Title

Areas of endemism in the Nearctic: a case study of 1339 species of Miridae (Insecta: Hemiptera) and their plant hosts

Permalink

https://escholarship.org/uc/item/2ff2p97p

Journal

Cladistics, 33(3)

ISSN

0748-3007

Authors

Weirauch, Christiane
Seltmann, Katja C
Schuh, Randall T
et al.

Publication Date

2017-06-01

DOI

10.1111/cla.12169

Cladistics

Cladistics 33 (2017) 279–294

10.1111/cla.12169

Areas of endemism in the Nearctic: a case study of 1339 species of Miridae (Insecta: Hemiptera) and their plant hosts

Christiane Weirauch[a,]*, Katja C. Seltmann[b,c], Randall T. Schuh[b], Michael D. Schwartz[b], Christine Johnson[b], Mary Ann Feist[d] and Pamela S. Soltis[e]

[a]*Department of Entomology, University of California Riverside, 900 University Avenue, Riverside, CA 92521, USA;* [b]*Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA;* [c]*Santa Barbara, Cheadle Center for Biodiversity & Ecological Restoration, University of California, Harder South Building 578, MS-9615, Santa Barbara, CA 93106, USA;* [d]*Wisconsin State Herbarium, University of Wisconsin-Madison, Birge Hall, 430 Lincoln Drive, Madison, WI 53706, USA;* [e]*Laboratory of Molecular Systematics & Evolutionary Genetics, Florida Museum of Natural History, University of Florida, 3215 Hull Road, Gainesville, FL 93106, USA*

Accepted 16 May 2016

Abstract

Areas of endemism are essential first hypotheses in investigating historical biogeography, but there is a surprising paucity of such hypotheses for the Nearctic region. Miridae, the plant bugs, are an excellent taxon to study in this context, because this group combines high species diversity, often small distribution ranges, a history of modern taxonomic revisions, and comprehensive electronic data capture and data cleaning that have resulted in an exceptionally error-free geospatial data set. Many Miridae are phytophagous and feed on only one or a small number of host plant species. The programs ndm/vndm are here used on plant bug and plant data sets to address two main objectives: (i) identify areas of endemism for plant bugs based on parameters used in a recent study that focused on Nearctic mammals; and (ii) discuss hypotheses on areas of endemism based on plant bug distributions in the context of areas identified by their host plant species. Given the narrow distribution ranges of many species of Miridae, the analytical results allow for tests of the prediction that areas of endemism for Miridae are smaller and more numerous, especially in the Western Nearctic, than are those of their host plants. Analyses of the default plant bug data set resulted in 45 areas of endemism, 35 of them north of Mexico and many located in the Western Nearctic; areas in the Nearctic are more numerous and smaller than those identified by mammals. The host plant data set resulted in ten areas of endemism, and even though the size range of areas is similar between the Miridae and plant data sets, the average area size is smaller in the Miridae data set. These results allow for the conclusion that the Miridae indeed present a valuable model system to investigate areas of endemism in the Nearctic.

Introduction

Areas of endemism, identified by the congruent distribution of two or more taxa of animals or plants, have long been recognized as one of the fundamental components of investigations into the historical biogeography of taxa or biogeographic regions (Morrone, 1994). Despite continued development of methods in parsimony- and likelihood-based approaches to historical biogeography (Ronquist, 2001; Nylander et al., 2008; Matzke, 2013, 2014; Yu et al., 2015) and a vast body of literature applying these methods, objective approaches to identifying areas of endemism have received comparatively little attention in theory and practice (Szumik et al., 2002; Torres-Miranda et al., 2013; Morrone, 2014a; Escalante, 2015). Among the more widely used methods are Parsimony Analysis of Endemicity (Morrone, 1994, 2014a,b), Biotic Elements Analysis (Hausdorf and Hennig, 2003) and Endemicity Analysis (EA; Szumik et al., 2002; Szumik and Goloboff, 2004), of which the last has been applied to a fairly broad range of taxa and biogeographic regions (Domínguez et al., 2006; Ferrari et al., 2010; Aagesen

---

*Corresponding author.
E-mail address: christiane.weirauch@ucr.edu

et al., 2012) and has been shown to outcompete the other two approaches when compared using hypothetical species distributions (Casagranda et al., 2012).

Thus far, the majority of EA analyses have focused on identifying areas of endemism across either the entire Neotropical region (e.g. Prado et al., 2015) or some of its subregions or provinces (e.g. Szumik et al., 2012; Del Ventura et al., 2013). Among the exceptions to this pattern are a worldwide analysis that uses specimen information from > 4200 species of Tipulidae (Ribeiro et al., 2014), a study that investigates areas of endemism among gypsum-tolerating plants in Spain (Martínez-Hernández et al., 2015), and analyses that are based on mammal distribution data and examine the southern limit of the Nearctic region (Escalante et al., 2010) and areas of endemism within that region (Escalante et al., 2013). Given the paucity of studies that have aimed at objectively identifying areas of endemism in the Nearctic (especially north of Mexico), the analysis by Escalante et al. (2013) is of particular interest. That study assembled a data set comprising 710 species of mammals represented by 245 818 records (i.e. unique combinations of taxon names and georeferences) ranging from Canada to Panama; pruned that data set to include only the 652 species represented by more than five records; modelled species habitats using MAXENT; overlaid these ranges with 2-degree grid cells; and then applied EA methods using ndm/vndm (Goloboff, 2011). Based on the selected parameters (e.g. saving sets with endemicity scores of > 2.0; calculating consensus at 40% of similarity with any area), the analysis resulted in 76 consensus areas, with only 21 of them located in, or mostly located in, the Nearctic region north of Mexico (here referred to as North America). Most of the obtained consensus areas entirely contained within North America are large and most of the areas with high endemicity scores are found in the Western United States, although exceptions include Florida and an area described as the eastern United States. The authors concluded that areas of endemism in North America are likely to be more complex than previously assumed. However, many mammal species in North America have fairly broad distribution ranges (Wilson and Reeder, 2005; Kays and Wilson, 2009) and this picture could potentially become even more complex, but in all likelihood also more refined, if EA analysis was based on groups of animal or plant species with well-documented, but smaller, distribution ranges.

We argue that Miridae (Insecta: Hemiptera), the plant bugs, may be such a group (Fig. 1). With more than 11 000 described species worldwide (Cassis and Schuh, 2012), Miridae are one of the largest family-level clades among the hemimetabolous insects. More than 2000 species of plant bugs are documented to occur in the Nearctic region (Schuh, 2002–2013). Importantly, Nearctic Miridae have been the focus of a large body of revisionary and monographic taxonomic work during the past century, and even though undescribed species are still being discovered and incorporated into taxonomic revisions in North America, the fauna is well documented (Cassis and Schuh, 2012). Building on seminal studies by Van Duzee (Van Duzee, 1916, 1918), and more importantly those of Knight (Knight, 1925, 1927, 1962, 1968), a number of researchers have made significant contributions to the study of Nearctic Miridae, especially starting in the 1970s. These contributions have come in the form of both dedicated field work across the region, resulting in vast holdings of well-preserved plant bug specimens in several institutions (e.g. American Museum of Natural History (R.T. Schuh, M.D. Schwartz, G.M. Stonedahl), United States National Museum (T.J. Henry), Texas A&M University Insect Collection (J.C. Schaffner) and several smaller university-based collections) and a large body of standardized and well-illustrated taxonomic publications that have established clear species concepts and provided a wealth of precise specimen locality data (Henry and Kim, 1984; Schwartz, 1984; Schuh and Schwartz, 1985, 1988; Schwartz and Stonedahl, 1986, 1987; Stonedahl and Schuh 1986; Stonedahl and Schwartz, 1986, 1988, 1996; Stonedahl, 1988, 1990; Asquith, 1991; Stonedahl and Henry, 1991; Schwartz and Foottit, 1998; Schwartz and Schuh, 1999; Schuh, 2000a,b, 2001). These efforts recently culminated in two large-scale U.S. National Science Foundation supported projects, the “Plant Bug Planetary Biodiversity Inventory (PBI)” (http://research.amnh.org/pbi/) and the Thematic Collections Network (TCN) “Plants, Herbivores, and Parasitoids: A Model System for the study of Tri-Trophic Associations (Tri-Trophic Database or TTD)” (http://tcn.amnh.org/). The Plant Bug PBI project, although focusing on only two of the eight subfamilies of Miridae and emphasizing field work and revisionary systematics in dramatically understudied regions such as Australia and South Africa, has advanced systematics of Nearctic Miridae through taxonomic revisions (Schuh, 2004a,b, 2006, 2008; Schuh and Schwartz, 2004, 2005; Schwartz, 2004, 2005, 2006, 2011; Schwartz and Stonedahl, 2004; Weirauch, 2006a,b, 2009; Henry, 2007; Forero, 2008; Schaffner and Schwartz, 2008; Wyniger, 2010, 2011), additional field work in the Western Nearctic including Mexico and, uniquely, through the onset of electronic specimen data capture using the PBI instance of the Arthropod Easy Capture database software application (AEC) (Schuh et al., 2010; Schuh, 2012). More recently, electronic data capture of Nearctic Hemiptera including Miridae was one of the main objectives of the TTD-TCN project, and due to the combined efforts of these projects, more than 295 723 individual specimen records for plant bugs are now publicly available through the Integrated Digitized Biocollections

