complete inventory for an area the species accumulation curve climbs asymptotically to the true species richness and taxa that are *rare* have been observed more than once. The expected richness function in ESTIMATES is called *Sobs (Mao Tau)*. A Coleman curve is calculated by randomly reassigning specimens to samples and then recalculating the species accumulation curve, thus removing any clumping in the data. The present study used the incidence-based *Chao2* estimator and

incidence-based coverage estimator *ICE*, which depends on the presence and distribution of rare taxa, to estimate the lower bounds of species richness and to assess the degree of under-sampling. The present study used the default values in ESTIMATES, that is, 50 randomisations for estimators and 10 for the upper abundance limit for rare taxa. Input files for ESTIMATES were obtained by using the point-to-grid function for species richness in DIVA-GIS, which produces a grid file of the presence (1) or absence (0) of each species in each grid cell. The parameter field was set for all species, thereby generating 492 separate grid files. These grid files were stacked together in DIVA and exported as a matrix in text file format. Only cells that had at least one observation were used ($n = 325$) and these records were reformatted for input into ESTIMATES.

Results

After data cleaning the original 87 637 specimen entries, 43 738 records, comprising 7069 collections, were suitable for inclusion in this analysis. Most (61.3%) records were for specimens collected in the 1960s, followed by the 1970s (34.0%), and the most recent record was for 1982. Only 3.7% of records were from before the 1960s, with the oldest being for *Anopheles aquasalis* Curry by F. Urich in Trinidad in 1899. Not included are 7155 records from Egypt, Israel and Kenya, 27 692 of uncertain taxonomy, 153 miscellaneous, and 8899 records that were not georeferenced. The latter category was mostly from Panama (2864) followed by Brazil (1627), Colombia (937), Ecuador (745), Mexico (429), Nicaragua (414), Trinidad-Tobago (361), and Venezuela (238). When repeat specimen collections (i.e.

multiple records of a species during a single collection) of the cleaned data were combined, this resulted in 12 505 unique species-collection records. When multiple collections at the same location on different dates were combined, this resulted in 6773 unique species-location records. These records comprised 1853 locations from 42 countries in the Neotropical region (Fig. 1). Points that originally fell outside country polygons were most noticeable for islands and border areas but these were often due to the combined error of the input coordinates and the inaccuracy of country polygons. The database is available at http://www.mosquitomap.org and in future will be available with additional fields, through the Smithsonian Institution's collections online database.

There are 492 species listed and 111 (23%) of these were singletons, being only recorded from one location (Fig. 1). Highest density of observations per 1° grid cell was recorded for the Canal zone in Panama, northern Venezuela, and French Guiana. Figure 2 shows species ranked according to the number of locations where they were collected. *Anopheles albimanus* Wiedemann appears at the most sites. Table 1 shows the number of locations and the species number per country. The numbers of species recorded for countries according to the SCC are included in Table 1. In a number of cases the present data set has more species than were recorded for the SCC, although the SCC is regularly updated and may change to eliminate this discrepancy. Compared with the species number recorded for countries in the SCC, the MIMP database under-samples Argentina, The Bahamas, Bermuda, Bolivia, Cuba, and Peru, in particular.

Figure 3 shows the log–log species–area relationship using the number of species per country in the SCC (or the MIMP database, whichever is the higher number according to Table 1).

Fig. 1. Collection locations ($n = 1853$) in the MIMP database. Larger circles indicate location of species collected at only one location. Note that many record locations are not visible due to the scale.
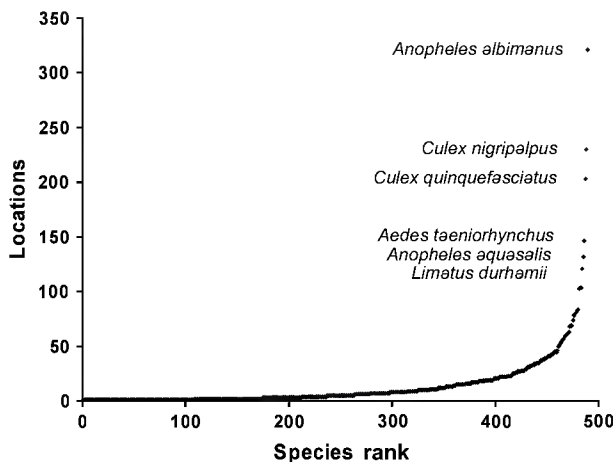
Fig. 2. Species in the MIMP database ranked according to the number of locations where they were collected.

A trend of increasing species number with increasing land area was found, as was shown by Foley *et al.* (2007) for all countries. A linear regression of log area (*x*) against log species number (*y*) for each country showed a highly significant positive relationship ($y = 0.3547x + 0.1064$, $r^2 = 0.6685$, $F_{1,40} = 80.67$, $P < 0.0001$, residual mean-square error $= 0.133$, SE of intercept $= 0.1784$; SE of slope $= 0.03949$). This indicated that 66.85% of the variation in species richness is explained by area. Islands in the study area have generally smaller land area and less species compared with mainland countries. For mainland countries, Chile appears to have the least species for its land area. By contrast, the island nation of Trinidad-Tobago has a relatively species-rich mosquito fauna for its land area.

The results of the species distribution modelling are shown in terms of predicted grid-cell area in Fig. 4. Approximately 200 species had a predicted species range of only one grid cell or less, and it is likely that more sample points are needed for these species to adequately model their distribution. *Aedeomyia squamipennis* (Lynch Arribalzaga) had the largest predicted species range. Generally *Anopheles* species had the largest species ranges (Fig. 5), whereas species of *Deinocerites* and *Wyeomyia* had the smallest predicted species range.

Figure 6 shows number of species for each 1-degree grid cell that contained observations. The most species rich grid cells often contained the greatest number of observations (data not shown). Figure 7 shows the result of summing the potential species ranges of all species. Greatest species richness occurred in Trinidad-Tobago, French Guiana, Colombia, and Brazil. Greatest overlap of these two methods occurred on the border of Costa Rica and Nicaragua and in French Guiana.

Figure 8 shows potential endemicity including a close-up centred on Panama. Only models that resulted in two or more grid cells were included, which could preferentially eliminate species that were recorded from only one location. Hotspots may be influenced by collecting effort but Fig. 8 shows that hotspots often occur in unsampled areas.

The curves resulting from the ESTIMATES analysis are given in Fig. 9. ESTIMATES advised that the coefficient of variation was > 0.5, and recommended re-computing *Chao2* using the Classic instead of the Bias-Corrected option and reporting the larger of *Chao2* and *ICE* (which in this case was *ICE*) as the best estimate for incidence-based richness. By the end of the samples, the *ICE* estimator was only 3% higher than the *Chao2* value but was 34% higher than the *Sobs* curve. According to Heyer *et al.* (1999), for a complete inventory the estimators and *Sobs* coincide and asymptote together, whereas for a relatively under-sampled taxon the estimator curves are much higher (e.g. 65%) than the observed curves. In the most under-sampled taxa, the *Sobs* curve may also be linear (Heyer *et al.*, 1999), but this was not observed in the present study. A discrepancy between Coleman and *Sobs* curves in Fig. 9 is evidence of patchiness in the distribution of data points, especially for rare species.

Discussion

The present database is the largest digitised collection of georeferenced species occurrence records for vouchered Neotropical mosquitoes. Analyses of species richness were undertaken to demonstrate the utility of these data to answer fundamental questions about mosquito biogeography, and to encourage others to make available their georeferenced mosquito collection records. A worldwide database of georeferenced mosquito collection records would enable new insights into global patterns of mosquito biodiversity and survey history.

Jablonski *et al.* (2006) identified origination rate, extinction rate, and migration as the fundamental determinants of spatial patterns of biodiversity. The species–area relationship for a country is expected to primarily reflect the sampling effort, evolutionary history, and the intrinsic ability of the land to support different mosquito species. Variation in the productivity, heterogeneity and stability of the environment may also be important for species richness. A latitudinal gradient of increasing biodiversity from the poles to the equator has been noted for many organisms (see Jablonski *et al.*, 2006), including mosquito species (Foley *et al.*, 2007). The underlying mechanism for this phenomenon may vary depending on the organism. Allen and Gillooly (2006) showed that for fossil ocean plankton, species richness and speciation rates both peak near the equator even after controlling for sampling effort and habitat area. Jablonski *et al.* (2006) showed that for genera and subgenera of marine bivalves, taxa have preferentially originated in the tropics and expanded toward the poles without losing their tropical presence. Weir and Schluter (2007), however, found that for birds and mammals faster speciation at higher latitudes contributes to the latitudinal diversity gradient.

The latitudinal biodiversity gradient, spatial scale, and the species–time relationship interact with the species–area relationship (e.g. Turner & Tjørve, 2005; Drakare *et al.*, 2006; White *et al.*, 2006) but despite this complication, two thirds of the variation in mosquito species richness for countries in this study is explained by area. Foley *et al.* (2007) showed a similar species–area relationship using worldwide country species data. A pattern of increasing species numbers with area was promoted by MacArthur and Wilson (1967) through their theory of

Table 1. Number of locations and mosquito species for countries in the Mosquito Information Management Project database compared with the number of species recorded in the SCC (accessed 17 June 2006). Where species number in MIMP is for one island of a country the number is preceded by >. Area is in square km.

Country	SCC	MIMP	% species	No. locations	Land area	Locations/area
Anguilla	–	5	>100*	10	102	0.0980‡
Antigua and Barbuda	7	>16	>100*	38	442.6	0.0859‡
Argentina	183	37	20.2†	20	2 736 690	0.0000
The Bahamas	19	4	21.1†	7	10 070	0.0007
Barbados	5	8	160.0*	37	431	0.0858‡
Belize	91	49	53.8	30	22 806	0.0013
Bermuda	1	–	0.0†	0	53.3	0.0000
Bolivia	157	9	5.7†	11	1 084 390	0.0000
Brazil	447	141	31.5	135	8 456 510	0.0000
Cayman Is.	12	20	166.7*	15	262	0.0573‡
Chile	13	5	38.5	9	748 800	0.0000
Colombia	251	147	58.6	94	1 038 700	0.0001
Costa Rica	154	128	83.1	215	50 660	0.0042
Cuba	48	1	2.1†	7	110 860	0.0001
Dominica	9	19	211.1*	60	754	0.0796‡
Dominican Republic	41	40	97.6	76	48 380	0.0016
Ecuador	118	69	58.5	65	276 840	0.0002
El Salvador	69	22	31.9	15	20 720	0.0007
French Guiana	224	123	54.9	62	89 150	0.0007
Grenada	17	22	129.4*	39	344	0.1134‡
Guadeloupe	14	14	100.0	26	1706	0.0152‡
Guatemala	105	67	63.8	69	108 430	0.0006
Guyana	84	47	56.0	16	196 850	0.0001
Haiti	27	17	63.0	15	27 560	0.0005
Honduras	82	47	57.3	24	111 890	0.0002
Jamaica	57	28	49.1	51	10 831	0.0047
Martinique	6	6	100.0	5	1060	0.0047
Mexico	211	108	51.2	107	1 923 040	0.0001
Montserrat	5	18	360.0*	40	102	0.3922‡
Saint Kitts and Nevis	3	>10	>100.0*	19	261	0.0728
Nicaragua	81	57	70.4	39	120 254	0.0003
Panama	264	172	65.2	214	75 990	0.0028
Paraguay	63	29	46.0	20	397 300	0.0001
Peru	130	18	13.8†	17	1 280 000	0.0000
Puerto Rico	35	15	42.9	31	8870	0.0035
Saint Lucia	13	7	53.8	13	606	0.0215‡
St Vincent/Grenadines	–	>1	>100.0*	1	389	0.0026
Suriname	161	52	32.3	20	161 470	0.0001
Trinidad and Tobago	125	56	44.8	70	5128	0.0137‡
Uruguay	54	–	–	0	173 620	0.0000
Venezuela	238	98	41.2	95	882 050	0.0001
Virgin Islands	14	13	92.9	17	346	0.0491‡