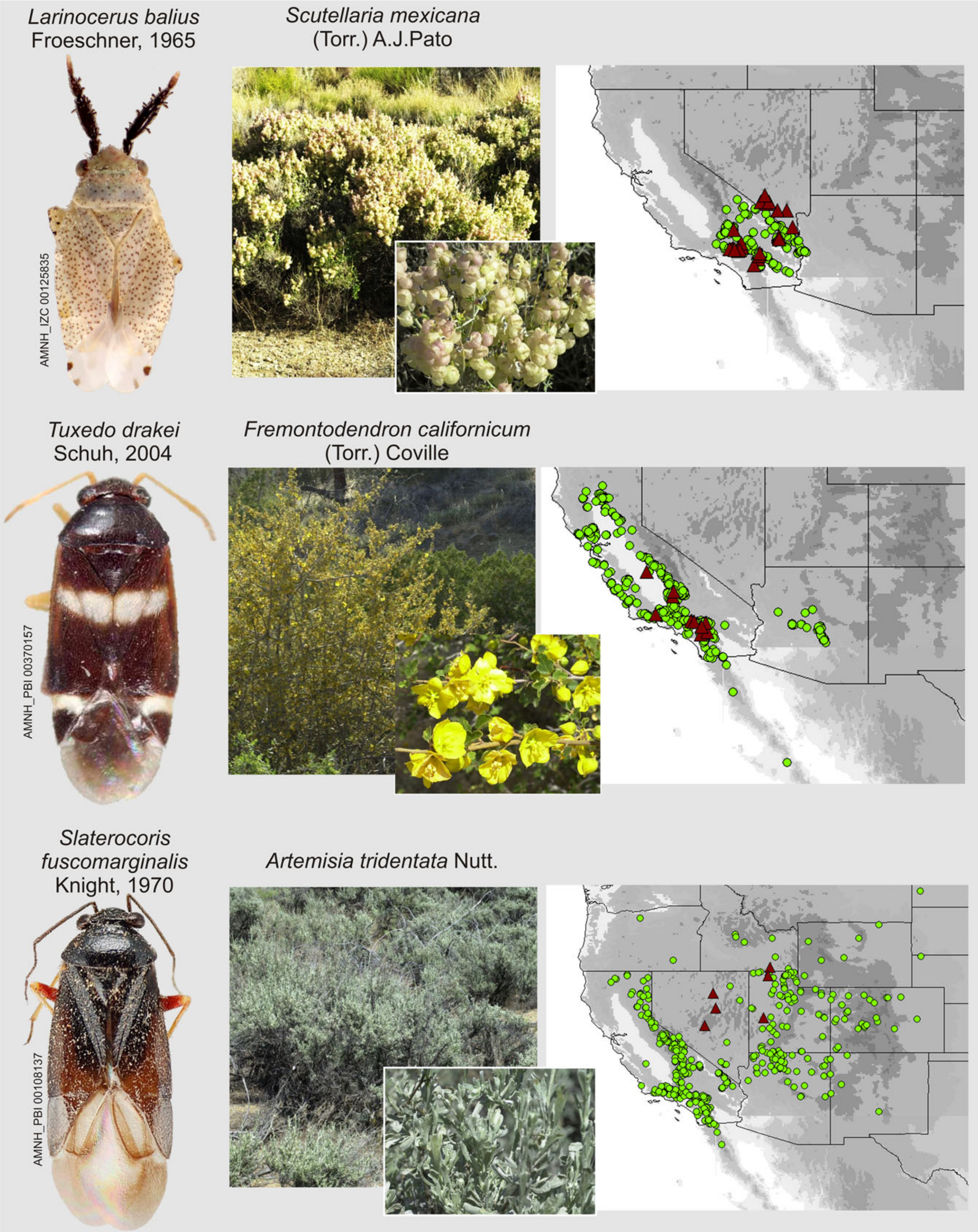

Fig. 1. Three species of Nearctic Miridae in the subfamilies Phylinae and Orthotylinae, their host plants and distribution ranges (plant bug species: maroon triangle; plant species: green circle), illustrating the relatively smaller distribution ranges of these plant bugs compared to their host plant. [Colour figure can be viewed at wileyonlinelibrary.com].

(iDigBio) portal (http://www.idigbio.org/portal/) and the Heteroptera Species Pages (http://research.amnh.org/pbi/heteropteraspeciespage/). Because of its history, this data set is uniquely suitable for analyses of areas of endemism: it is exceptionally "clean" with respect to species identifications as well as locality data, the latter due to extensive georeferencing efforts that have spanned the past 12 years.

Miridae have additional properties that make them an excellent model for the study of areas of endemism in the Nearctic, an area of about 24 million km^2 (http://data.worldbank.org/indicator/AG.SRF.TOTL.K2). As documented in numerous publications and dynamic maps produced from the electronically available data (see, e.g., http://www.discoverlife.org/tttcn/), many Nearctic species of Miridae have small distribution ranges, especially in the Western Nearctic including Mexico (e.g. Schwartz, 2004, 2005; Schuh, 2006, 2008). In addition, most Miridae are phytophagous and many species show high degrees of specialization on a small number of host plant species (stenophagy), with numerous species being monophagous (i.e. breed and feed on only one host species) (e.g. Schwartz, 2005; Schuh, 2008). Because these stenophagous habits have long been known, targeted plant bug field work during the past 50 years has also resulted in a substantial amount of vouchered host plant information associated with plant bug collection events. Worldwide and in North America, more than 43% of plant bug specimen records in the AEC database are therefore connected to host plant information. When overlaying plant bug and host plant species distributions, it also becomes evident that many of the stenophagous and monophagous plant bug species tend to have smaller distribution ranges than their host plants (Fig. 1).

In light of the above observations, the present study has two main objectives. The first is to identify areas of endemism for plant bugs based on the parameters ("default dataset") used in Escalante et al. (2013) that focused on Nearctic mammals and compare these two sets of hypotheses. Given the narrow distribution ranges of many species of Miridae, we predict that areas of endemism are smaller and more numerous for Miridae in North America, and especially in the Western United States, than they are for mammals. The second objective is to explore hypotheses on areas of endemism based on plant bug distributions in the context of areas identified by their host plant species. Distribution ranges of Miridae are often smaller than those of their host plants and we predict that areas of endemism defined by Miridae will also be smaller than those identified by their host plants. To more thoroughly utilize and analyse available plant bug records and the NDM methodology, we also investigate the effects of different minimum numbers of records per species, a smaller grid-cell size, and different levels of strictness in defining areas of endemism on number and composition of resulting areas of endemism.

Material and methods

Data sets and data cleaning

Miridae specimen records for Canada, the United States and Mexico were downloaded from the AEC database. The R and PHP software, and resulting MySQL database, for acquiring and cleaning specimen data from AEC and iDigBio is available on GitHub (Seltmann, 2016a,b; https://dx.doi.org/10.5281/zenodo.44387). We limited the Miridae data set to include only the four largest subfamilies of Miridae (Deraeocorinae, Mirinae, Orthotylinae and Phylinae) that represent the majority of specimen records in the AEC database (287 432 in the Nearctic), are reliably identified, and have better-understood host associations than the four smaller and less studied subfamilies. Three data sets based on different minimum numbers of unique locality records per species were created: the largest data set included 1566 species and 61 784 unique records, with all included species having three or more records. The default data set—the data set most closely matching the data set criteria used by Escalante et al. (2013) for mammals—included 1339 species with five or more unique localities, resulting in 61 016 records. All unique localities included in this default Miridae data set are presented as a map (Fig. 2a) that provides evidence for broad plant bug sampling across the United States as well as Mexico, but also highlights potentially undersampled areas, such as Baja California Sur, northern Mexico, and the upper Great Plains. The smallest and most restricted data set included the 1004 species for which ten or more unique records were available and comprised 58 820 records. Localities, represented by latitude and longitude geographical coordinate data, were rounded to three decimal places prior to comparison for uniqueness. Miridae species distribution maps and data are available through figshare (https://dx.doi.org/10.6084/m9.figshare.2059506) (Seltmann et al., 2016b).