*Country records where the SCC could be updated.
†Where the present database is least representative of known species.
‡Where the density of collection sites is high.

island biogeography. Chile was shown to have relatively few species for its area, possibly due in part to its higher latitude compared with the other countries in the analysis. In contrast, Trinidad-Tobago appears to have many species for its area.

Inspection of Fig. 3 suggests that the expected species number for an area the size of Trinidad-Tobago is about 30, compared with the 122 species recorded. Foley *et al.* (2007) demonstrated that island countries have higher endemicity and are possibly better sampled for mosquitoes than are mainland countries.

Trinidad-Tobago is species rich, however, even compared with other island nations. The species richness of Trinidad-Tobago has also been shown for bats and passerine birds (Koopman, 1958). These islands are considered continental, and would have been joined to the mainland during the Pleistocene (Koopman, 1958). Only 33 species, however, are common between Trinidad-Tobago and Venezuela (the closest continental area), including five of seven *Wyeomyia*, three of eight *Anopheles*, six of 13 *Culex*, and five of seven *Aedes*. Trinidad showed a high
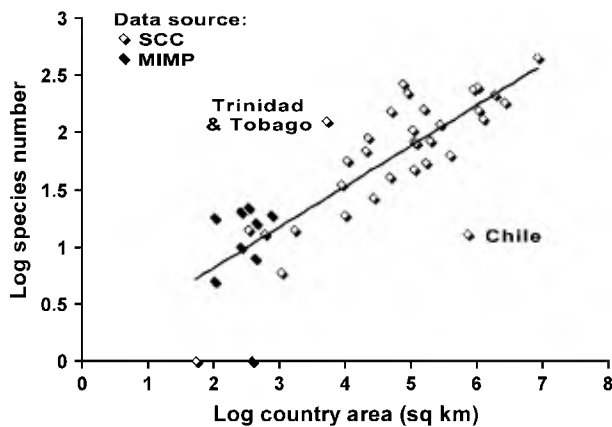
Fig. 3. Mosquito log–log species–area relationship for 42 countries covered by the MIMP database. Species numbers were obtained from the SCC (or the MIMP database, whichever was highest). A linear regression line of best fit is shown.

potential species richness and endemism based on ENM, supporting the idea of a generalised environmental suitability for species generation and/or survival on that island.

Potential hotspots in species endemism and richness could provide a guide to the most productive locations for future mosquito biodiversity surveys. Some Neotropical areas are known hotspots in diversity and endemism for plants and vertebrates, that is Brazil's Atlantic Forest, the Caribbean islands, Tropical Andes, and Mesoamerica (Myers *et al.*, 2000). Belkin *et al.* (1965) regarded the mosquito fauna of an area he described as Middle America, to be the most endemic in the world. Comparison of maps of potential mosquito species richness and endemism generated from ENM (Figs 7 and 8) revealed much congruence. Using birds as an example, however, Orme *et al.* (2005) found that areas of species richness and endemism are not usually congruent.

Compared with the arthropod, vertebrate, and plant species analysed in Heyer *et al.* (1999), the mosquito curves in Fig. 9
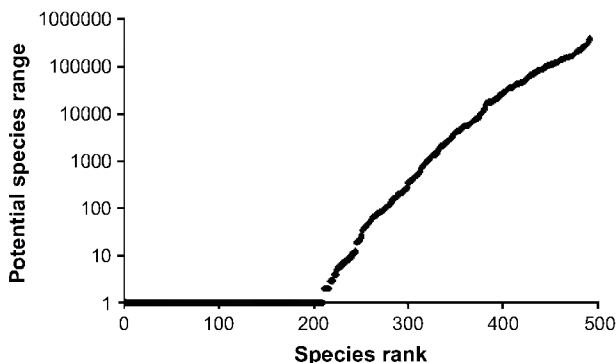
Fig. 4. Log potential species range (number of 2.5 min grids) according to BIOCLIM ecological niche modelling for mosquito species in the MIMP database. Only predicted species ranges of two to 100 000 grid cells were used in the analysis of endemicity (Fig. 8).

suggests an inventory that is intermediate in completeness. Values for *Sobs* and *ICE* were still rising for the mosquito data rather than reaching an asymptote, and the uniques curve is level rather than declining. For a complete inventory, values for uniques tend toward zero, as they will have been observed multiple times. It is possible that the rarity of many mosquito species is an artefact, perhaps by non-random sampling, which distorts the results. This possibility and the patchiness in the distribution of data points, especially for rare species, suggest that care should be exercised in the interpretation of species richness estimates from the mosquito data.

One of the few studies of mosquito species richness was conducted by Montes (2005) in the Serra do Cantareira State Park of Brazil. That author found that the forest environment had the highest species richness and the peri-domestic environment had the most dominant species (*Culex vaxus* Dyar). Loss of species and an increase in dominance are a typical outcome of environmental degradation (Magurran, 2004). Lande *et al.* (2000) noted that species accumulation curves might be unreliable when there is a mixture of assemblages that differ in species richness and evenness. Magurran (2004) noted that samples taken from an assemblage where one species dominates and the others are rare would tend to under-estimate richness. The most commonly encountered species in the present database was *Anopheles albimanus* and 23% of species were recorded from just one location, suggesting that species evenness was low. Researchers should be cognizant of the impact of humans on the diversity and composition of mosquito species and the potential effect of development on species richness estimators.

In addition, climate change severely impacts the species ranges of many organisms and is an important force for change, now and in the future (Parmesan & Yohe, 2003). It is evident that anthropogenic changes in habitats and climate affect invertebrate diversity (Thomas *et al.*, 2004) and community processes such as insect pollination (Biesmeijer *et al.*, 2006). Changes in land cover and/or climate may have important effects on the distribution and intensity of mosquito-borne diseases on regional scales (e.g. Martens *et al.*, 1999; Lindsay & Bayoh, 2004; Zhou *et al.*, 2004; Munga *et al.*, 2006). Rubio-Palis and Zimmerman (1997) concluded that irrigation, clearing of tropical forest for subsistence agriculture, animal husbandry and mineral exploitation, and construction of dams complicate an ecoregional approach to classifying vector distribution in the Neotropics. It is likely that the distributions of non-vector species of mosquito are also affected by the activities of a growing human population in the Neotropical region. Databases of mosquito collection data may assist in understanding the scope and intensity of human-mediated environmental changes on a regional or global scale.

A number of assumptions and limitations are inherent in the present study. For instance, it was assumed that sampling is uniform and species-blind, that is a complete species inventory. Hijmans *et al.* (2000) identified four types of bias that could apply in the present case, namely: species bias (e.g. oversampling species of *Anopheles* due to greater abundance or in connection to malaria studies); species–area bias (e.g. oversampling island endemics compared with mainland species); hotspot bias (e.g. over-sampling areas where previous studies
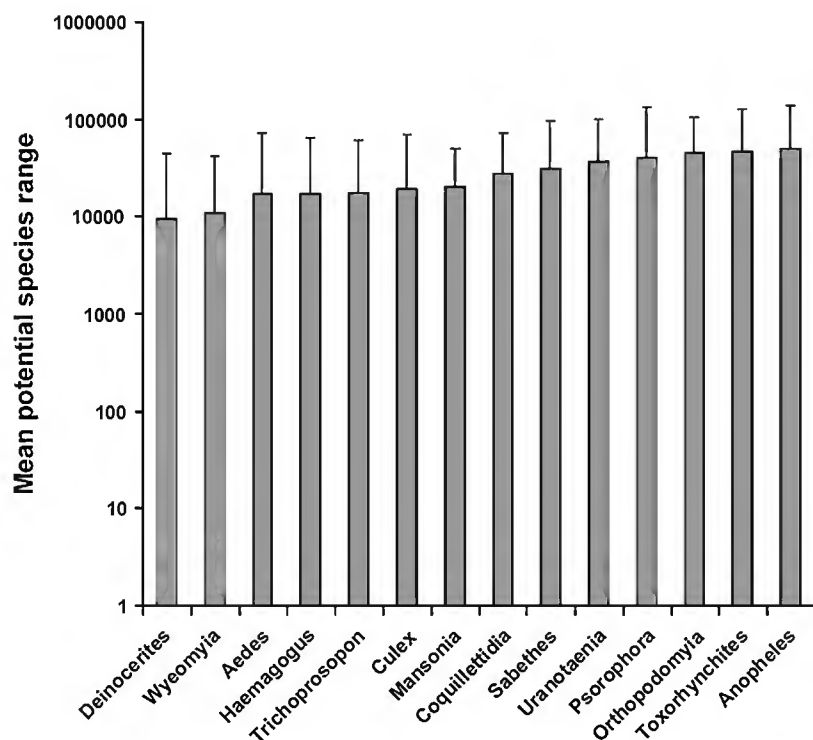
Fig. 5. Mean log potential species range (2.5 min grids) + 1 SD according to BIOCLIM ecological niche modelling in DIVA-GIS for mosquito genera in the MIMP database. Only data for genera with five or more species are shown.

indicated a high species richness); and infrastructure bias (e.g. over-sampling near roads and towns). An additional source of bias is reporting bias (i.e. not reporting unsuccessful attempts to collect mosquitoes). Sampling bias will favour certain species, so the database should be viewed as a record of species presence rather than absence.

It was beyond the scope of this study to evaluate the ecological realism of distribution models. Estimates of species ranges
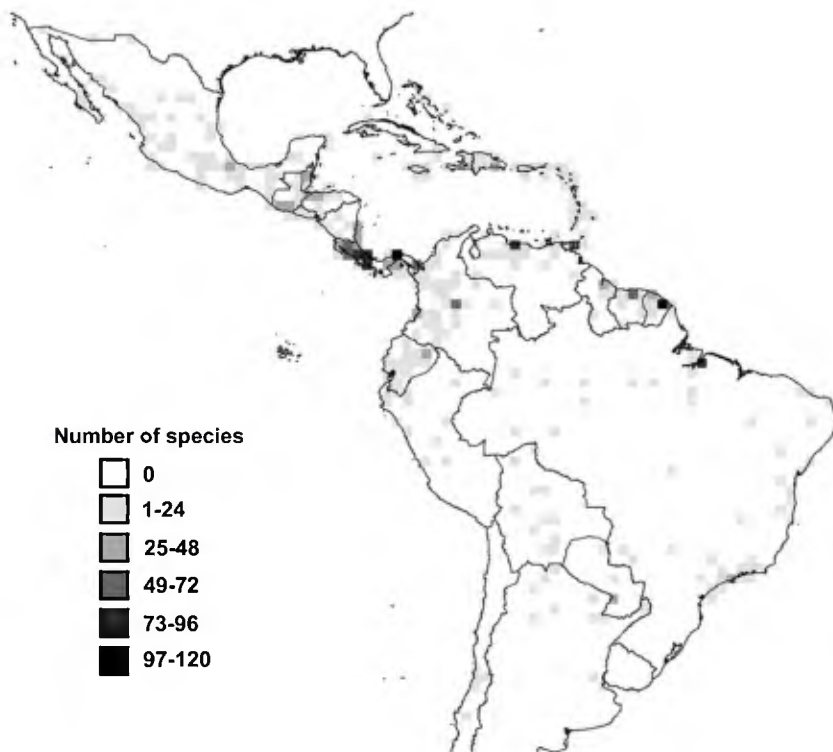
Fig. 6. Observed species richness for 1-degree grids calculated in DIVA-GIS from 1853 locations.