We generated a list of plant species that have been documented as host for any of the Miridae included in the default (1339 species) data set. Plant hosts were obtained from AEC and are direct observations recorded on labels associated with the plant bug specimens, with the requirement that the plant be listed as a host for more than two independent specimen collecting events, with an overall frequency greater than 15%. A collecting event is defined as the combination of collecting date, location and host plant. Recorded plant host names were checked against the iPlant Taxon Name Resolution Service (Boyle et al., 2013; Taxonomic Name Resolution Service, 2015) for name status and

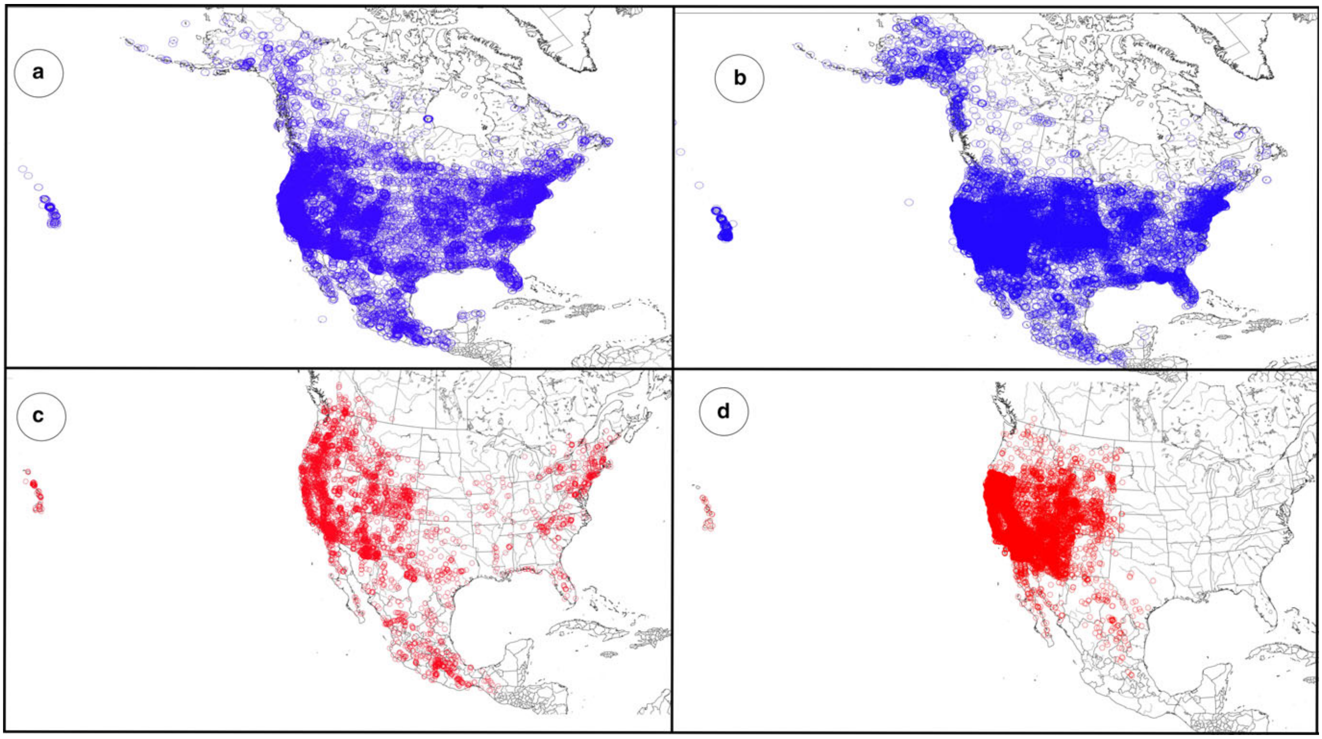

Fig. 2. Maps illustrating all unique localities included in the analysis, documenting dense sampling and localities of scoring species: (a) all unique localities for the default Miridae data set; (b) all unique localities for the plant data set; (c) unique localities of scoring plant bug species; and (d) unique localities of scoring plant species. [Colour figure can be viewed at wileyonlinelibrary.com].

synonyms were merged in the analysed plant data set. Plant specimen records were downloaded from iDigBio using ridigbio (Michonneau et al., 2015), an extension to the iDigBio search Application Program Interface. As additional quality control, we checked if the recorded latitude and longitude mapped to the state reported in the database and discarded ~1200 specimen records that represented specimens planted in botanic gardens. After cleaning, the data set was limited to plant species represented by five or more unique plant localities, conforming to the default insect AOE analysis, and resulting in 313 plant species represented by 196 012 unique records. The unique localities included in this plant data set are shown in Fig. 2b, indicating overall dense sampling of plant localities with comparatively less comprehensive sampling in Mexico, Texas and the eastern half of the United States, except Florida and Northeastern USA. Host plant data and species distribution maps are available through figshare (https://dx.doi.org/10.6084/m9.figshare.2059503) (Seltmann et al., 2016a).

Analyses

The three Miridae data sets representing 1004, 1339 and 1566 species and the single plant data set (313 species) files were loaded into ndm/vndm v.3.0 (Goloboff, 2011). The following parameters were kept the same across the six Miridae and single host plant analyses: save number of areas was set to 10 000; searches were repeated ten times; the percentage of overlapping subsets was set to 0, the retention of suboptimal sets to 0, the random seed to 1, save sets with scores above 2.0, and the consensus option to “any” and 40%.

In the default analysis, we defined an area of endemism according to Nelson and Platnick (1981) and Platnick (1991) and as an area that has a minimum of two species with congruent distributions. We also performed two more restrictive analyses based on the largest data set (1566 species) that required a minimum of five and ten, respectively, congruent species distributions to qualify as an area of endemism (Table 1).

In addition to the default grid size of 2 degrees, a separate analysis was conducted with a grid size of 1 degree (Table 1). Trial runs with a grid size of 0.5 were discarded, because our data set does not have a sufficient number of specimen records for analyses at such fine scales, resulting in a small number of candidate (12) and consensus (seven) areas that are almost exclusively restricted to the densely sampled areas around the Southwest Research Station in Portal, Arizona, and certain areas in Southern California.

Tables 2 and 3 report the number of candidate areas for each consensus area for the default plant bug and

Table 1
Summary of Miridae and host plant analyses that differed with respect to the total numbers of included species and unique records, minimum number of unique localities per species, grid size and number of congruent species required for an area to qualify as endemic. Reported are numbers of examined and saved areas, sets read, number of distinct sets (with duplicate sets discarded) and the resulting number of consensus areas

Taxon	Analysis type	No. of species	No. of records	Min. no. of of unique localities	Grid size	Area of endemism spp. No.	No. of areas examined	No. of areas saved	No. of sets read	No. of sets	No. of consensus areas
Miridae	Default dataset/analysis	1339	61 016	5	2	2	1 065 975	519	4594	1121	45
	Smaller min. no. of records (3)	1566	61 784	3	2	2	1 084 465	487	4805	1149	55
	Larger min. no. of records (10)	1004	58 820	10	2	2	912 269	237	2328	605	33
	Larger no. of spp./area (5)	1566	61 784	3	2	5	939 980	331	3316	710	21
	Larger no. of spp./area (10)	1566	61 784	3	2	10	887 939	234	2259	365	12
	Smaller grid size	1339	61 016	5	1	2	1 453 656	70	788	185	19
Host plants	Default plant dataset	313	196 012	5	2	2	330 090	50	534	106	10

plant analyses, the number of species contributing to ("scoring") in each of the areas, as well as minimum and maximum endemicity scores for each consensus area. In the results section, we refer mostly to the highest endemicity scores ("maximum score") for a given consensus area.

Visualization of areas of endemism

ndm/vndm output files do not organize consensus area geographically, making it difficult to compare consensus areas between analyses. We therefore renumbered consensus areas derived from the analyses and organized them by geographical region, starting with large areas in Western North America and working South and East, with Hawaii listed last. ndm/vndm allows for the export of different types of files describing the consensus areas that can then be used to construct maps in DivaGIS 7.5.0 (http://www.diva-gis.org/). We exported both grid text files that describe the outline of the grid cells of each consensus area and species text files that list the coordinates of all species in support of that area. Latitude and longitude columns were reversed in the grid files to prepare the final maps. A base map showing country and state boundaries as well as elevational shading was imported into DivaGIS and all grid files for a given analysis then imported into that project ("Data > Import text to Line/Polygon"). Maps were exported as tiff files and compiled into plates in CorelDraw 13.0.0.667.

Results

Endemicity analysis of the different Miridae data sets resulted in between 12 and 55 consensus areas of endemism, with the default data set recovering 45 areas (see Tables 1 and 2). Using Area 16 (Southern California including Baja California) as an example, Figure 3 illustrates the connection between the five candidate areas resulting in one consensus area as seen in VNDM/NDM (Fig. 3a), the 13 scoring species for one of the candidate areas (Fig. 3b), and localities for the 16 scoring species across all candidate areas with the proposed consensus area of endemism superimposed (Fig. 3c,d). Sixteen representative areas of endemism chosen because of their high maximum endemicity scores and large numbers of candidate areas are shown in Figure 4; maps illustrating all 45 areas are provided as Fig. S1. Consensus areas in the default data set were based on between 1 and 770 candidate areas, with between 1 ($n = 14$) and 184 (Area 1; Western North America) scoring species per area (~13.67 species on average; see Table S1 for a list of scoring species per area). A total of 610 species out of the 1339 species included in the default data set

Table 2
The 45 consensus areas of endemism based on the default Miridae data set, reporting the area identification number (AOE#), number of areas included in the consensus, number of scoring species, minimum and maximum endemicity scores and a brief geographical description of the area

AOE#	No. of areas	No. of species	Min score	Max score	Area description
1	770	184	2.03	34.77	Western North America
2	46	30	2.1	4.81	Western North America
3	22	25	2.8	5.29	Western North America, excl. southern parts
4	85	46	2.21	7.79	Western North America, excl. northern parts
5	1	3	2.33	2.58	Western North America, excl. warm deserts in Arizona and New Mexico
6	45	6	2.13	2.64	Southwestern United States
7	3	7	2.62	3.12	Western North America, interior (expanded Great Basin)
8	4	7	2.39	2.64	Intermontane distributions, incl. Southwestern United States
9	47	22	2.28	7.47	Pacific Northwest and Northern California
10	3	6	2.83	3.58	Marine West Coast Forest and Western Cordillera
11	2	5	3.04	3.29	Montane California (Sierra Nevada, Coast Range, parts of Great Basin)
12	2	5	3.06	3.31	Northern Rocky Mountains and Northern Great Basin
13	9	15	2.71	4.97	Southwestern warm deserts and Mediterranean California
14	2	5	2.33	2.83	Sky Island and Sierra Nevada (disjunct)
15	1	9	5.15	5.41	Pacific Northwest and Northern California, without wet areas
16	5	16	5.01	8.82	Southern California incl. Baja California
17	2	4	2.37	2.62	Intermontane distributions (disjunct)
18	1	4	2.76	3.01	Intermontane distributions, mostly southern parts
19	2	5	2.19	2.94	Intermontane distributions along United States-Mexico border
20	45	7	2.52	4.02	Intermontane distributions, between Mexico and Wyoming
21	1	4	2.85	3.1	Intermontane distributions (disjunct), Southwestern United States
22	5	16	2.15	4.11	Intermontane distributions, Southern Cordillera
23	2	5	2.06	2.56	Eastern Great Basin
24	3	4	2.11	2.61	Intermontane distributions (disjunct)
25	1	4	2.49	2.74	Front Range of the Rocky Mountains
26	3	9	2.73	4.48	Southern half of South Central Semi-arid Prairies
27	1	3	2.18	2.43	Southern Plains
28	1	4	2.6	2.85	Southwestern deserts (disjunct)
29	1	3	2.1	2.35	Sonoran Desert and Baja Californian Desert
30	1	3	2.1	2.35	Mexican Plateau (disjunct)
31	3	7	3.56	4.06	Mexican Plateau (disjunct)
32	4	7	2.1	3.6	Mexican Plateau
33	19	28	3.26	5.92	Montane Mexico: highlands, temperate sierras, and tropical dry forest
34	1	3	2.16	2.41	Montane Mexico: highlands, temperate sierras, and tropical dry forest
35	2	10	3.12	5.56	Transmexican Volcanic Belt
36	19	22	4.38	13.4	Transmexican Volcanic Belt
37	1	2	2	2.25	Western Pacific Coastal Plains, Hills and Canyons
38	1	3	2.19	2.44	Eastern North America, excl. Florida (disjunct)
39	14	22	2.1	5.14	Northeastern United States
40	7	8	3.19	4.44	Appalachian
41	11	15	2.24	3.49	Appalachian and Southeastern Plains
42	1	5	2.99	3.24	Eastern North America, excl. Florida (disjunct)
43	2	4	2.08	2.33	Eastern North America, excl. southern parts (disjunct)
44	1	3	2.12	2.37	Florida (Southeast Coastal Plain)
45	2	10	3.5	7.3	Hawaii

contributed to the score of one or several areas of endemism. Unique localities of scoring species are provided as Fig. 2c, documenting that most of the scoring species have ranges in the the Western Nearctic and Mexico, but to a lesser degree in the Eastern United States including Florida. Sizes of areas of endemism in the default Miridae data set ranged from 2 (Area 37; Western Pacific Coastal Plains in Mexico) to 81 (Area 2; Western North America) grid cells, with an average of 22.9 (median of 18.5). Thirty-five of the 45 areas are located entirely or mostly in the Nearctic north of Mexico, one is found in Hawaii, and nine areas are located entirely or almost entirely in Mexico. With a maximum endemicity score of 34.77, a large area of endemism that we here refer to as Western North America (Area 1; Fig. 4a) is the highest scoring area, followed by areas in the Transmexican Volcanic Belt (Area 36; maximum score 13.4; Fig. 4b), Southern California including Baja California (Area 16; maximum score 8.82), the Southwestern U.S. including Baja California and northern Mexico (Area 4; maximum score 7.79; Fig. 4B), an area in the Pacific Northwest (Area 9; maximum score 7.47; Fig. 4e) and Hawaii (Area 45; maximum score 7.3). The highest

Table 3
The ten consensus areas of endemism based on the plant data set, reporting the area identification number (AOE#), number of areas included in the consensus, number of scoring species, minimum and maximum endemicity scores and a brief geographical description or eco-region name of the area

AOE #	No. of areas	No. of species	Min score	Max score	Area description
1	2	4	2.08	2.33	Western North America, excl. northern parts
2	71	18	2.18	14.35	Western North America, excl. northern parts, incl. Baja California
3	23	16	2.46	5.43	Western North America, incl. intermontane areas and part of plains
4	3	5	2.15	2.9	Western North America
5	1	3	2.2	2.45	Southwestern United States, Baja California, Mexican Plateau
6	1	3	2.2	2.45	Western North America, excl. northern parts
7	2	11	7.24	7.49	California
8	1	4	3	3.25	California and Pacific Northwest
9	1	4	3	3.25	Baja California Norte
10	1	2	2	2.25	Hawaii