Number of species

- ☐ 0
- ☐ 1-26
- ☐ 27-52
- ☐ 53-78
- ■ 79-103
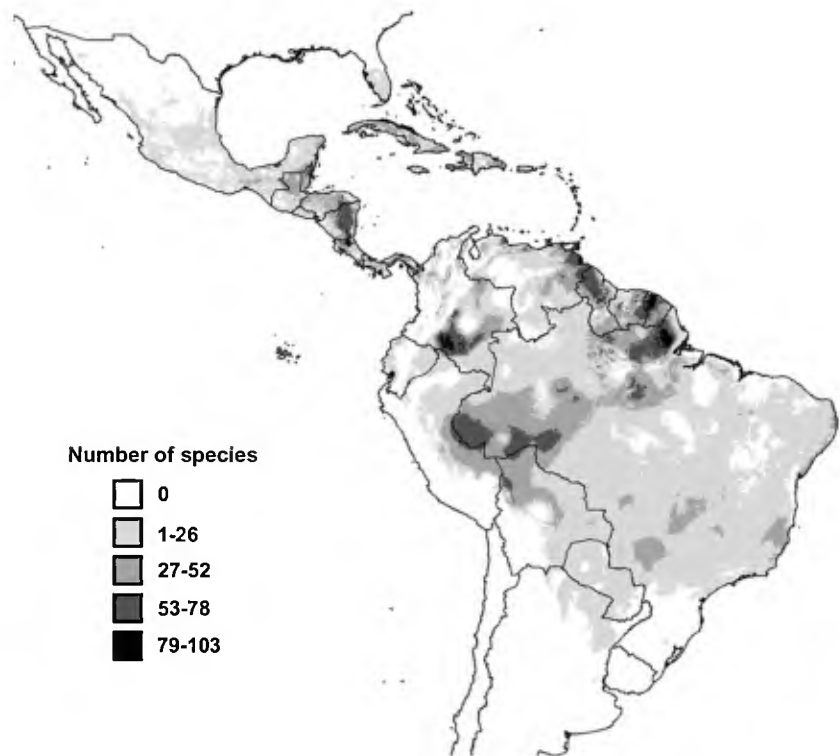
Fig. 7. Map of species richness created by summing the potential distribution (BIO-CLIM) of species from the MIMP database.

from ENM did not factor in historical effects on species dispersal and survival, and may therefore overestimate some current species ranges. Species whose predicted distribution covered a minimum of two grid cells were included in the esti-

mates of species richness and endemism but additional sampling will probably increase predicted range, thereby affecting these estimates. A better assessment of potential distribution will require greater sampling, additional environmental layers,

Fig. 8. Predicted hotspots in endemicity based on the MIMP database. Hotspots are shown by darker colour grids derived by summing grids of species potential range (*n* = 255), weighted so that species predicted to have relatively restricted range were given greater weight than species predicted to be more widespread. For display purposes, hotspots were accentuated by calculating a 9 × 9 neighbourhood size based on the maximum grid value. Inset shows collection locations on the endemicity map centred over Panama to show that potential hotspots in endemicity can occur beyond currently sampled areas.
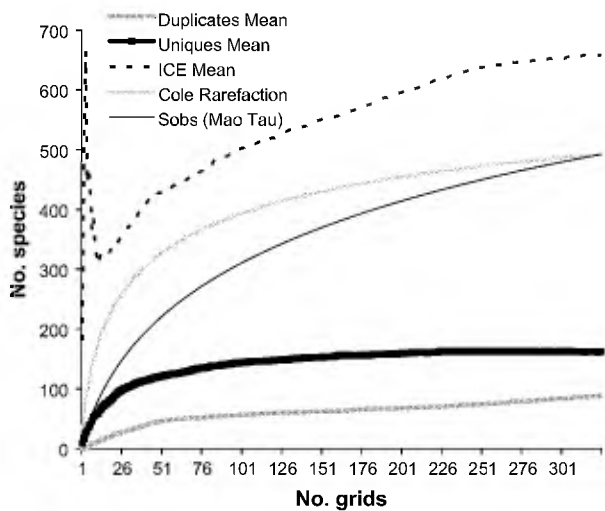
Fig. 9. Species richness estimators and patchiness indicators for mosquito species from the Mosquito Information Management Project database calculated with the program ESTIMATES. Sobs = empirical species accumulation curve; ICE = coverage estimator of Chao and Lee (1992); Cole Rarefaction = Coleman curve, a patchiness indicator, of Coleman (1981); uniques = number of species occurring once in just one grid cell; duplicates = number of species occurring in just two grid cells.

and a statistical treatment of the reliability of distribution models, as is available in the modelling procedure Genetic Algorithm for Rule Set Prediction (Stockwell & Noble, 1992) available in the DESKTOP GARP software.