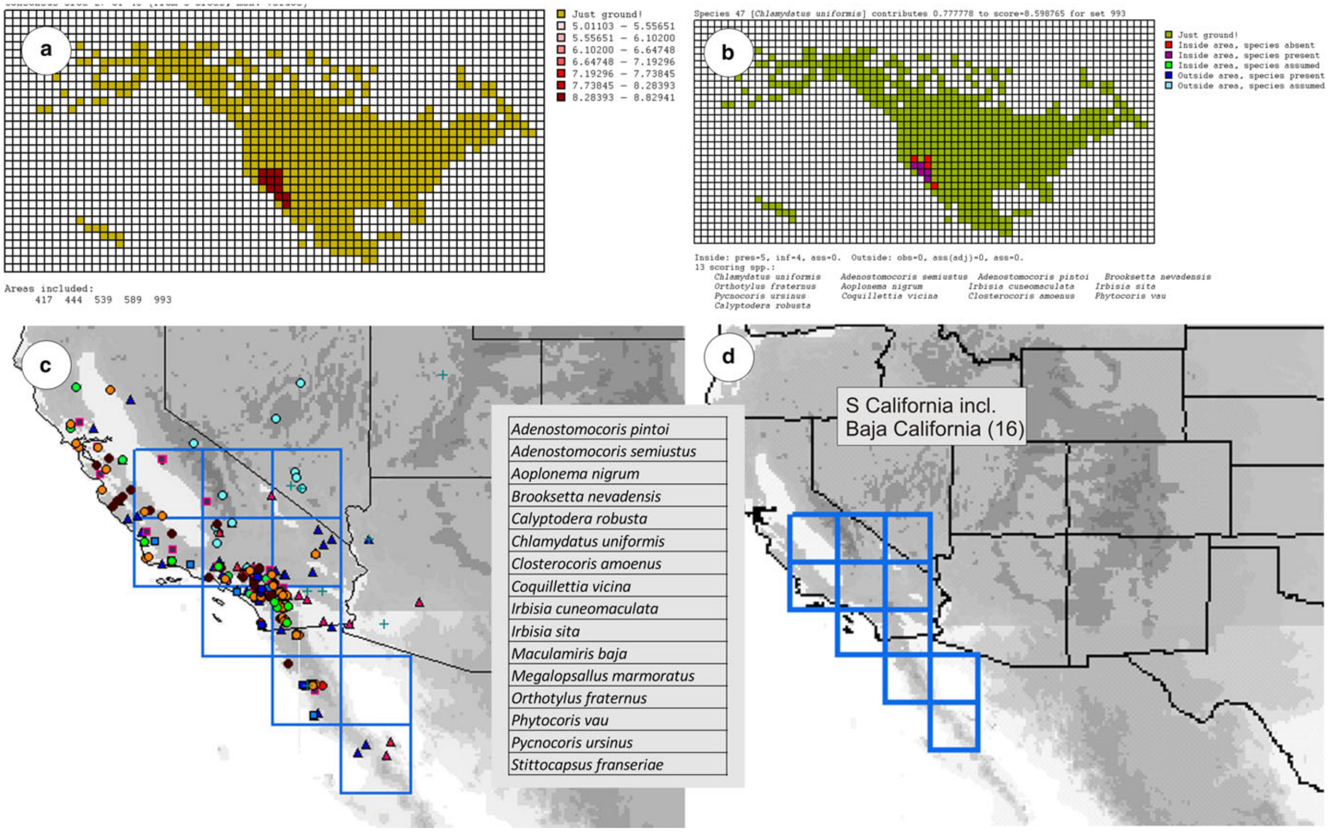

Fig. 3. Candidate areas, scoring species and their distribution ranges for consensus Area 16, Southern California and Baja California. (a) Five candidate areas resulting in one consensus area as seen in the program VNDM/NDM; (b) the composition of scoring species for one of the candidate areas; (c) and (d), localities for the 16 scoring species plotted using DIVA-GIS with the proposed consensus area of endemism superimposed. [Colour figure can be viewed at wileyonlinelibrary.com].

scoring area in Eastern North America is Area 39 (maximum score 5.14; Fig. 4b) located in the Northeastern United States. Overall, eight of the areas of endemism recovered in this analysis are located in the Eastern United States (areas 38–44), with the remainder located in the Western United States, Mexico and Hawaii. Even though a number of areas are unique to a given geographical space (e.g. Hawaii (45), Florida (44), Front Range of the Rocky Mountains (25; Fig. 4D), other areas broadly overlap with each other. Among the examples are two Appalachian areas—40 (23 grid cells) and 41 (26)—that differ because Area 41 includes several additional grid cells located south of the southern Appalachians, but also large areas of

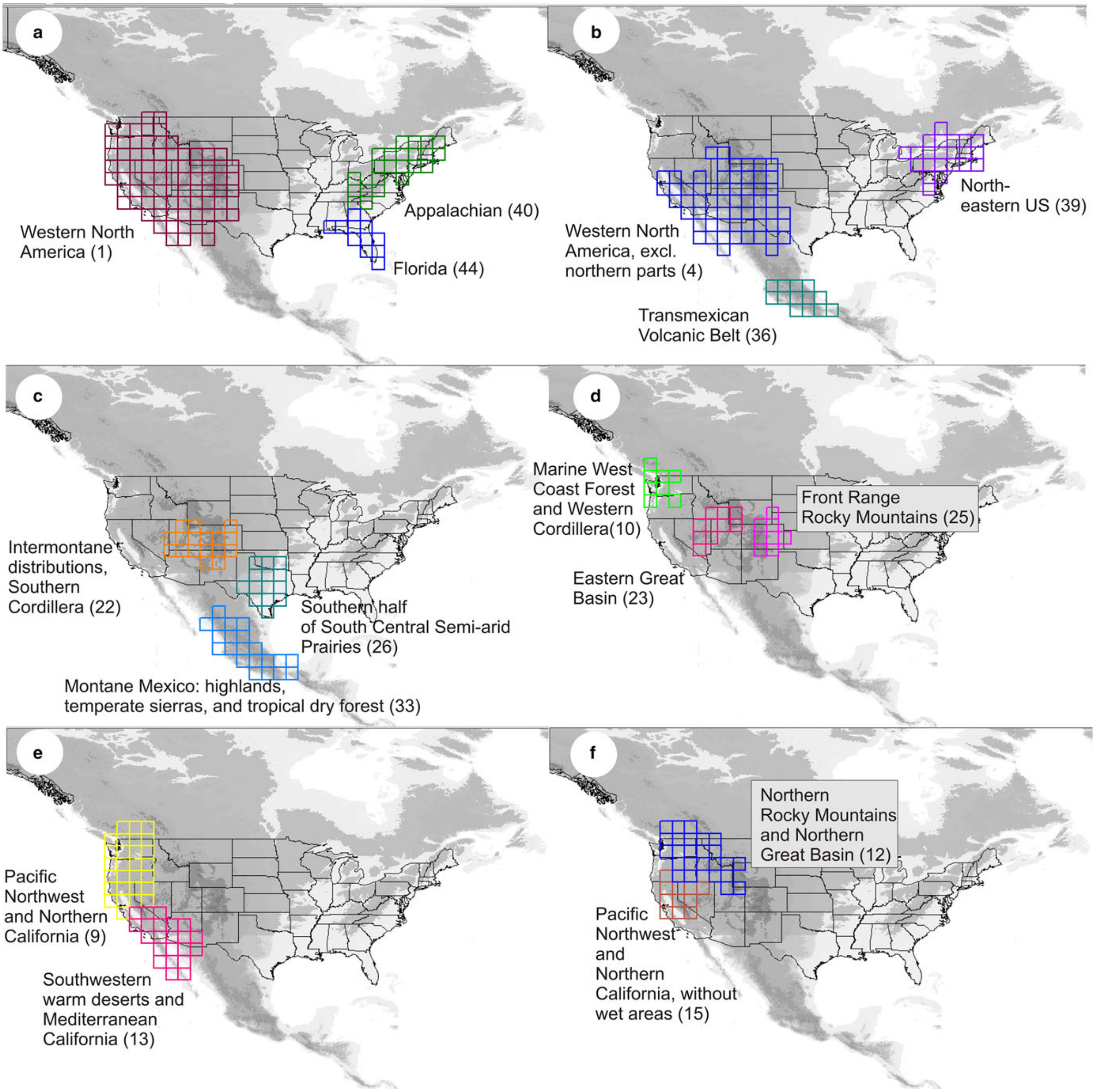

Fig. 4. Sixteen-area subset of the 45 consensus areas based on the Miridae default data set, showing major areas of endemism recovered across the Nearctic region, with emphasis on high-scoring consensus areas or areas representing a substantial number of candidate areas and scoring species. (a) Western North America (1), Appalachian (40) and Florida (Southeast Costal Plains; 44). (b) Western North America, excluding northern parts (4), Transmexican Volcanic Belt (36) and Northeastern United States (39); (c) Intermontane distributions, Southern Cordillera (22), Southern half of South Central Semi-arid Prairies (26), and Montane Mexico: highlands, temperate sierras and tropical dry forest (33); (d) Marine West Coast Forest and Western Cordillera (10), Eastern Great Basin (23) and Front Range of the Rocky Mountains (25); (e) Pacific Northwest and Northern California (9) and Southwestern warm deserts and Mediterranean California (13); (f) Northern Rocky Mountains and Northern Great Basin (12) and Pacific Northwest and Northern California and adjacent Nevada, without wet areas (15). [Colour figure can be viewed at wileyonlinelibrary.com].

endemism in the Western United States that differ only by a small number of grid cells (e.g. areas 1–5).

Endemicity analysis of 313 associated plant species with a sufficient number of unique localities (five) resulted in 185 candidate and ten consensus areas, four of which are shown in Fig. 5 (see Fig. S2 for maps of all areas). Consensus areas were derived from between 1 and 71 candidate areas, with between two (e.g. Area

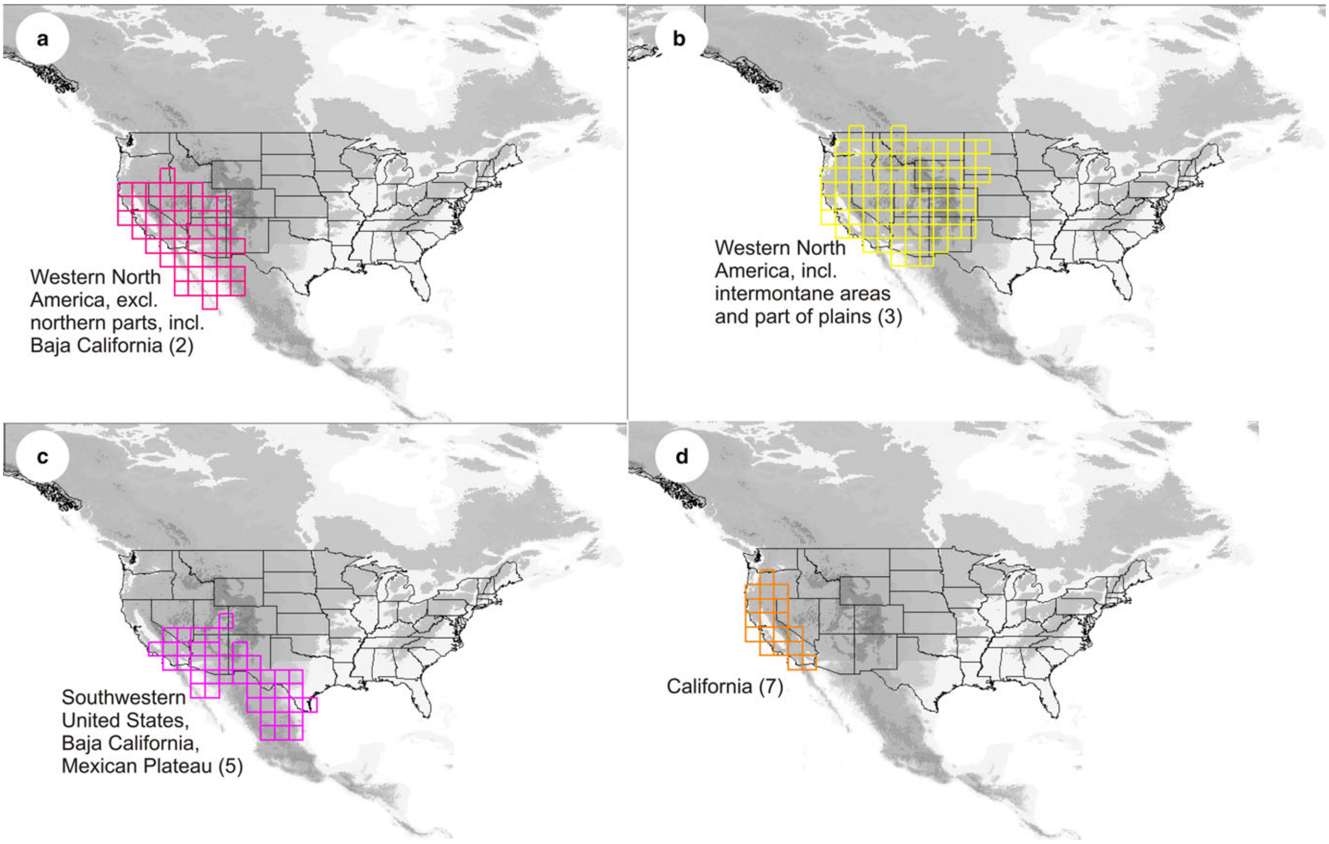

Fig. 5. Four-area subset of the ten consensus areas based on the plant data set. (a) Western North America, excluding northern parts, including Baja California (2); (b) Western North America, including intermontane areas and part of plains (3); (c) Southwestern United States, Baja California, Mexican Plateau (5); (d) California (7). [Colour figure can be viewed at wileyonlinelibrary.com].

10) and 18 (Area 2; Western North America) scoring species per area (seven species on average; see Table 3 for lists of species). Of the 313 plant species included in this analysis, only 59 scored in one or several of the consensus areas. Unique localities of scoring species are provided as Fig. 2D, showing almost complete restriction to the Western United States and Mexico, with few localities in Mexico and none in the Eastern United States. Sizes of areas of endemism ranged from 3 (Area 10; Hawaii) to 87 (Area 3: Western North America, including intermontane areas and part of the high plains) grid cells with an average of 29.3 cells per area (median of 25.5). Nine of the areas are located in the Western Nearctic and one comprises most of Hawaii (Area 10); areas in the Eastern United States are notably absent from this analysis. Three of the Western Nearctic areas cover most (Area 3) or large parts (e.g. Area 4) of that region and three comprise parts of the Southwestern United States and parts of Mexico (e.g. areas 1 and 6). Two areas roughly correspond to the California Floristic Province (areas 7 and 8) and a small area straddles the border between California and Baja California Norte (Area 9). Maximum endemicity scores across all consensus areas recovered from this analysis range between 2.25 and 14.35 (Area 2 (Western North America) has highest maximum score), with the only additional maximum scores above 5 being recovered by California (Area 7; maximum score 7.49) and the largest recovered consensus area, the Western Nearctic (Area 3; maximum score of 5.43).

Compared to the default Miridae data set and analysis, the five secondary plant bug data sets and analyses (Table 1) vary widely in the number, but not as much in the composition of the recovered consensus areas of endemism (Table 1). The data set with the largest number of included plant bug species (1566) and a smaller minimum of unique localities resulted in 55 consensus areas (Fig. S3), of which most are broadly overlapping with areas also identified in the default data set. However, some areas are unique to this data set, most notably areas 34 and 35 that comprise only the southern part of Texas and adjacent areas in Mexico, but also Area 11 that is located in Alaska and the small Pacific Northwest areas 12 and 13 that comprise Vancouver Island and the Olympic peninsula, as well as some inland areas. Maximum scores across all consensus areas in this analysis range from 2.12 to 36.02

(Area 1, Western North America), with most of the high scoring areas also being recovered in the default analysis.

The data set that included only 1004 species with ten as the minimum number of unique localities resulted in 33 consensus areas of endemism (Table 1, Fig. S4). The distribution of areas roughly corresponds to the default and 1556 species data set analyses in recovering the greatest number of areas in the Western Nearctic including Mexico, with a few areas, including Florida, found in the East. Some areas are virtually identical between the 1004 spp. and default data sets (e.g. Area 2 (default) and Area 2 (1004 spp.)), but others are unique. As an example, Area 5 comprises the California Floristic province that was also recovered in the plant, but not the default mirid, analysis. Maximum endemicity scores across all consensus areas recovered in this analysis ranged between 2.32 and 22.65 (Area 3 (Western North America) is the highest scoring area with 22.65), with again most of the higher scoring areas also largely recovered in the default and 1556 spp. analyses. The averages of maximum endemicity scores (across all consensus areas in a given analysis) were very similar between these three analyses (i.e. 4.74 for the default analysis, 4.59 for the 1556 spp. analysis and 4.33 for the 1004 spp. analysis).

The two analyses that defined areas of endemism as possessing five and ten congruent species per area, respectively, resulted in fewer areas of endemism, but much higher average maximum scores compared to the default analysis. The first analysis (five species/area) resulted in 21 consensus areas, many of which are again congruent with areas recovered in the default analysis (Fig. S5). Maximum scores range between 3.56 and 36.02 (Area 1 (Western North America) is the highest scoring area with 36.02), with an average of 7.67, and 11 of the areas receiving a maximum score of 5 and higher. Other than Florida, only two additional areas were found in Eastern North America (areas 18 and 19). Several areas were recovered along the Mexican Plateau and Transvolcanic Belt, the Southern half of the South Central Semi-arid Prairies and subsets of the Western Nearctic. This subset of the Western Nearctic also includes two areas in the Pacific Northwest, similar to areas also recovered in the 1556 spp. analysis. The second analysis that used a cut-off of ten congruent species to define areas of endemism, resulted in only 12 consensus areas, with maximum scores ranging from 6.93 to 36.02 (Area 1 (California) is the highest scoring area with 36.02) and an average of 13.29 (Fig. S6). A comprehensive Western North America area as seen in most other analyses was not recovered. Instead, subsets of this area, including California (Area 1) or Southwestern deserts including Sonoran desert and California (Area 3), receive high endemicity scores, with additional areas being restricted to the Mexican Plateau and Hawaii.