Another limiting factor is that mosquito taxonomy is still largely at the alpha stage of species discovery and description (Zavortink, 1990). Belkin *et al.* (1965) speculated that over 1000 species would eventually be described from Middle America. Middle America included the southern U.S.A. and only a part of Colombia, Ecuador, and Peru, so it does not exactly coincide with the country coverage of the present database. As of 22 September 2006, however, the SCC reported 852 different species from the countries covered by Belkin *et al.* (1965), that is U.S.A. plus the 42 countries in the present database, less: Argentina, Bolivia, Brazil, Chile, Paraguay, and Uruguay. The SCC reported 998 different species from the 42 countries covered by the present database, but this is far from a complete species inventory for the region. The present data set includes collection details for about half that number of species. About half the records in the original MIMP database, however, were not included in the data set presented here due to a lack of a georeference or because of uncertain species identification.

Faran *et al.* (1984) discussed the limitations of the MIMP database, which included variable coverage (good for the Canal Zone of Panama and poor for Amazonia and Patagonia), potential bias in the type of habitat surveyed (towards ground pools), and a lack of uniformity in collection method. Most records in the MIMP database are larval collections and Fig. 2 indicates that the five most commonly collected species were ground pool breeders, which may indicate a collection bias towards ground pools. Larvae of the sixth most commonly collected mosquito,

Limatus durhamii Theobald, however, are usually found in natural container habitats, which suggest that this bias, if present, is not absolute.

Many uses of primary species-occurrence data are possible (Chapman, 2005c) and methods are available to ensure data quality (e.g. Chapman, 2005 a,b). The database could be improved by including estimates of georeference error, e.g. by the point-radius method (Wieczorek *et al.*, 2004). Almost 9000 mosquito records in the original MIMP database lacked a georeference but could be added with new software (e.g. BIOGEOMANCER, http://www.biogeomancer.org/) that can parse text location descriptions in batch mode and calculate geographic coordinates. Most (38 of 42) countries had records that lacked a georeference; no country bias could be discerned. Inclusion of these records may increase the number of species for each country but probably not substantially. For example, the data for Argentina, the least representative for known species (Table 1), would gain 22 records (4%) and two additional species (a 5% increase). Over 7000 georeferenced records from Egypt, Israel, and Kenya were present in the original database but were not included in this study.

The addition to this database of other georeferenced collection records from the Neotropics and other parts of the world would advance the reliability of species richness estimates and knowledge of global mosquito biogeography. Thousands of mosquito collection records remain to be digitised at the WRBU, and many more exist in other museums and research institutes around the world. The availability of mosquito collections and associated data may reduce the cost of studying the transmission of many mosquito-borne pathogens (Suarez & Tsutsui, 2004). Despite their value, collection records are often not digitised, are not easily accessible, and as was learnt for the MIMP database, vulnerable to loss. Species occurrence data for many organisms, however, are increasingly available electronically (Graham *et al.*, 2004), and the MIMP database and cleaned subset presented here will be accessible through the Smithsonian Institution's online collections database. An online database of worldwide mosquito collection records would have many benefits. For example, it would allow the high resolution prediction of the potential distribution for each species, allow powerful insights into mosquito community structure and ecological and climatic correlates to species occurrence (ecological niche), enable predictions about the potential spread of exotic mosquito introductions, help identify cryptic evolutionary lineages that differ in geographic or ecological space, allow the location of biodiversity hotspots, and enable the identification of under-sampled areas in need of further mosquito collecting. An online repository of mosquito collection information may also encourage the standardisation of collection reporting, and the digitising and contribution of past collection records. Suarez and Tsutsui (2004) concluded that the rate collection records are entered into databases and made accessible must be increased to benefit society. Mapping tools and the Internet are now available to enrich the information content and global reach of mosquito collection databases by enabling online users to analyse records and visualise results in spatially explicit ways.

Acknowledgements

References

Allen, A.P. & Gillooly, J.F. (2006) Assessing latitudinal gradients in speciation rates and biodiversity at the global scale. *Ecology Letters*, **9**, 947–954.

Biesmeijer, J.C., Roberts, S.P.M., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T. *et al.* (2006) Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science*, **313**, 351–354.

Belkin, J.N. (1962) *The Mosquitoes of the South Pacific (Diptera, Culicidae)*, Vol. 2, pp. 1–412. University of California Press, Berkeley, California.

Belkin, J.N. & Heinemann, S.J. (1973) Collection records of the project 'Mosquitoes of Middle America' 1. Introduction; Dominican Republic (RD0). *Mosquito Systematics*, **5**, 201–220.

Belkin, J.N. & Heinemann, S.J. (1975a) Collection records of the project 'Mosquitoes of Middle America' 2. Puerto Rico (PR, PRA, PRX) and Virgin Is. (VI, VIA). *Mosquito Systematics*, **7**, 269–296.

Belkin, J.N. & Heinemann, S.J. (1975b) Collection records of the project 'Mosquitoes of Middle America' 3. Bahama Is. (BAH), Cayman Is. (CAY), Cuba (CUB), Haiti (HAC, HAR, HAT) and Lesser Antilles (LAR). *Mosquito Systematics*, **7**, 367–393.

Belkin, J.N. & Heinemann, S.J. (1976a) Collection records of the project 'Mosquitoes of Middle America' 4. Leeward Islands: Anguilla (ANG), Antigua (ANT), Barbuda (BAB), Montserrat (MNT), Nevis (NVS), St. Kitts (XIT). *Mosquito Systematics*, **8**, 123–162.

Belkin, J.N. & Heinemann, S.J. (1976b) Collection records of the project 'Mosquitoes of Middle America' 5. French West Indies: Guadeloupe (FWI) and Martinique (FWIM, MAR). *Mosquito Systematics*, **8**, 163–193.