The final data set (“smaller grid size”) explored the effect of selecting a smaller grid size for the default Miridae data set and analysis, and resulted in 19 consensus areas (Fig. S7). Compared to the default data set, these areas are typically small and all are restricted to the Western Nearctic including Mexico and Hawaii. Several areas are disjunct, including areas 4 (California), 6 (Pacific Northwest), and 10 and 11 (Southeast corner of Arizona = Southwest Research Station and small areas in Texas and New Mexico). Maximum endemicity scores range between 2.25 and 10.47 (Area 7 is the highest scoring area with 10.47), with an average of 4.07.

Discussion

Comparing consensus areas of endemism derived from our default Miridae data set and analysis with areas recovered by Escalante et al. (2013), we find several differences and similarities. First, Escalante et al. (2013) recovered 21 consensus areas located in North America north of Mexico, whereas the default Miridae data set resulted in a higher number (i.e. 35) of such areas. This finding corroborates our prediction that the smaller distribution ranges observed in North American Miridae could result in a larger number of areas of endemism. Second, several consensus areas broadly overlap between the plant bug and the mammal data sets, six of which are illustrated in Figure 6 and comprise an expanded Great Basin, Tamaulipas, Florida, the Pacific Northwest and Northern California, Southern California including Baja California, and Eastern North America. Some but not all of these shared consensus areas received high maximum scores in the Miridae data set. This observation shows that certain areas of endemism in North America, but not necessarily those with the highest scores, are identified by groups of organisms as drastically different as insects and mammals. Third, and despite these shared areas, the Miridae data set resulted in a substantial number of consensus areas without counterpart in the mammal data set. Among these are high scoring areas such as 22 (intermontane distributions, Southern Cordillera; maximum score 4.11), 26 (southern half of South Central Semi-arid Prairies; maximum score 4.48) and 39 (Northeastern United States; maximum score 5.14), but also numerous lower scoring areas, especially in the Southwestern United States (e.g. areas 18, 20, 23). Many of these areas are relatively small and more finely partition the larger consensus areas in the Western Nearctic, such as areas 1–5. Depending on the target area and research question, biogeographical analyses will use areas of endemism of vastly

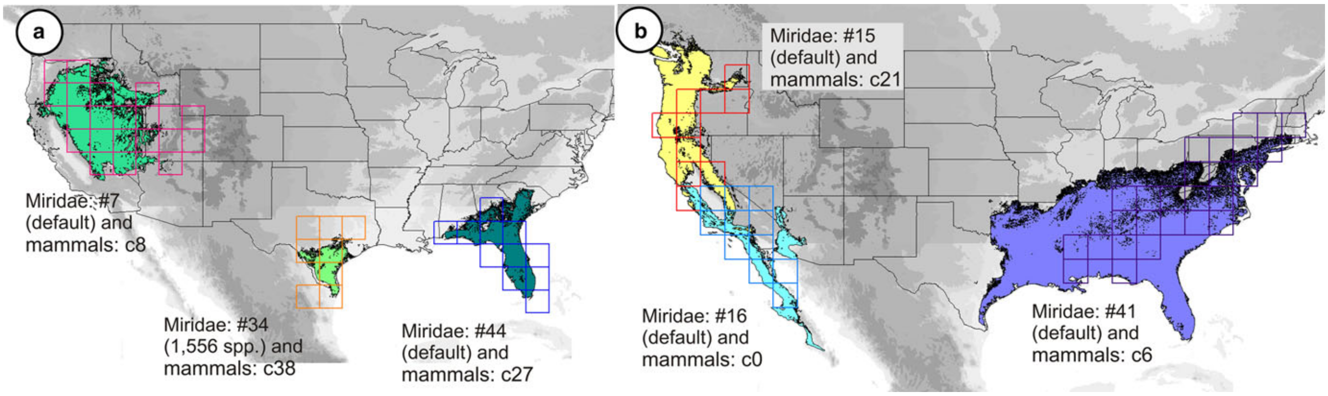

Fig. 6. Five consensus areas (c8, c38,c27, c21 and c0) and one region (c6) identified by Escalante et al. (2013) based on mammal distributions (shaded areas) superimposed with six consensus areas recovered from the default Miridae data set (grids), showing broad congruence with respect to area of endemism hypotheses. [Colour figure can be viewed at wileyonlinelibrary.com].

different size: for example, to analyse relationships across the Holarctic, identifying broad areas of endemism such as Western Nearctic and Eastern Nearctic may be appropriate (Enghoff, 1995; Sanmartin et al., 2001), but not for analyses that could explain the biogeographical history of Western Nearctic biota. We argue that the unique areas of endemism identified in our Miridae analyses provide a good starting point for investigations into the historical biogeography of North America at a finer scale and especially within the Western Nearctic, but could also have implications for conservation studies that build on such endemicity analyses.

Even though the mammal data set appears to be somewhat less well suited to inform areas of endemism in North America north of Mexico in comparison to the Miridae data set, it is clearly superior in Central America and the Mexican Transition Zone. Numerous consensus areas in these regions were identified by mammal distributions, including some small, but widely recognized, areas such as the Sierra Madre del Sur (Liebherr, 1994; Escalante et al., 2013; their fig. 4). This strength of the mammal data set in that region is not surprising given the diversity and often small distribution ranges of mammals in the Neotropical region compared to North America (Wilson and Reeder, 2005). However, similar to the observations above, several larger consensus areas are shared between the Miridae and mammal analyses (e.g. Mexican Plateau (Miridae Area 32); Transmexican Volcanic Belt (Area 36)), further strengthening areas of endemism that have been long recognized based on a wide variety of taxa (Marshall and Liebherr, 2000; Morrone, 2014b). The lack of specimen data for Miridae south of Mexico prevented us from including these areas in our analyses, while at the same time recognizing that understanding the division between the Nearctic and the Neotropics was one of the goals of an earlier analysis focusing on mammals (Escalante et al., 2010). The taxonomy and distribution of Central American Miridae are much less comprehensively documented than those of Nearctic taxa, and we expect that assembling a data set comparable to our North America-focused data set will be many years in the making. Furthermore, it is no accident that knowledge of the Central American fauna lags behind that of Mexico and North America, because revisionary work—focused primarily on the Nearctic—has shown that in most groups of Miridae there is little similarity in the faunas of the two areas (Schuh and Schwartz, 1985, 1988).

When comparing areas of endemism defined by Miridae to areas identified by their host plants, we found striking differences. Areas of endemism recovered by the host plant data set are entirely restricted to the Western Nearctic and most areas are large (Fig. 5). Both the average size of areas of endemism (as determined by numbers of grid cells) and the areas' absolute sizes are larger for the host plants than for the Miridae. These findings corroborate our prediction that given the smaller distribution ranges of many species of Miridae compared to their host plants, areas of endemism identified by Miridae should also be smaller than those defined by these host plants. However, this result does not suggest that plants in general are less suitable for the prediction of areas of endemism compared to other groups of organisms. The selection of plant species used in our data set is rather restrictive in being solely governed by a plant's association with one of the Miridae species included in the default data set and a minimum of five unique localities. An inclusive list of North American plant species, or at least a larger number or a different subset of these species, would likely result in a very different picture. In addition, and despite the fact that unique localities for plants appear to cover the entire range (Fig. 2b), plant records were likely to be insufficient in certain areas. This is almost certainly true for Mexico, where even the unique locality plot is quite sparse, but also in the

Eastern United States, where our plant data set failed to identify any area of endemism. We concluded that our host plant data set is of limited value in identifying areas of endemism in North America beyond broader areas also recovered in the Miridae data set, such as Western North America and California, and for comparing the distributions of the bugs and their hosts.

Different Miridae data sets and parameters used in the analyses allowed us to further explore the strengths and weaknesses of the Miridae data set (Table 1). The larger, but less densely sampled data set (1566 spp.; 61 784 unique records; minimum of three records per species) resulted in a larger number of consensus areas, whereas the smaller, more restrictive data set (1004 spp.; 58 820 unique records; minimum of ten records per species) in fewer areas. A large number of areas are almost identical (e.g. Figs S1, S3 and S4: Southern Plains (Area 27 in default data set)) or at least broadly congruent between these three data sets. However, other areas are unique to one of the three data sets and will require further exploration, especially those areas that have high or fairly high endemicity scores. Defining an “area of endemism” more strictly by requiring five or ten, respectively, congruent species had a more severe impact on analyses. The two more restrictive analyses resulted in much smaller numbers of areas, but on average much higher endemicity scores (see Figs S5 and S6). Areas are also smaller and most (Fig. S5) if not all (Fig. S6) are restricted to the Western Nearctic. Some of the areas are congruent with those found in the default data set (e.g. Western North America and California), but others were not recovered in the default data set, for example, two very small areas in the Pacific Northwest that (areas 5 and 6 in the “five endemic species” analysis) that are subsets of Area 10 of the default data set. Additional smaller areas of endemism in these two data sets are found in the Southwestern United States (e.g. areas 9 and 4 in the five and ten spp. analyses, respectively) and could be of value when analysing biogeographical relationships in the Western Nearctic. A drawback of these analyses is the absence of consensus areas in the Eastern half of the United States that might be due to the lack of unique locality records or to the larger number of more widespread species in that region (see Figs S5 and S6). The final data set used 1-degree instead of 2-degree grid cells, resulting in an analysis that has the potential to recover areas of endemism on a finer scale. Some of the areas are congruent with areas in the default data set and indeed provide a somewhat more refined boundary to areas such as California (e.g. Fig. S7, areas 2 (California except Coastal ranges) and 3 (all of California)). However, several areas are very small and others are highly disjunct (e.g. Fig. S7, areas 10 and 11), indicating that our data set does not include sufficient unique localities in order to perform well at this smaller grid size.