Belkin, J.N. & Heinemann, S.J. (1976c) Collection records of the project 'Mosquitoes of Middle America' 6. Southern Lesser Antilles: Barbados (BAR), Dominica (DOM), Grenada (GR, GRR), St. Lucia (LU), St. Vincent (VT). *Mosquito Systematics*, **8**, 237–297.

Belkin, J.N., Schick, R.X., Galindo, P. & Aitken, T.H.G. (1965) Mosquito studies (Diptera, Culicidae). I. A project for a systematic study of the mosquitoes of Middle America. *Contributions of the American Entomological Institute*, **1**, 1–78.

Chao, A. & Lee, S.M. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210–217.

Chapman, A.D. (2005a) *Principles and Methods of Data Cleaning – Primary Species and Species-Occurrence Data, Version 1.0.* Report for the Global Biodiversity Information Facility, Copenhagen, Denmark.

Chapman, A.D. (2005b) *Principles of Data Quality, Version 1.0.* Report for the Global Biodiversity Information Facility, Copenhagen, Denmark.

Chapman, A.D. (2005c) *Uses of Primary Species-Occurrence Data, Version 1.0.* Report for the Global Biodiversity Information Facility, Copenhagen, Denmark.

Coleman, B.D. (1981) On random placement and species–area relations. *Mathematical Bioscience*, **54**, 191–215.

Colwell, R.K. (2005) *EstimateS: Statistical Estimation of Species Richness and Shared Species from Samples. Version 7.5.* (User's Guide and application at: http://purl.oclc.org/estimates).

Creighton, R.A. & Crockett, J.J. (1971) SELGEM: A system for collection management. Smithsonian Institution, Washington DC. *Smithsonian Institution Information Systems Innovations*, **2**, 1–35.

Drakare, S., Lennon, J.J. & Hillebrand, H. (2006) The imprint of the geographical, evolutionary and ecological context on species–area relationships. *Ecology Letters*, **9**, 215–227.

Elith, J., Graham, C.H., Anderson, R.P., Dudik, M., Ferrier, S., Guisan, A. *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.

Faran, M.E., Burnett, C., Crockett, J.J. & Lawson, W.L. (1984) A computerized mosquito information and collection management system for systematic research and medical entomology. *Mosquito Systematics*, **16**, 289–307.

Foley, D.H., Rueda, L.M. & Wilkerson, R.C. (2007) Insight into global mosquito biogeography from country species records. *Journal of Medical Entomology*, **44**, 554–567.

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, **19**, 497–503.

Harbach, R.E. (1998) Phylogeny and classification of the Culicidae (Diptera). *Systematic Entomology*, **23**, 327–370.

Heinemann, S.J. (1980) A clarification of the format used in the collection records series of the project 'Mosquitoes of Middle America'. *Mosquito Systematics*, **12**, 333–334.

Heinemann, S.J. & Belkin, J.N. (1977a) Collection records of the project 'Mosquitoes of Middle America' 7. Costa Rica (CR). *Mosquito Systematics*, **9**, 237–287.

Heinemann, S.J. & Belkin, J.N. (1977b) Collection records of the project 'Mosquitoes of Middle America' 8. Central America: Belize (BH), Guatemala (GUA), El Salvador (SAL), Honduras (HON), Nicaragua (Nl, NIC). *Mosquito Systematics*, **9**, 403–454.

Heinemann, S.J. & Belkin, J.N. (1977c) Collection records of the project 'Mosquitoes of Middle America' 9. Mexico (MEX, MF, MT, MX). *Mosquito Systematics*, **9**, 483–535.

Heinemann, S.J. & Belkin, J.N. (1978a) Collection records of the project 'Mosquitoes of Middle America' 10. Panama, Including Canal Zone (PA, GG). *Mosquito Systematics*, **10**, 119–196.

Heinemann, S.J. & Belkin, J.N. (1978b) Collection records of the project 'Mosquitoes of Middle America' 11. Venezuela (VZ); Guianas: French Guiana (FG, FGC), Guyana (GUY), Surinam (SUR). *Mosquito Systematics*, **10**, 365–459.

Heinemann, S.J. & Belkin, J.N. (1978c) Collection records of the project 'Mosquitoes of Middle America' 12. Colombia (COA, COB, COL, COM). *Mosquito Systematics*, **10**, 493–539.

Heinemann, S.J. & Belkin, J.N. (1979) Collection records of the project 'Mosquitoes of Middle America' 13. South America: Brazil (BRA, BRAP, BRB), Ecuador (ECU), Peru (PER), Chile (CH). *Mosquito Systematics*, **11**, 61–118.

Heinemann, S.J., Aitken, T.H.G. & Belkin, J.N. (1980) Collection records of the project 'Mosquitoes of Middle America'. 14. Trinidad and Tobago (TR, TRM, TOB). *Mosquito Systematics*, **12**, 179–284.

Heyer, W.R., Coddington, J., Kress, W.J., Acevedo, P., Cole, D. Erwin, T.L. *et al.* (1999) Amazonian biotic data and conservation decisions. *Environment and Biodiversity*, **51**, 372–385.

Hijmans, R.J., Garrett, K.A., Huaman, Z., Zhang, D.P., Schreuder, M. & Bonierbale, M. (2000) Assessing the geographic representativeness of genebank collections: the case of Bolivian wild potatoes. *Conservation Biology*, **14**, 1755–1765.

Jablonski, D., Roy, K. & Valentine, J.W. (2006) Out of the tropics: evolutionary dynamics of the latitudinal diversity gradient. *Science*, **314**, 102–105.

Jarvis, A., Ferguson, M.E., Williams, D.E., Guarino, L., Jones, P.G., Stalker, H.T. *et al.* (2003) Biogeography of wild *Arachis*: assessing conservation status and setting future priorities. *Crop Science*, **43**, 1100–1108.

Kerr, J.T., Southwood, T.R.E. & Cihlar, J. (2001) Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 11365–11370.

Koopman, K.F. (1958) Land bridges and ecology in bat distribution on islands off the northern coast of South America. *Evolution*, **12**, 429–439.

Krzywinski, J., Wilkerson, R.C. & Besansky, N.J. (2001) Toward understanding anophelini (Diptera, Culicidae) phylogeny: insights from nuclear single copy genes and the weight of evidence. *Systematic Biology*, **50**, 540–556.

Lande, R., DeVries, P.J. & Walla, T. (2000) When species accumulation curves intersect: implications for ranking diversity using small samples. *Oikos*, **89**, 601–605.

Lindsay, S.W. & Bayoh, M.N. (2004) Mapping members of the Anopheles gambiae complex using climate data. *Physiological Entomology*, **29**, 204–209.

MacArthur, R.H. & Wilson, E.O. (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton, New Jersey.

Magurran, A.E. (2004) *Measuring Biological Diversity*. Blackwell Publishing, Malden, Massachusetts.

Martens, W.J.M., Kovats, R.S., Nijhof, S., deVries, P., Livermore, M.J.T., McMichael, A.J. *et al.* (1999) Climate change and future populations at risk of malaria. *Global Environmental Change*, **9**, S89–S107.

Montes, J. (2005) Fauna de Culicidae da Serra da Cantareira, Sao Paulo, Brasil. *Revista de Saude Publica*, **39**, 578–584.

Munga, S., Minakawa, N., Zhou, G., Mushinzimana, E., Barrack, O.-O.J., Githeko, A.K. *et al.* (2006) Association between land cover and habitat productivity of malaria vectors in western Kenyan highlands. *American Journal of Tropical Medicine and Hygiene*, **74**, 69–75.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nix, H.A. (1986) A biogeographic analysis of Australian elapid snakes. *Atlas of Elapid Snakes of Australia* (ed. by R. Longmore). Australian Flora and Fauna Series No 7. ACT: Australian Government Publishing Service, pp.4-15, Canberra, Australia.

Orme, C.D.L., Davies, R.G., Burgess, M., Eigenbrod, F., Pickup, N., Olson, V.A. *et al.* (2005) Global hotspots of species-richness are not congruent with endemism or threat. *Nature*, **436**, 1016–1019.

Parmesan, C. & Yohe, G. (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Pereira, H.M. & Cooper, H.D. (2006) Towards the global monitoring of biodiversity change. *Trends in Ecology and Evolution*, **21**, 123–129.

Qu, F.-Y. & Qian, G.-Z. (1989) Faunistic distribution and evolution of culicid mosquitoes. *Acta Zootaxonomica Sinica*, **14**, 468–474.

Raven, P.H. (ed.) (1980) *Research Priorities in Tropical Biology*. National Academy of Sciences, Washington, D.C. pp. xii + 116.

Rogers, D.J. & Randolph, S.E. (2003) Studying the global distribution of infectious diseases using GIS and RS. *Nature Reviews Microbiology*, **1**, 231–237.

Rubio-Palis, Y. & Zimmerman, R.H. (1997) Ecoregional classification of malaria vectors in the Neotropics. *Journal of Medical Entomology*, **34**, 499–510.

Shattuck, S.O. (1997) eGAZ, The electronic gazetteer. *Australian National Insect Collection News*, **11**, 9.

Soberon, J.M., Llorente Bousquets, J.E. & Onate, L. (2000) The use of specimen-label databases for conservation purposes: an example using Mexican papilionid and pierid butterflies. *Biodiversity and Conservation*, **9**, 1441–1466.

Stockwell, D.R.B. & Noble, I.R. (1992) Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. *Mathematics and Computers in Simulation*, **32**, 249–254.

Suarez, A.V. & Tsutsui, N.D. (2004) The value of museum collections for research and society. *BioScience*, **54**, 66–74.

Tatem, A.J., Hay, S.I. & Rogers, D.J. (2006) Global traffic and disease vector dispersal. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 6242–6247.

Thomas, J.A., Telfer, M.G., Roy, D.B., Preston, C.D., Greenwood, J.J.D., Asher, J., *et al.* (2004) Comparative losses of British butterflies, birds, and plants and the global extinction crisis. *Science*, **303**, 1879–1881.

Turner, W.R. & Tjørve, E. (2005) Scale-dependence in species–area relationships. *Ecography*, **28**, 721–730.

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E. & Steininger, M. (2003) Remote sensing for biodiversity science and conservation. *Trends in Ecology and Evolution*, **18**, 306–314.

Weir, J.T. & Schluter, D. (2007) The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science*, **315**, 1574–1576.

White, E.P., Adler, P.B., Lauenroth, W.K., Gill, R.A., Greenberg, D., Kaufman, D.M. *et al.* (2006) A comparison of the species-time relationship across ecosystems and taxonomic groups. *Oikos*, **112**, 185–195.

Wieczorek, J., Guo, Q. & Hijmans, R.J. (2004) The point-radius method for georeferencing locality descriptions and calculating associated uncertainty. *International Journal of Geographical Information Science*, **18**, 745–767.

Zavortink, T.J. (1990) Classical taxonomy of mosquitoes – a memorial to John N. Belkin. *Journal of the American Mosquito Control Association*, **6**, 593–599.

Zavortink, T.J. & Poinar, G.O. Jr (2000) *Anopheles (Nyssorhynchus) dominicanus* sp.n. (Diptera: Culicidae) from Dominican amber. *Annals of the Entomological Society of America*, **93**, 1230–1235.

Zhou, G., Minakawa, N., Githeko, A. & Yan, G. (2004) Association between climate variability and malaria epidemics in the East African highlands. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 2375–2380.