In summary, the first large insect data set with focus on the Nearctic region and one of few data sets applying endemicity analysis to North American taxa in general, has demonstrated the strengths, but also limitations, of insects as model subjects for endemicity analysis in this region. The development of a strong taxonomic foundation and considerable databasing efforts during the past three decades have allowed us to analyse a comprehensive and clean Miridae data set comprising a larger number of species and unique localities than was heretofore imaginable, with especially great species diversity and restricted distributional ranges in the Western United States and the Mexican Transition Zone. Because Miridae contain numerous species with small endemic distribution ranges, a substantial number of species have the potential to score in endemicity analyses. We showed that this expectation was met in our analysis, with close to half of the included plant bug species contributing to the score of one or several areas. The number of consensus areas of endemism resulting from our analyses was reasonable, based on a variety of criteria. Although some areas were broadly overlapping, many were unique. We observed a wide range of maximum endemicity scores, but recovered a substantial number of high scoring areas, especially in the Western Nearctic and Mexico, but also in Florida. Even though areas recognized among the different Miridae analyses were not identical, we observed considerable congruence among areas derived from different analyses, and we interpret this as an indication that proposed consensus areas are fairly robust with respect to data set and analytic parameters. These consensus areas could also be tested further by analyses employing higher taxa (e.g. genera) as opposed to species, as outlined by Szumik and Goloboff (2015). Increased sampling efforts in certain areas of the United States, but also in Mexico (e.g. along the U.S.–Mexico border and in Baja California Sur) would strengthen and refine our analyses, but we suspect that a substantial amount of additional data evenly sampled across the entire region would be required to allow for analyses using smaller grid sizes, and therefore resulting in more finely delimited areas of endemism.

Acknowledgements

Support for this study was provided by the NSF-funded EF#1115144, the Tri-Trophic Database Thematic Collection Network that is part of the Advancing the Digitization of Biological Collections program. We thank numerous collection managers, undergraduate students and other specimen data entry personnel for their assistance with specimen data capture and georeferencing. We also acknowledge all other principal and co-principal investigators as well as collaborators on this project, especially colleagues who participated in the NSF-sponsored data mining workshop held at the University of California, Riverside in 2014. Discussions with all participants and especially with Neil Cobb (Northern Arizona University), Jorge Soberón (University of Kansas) and Jacob Christian Cooper (University of Kansas) during this workshop majorly shaped the direction of this project.

References

Aagesen, L., Bena, M.J., Nomdedeu, S., Panizza, A., López, R.P., Zuloaga, F.O., 2012. Areas of endemism in the southern central Andes. Darwiniana, n.s. 50, 218–251.

Asquith, A., 1991. Revision of the Genus *Lopidea* in America North of Mexico (Heteroptera: Miridae: Orthotylinae). Koeltz Scientific Books, Königstein

Boyle, B., Hopkins, N., Lu, Z., Raygoza Garay, J.A., Mozzherin, D., Rees, T., Matasci, N., Narro, M.L., Piel, W.H., Mckay, S.J., Lowry, S., Freeland, C., Peet, R.K., Enquist, B.J. 2013. The Taxonomic Name Resolution Service: an online tool for automated standardization of plant names. BMC Bioinformatics 14, 16.

Casagranda, M.D., Taher, L., Szumik, C.A., 2012. Endemicity analysis, parsimony and biotic elements: a formal comparison using hypothetical distributions. Cladistics 28, 645–654.

Cassis, G., Schuh, R.T., 2012. Systematics, biodiversity, biogeography, and host associations of the Miridae (Insecta: Hemiptera: Heteroptera: Cimicomorpha). Ann. Rev. Entomol. 57, 377–404.

Del Ventura, F., Liria, J., Navarro, J.-C., 2013. Determinación de áreas de endemismo en mosquitos (Diptera: Culicidae) en Venezuela, mediante criterios explícitos de optimización. Bol. Malariol. Sal. Amb. 53, 165–182.

Domínguez, M., Roig-Juñent, S., Tassin, J.J., Ocampo, F.C., Flores, G.E., 2006. Areas of endemism of the Patagonian steppe: an approach based on insect distributional patterns using endemicity analysis. J. Biogeog. 33, 1527–1537.

Enghoff, H., 1995. Historical biogeography of the Holarctic: area relationships, ancestral areas, and dispersal of non-marine animals. Cladistics 11, 223–263.

Escalante, T., 2015. Parsimony analysis of endemicity and analysis of endemicity: a fair comparison. Syst. Biodivers. 13, 413–418.

Escalante, T., Rodríguez-Tapia, G., Szumik, C., Morrone, J.J., Rivas, M., 2010. Delimitation of the Nearctic region according to mammalian distributional patterns. J. Mammal. 91, 1381–1388.

Escalante, T., Morrone, J.J., Rodríguez-Tapia, G., 2013. Biogeographic regions of North American mammals based on endemism. Biol. J. Linn. Soc. 110, 485–499.

Ferrari, A., Paladini, A., Schwertner, C.F., Grazia, J., 2010. Endemism analysis of Neotropical Pentatomidae (Hemiptera, Heteroptera). Iheringia Sér. Zool. 100, 449–462.

Forero, D., 2008. Revision and phylogenetic analysis of the *Hadronema* group (Miridae: Orthotylinae: Orthotylini), with descriptions of new genera and new species, and comments on the Neotropical genus *Tupimiris*. Bull. Am. Mus. Nat. Hist. 312, 172pp.

Goloboff, P., 2011. NDM/VNDM ver. 3.0. Programs for identification of areas of endemism. Available at: http://www.zmuc. dk/public/phylogeny/endemism (Accessed 2 June 2011).

Hausdorf, B., Hennig, C., 2003. Biotic element analysis in biogeography. Syst. Biol. 52, 717–723.

Henry, T.J., 2007. Synopsis of the Eastern North American species of the plant bug genus *Parthenicus*, with descriptions of three new species and a revised key (Heteroptera: Miridae: Orthotylinae). Am. Mus. Nov. 3593, 1–30.

Henry, T.J., Kim, K.C., 1984. Genus *Neurocolpus* Reuter (Heteroptera-Miridae): taxonomy, economic implications, hosts, and phylogenetic review. Trans. Am. Entomol. Soc. 110, 1–75.

Kays, R.W., Wilson, D.E., 2009. Mammals of North America. Princeton University Press, Princeton, N.J., 816 pp.

Knight, H.H., 1925. Descriptions of thirty new species and two new genera of North American Miridae (Hemiptera). Bull. Brooklyn Entomol. Soc. 20, 33–58.

Knight, H.H., 1927. *Megalopsallus*, a new genus of Miridae with five new species from North America (Hemiptera). Annals Entomol. Soc. Am. 20, 224–228.

Knight, H.H., 1962. Ten new and six old species of *Lopidea* from North America (Hemiptera, Miridae). Iowa State J. Sci. 37, 29–41.

Knight, H.H., 1968. Taxonomic review: Miridae of the Nevada test site and the western United States. Brigham Young Univ. Sci. Bull.-Biol. Ser. 9, 1–282.

Liebherr, J.K., 1994. Biogeographic patterns of montane Mexican and Central American Carabidae (Coleoptera). Can. Entomol. 126, 841–860.

Marshall, C.J., Liebherr, J.K., 2000. Cladistic biogeography of the Mexican transition zone. J. Biogeogr. 27, 203–216.

Martínez-Hernández, F., Mendoza-Fernández, A.J., Pérez-García, F.J., Martínez-Nieto, M.I., Garrido-Becerra, J.A., Salmerón-Sánchez, E., Merlo, M.E., Gil, C., Mota, J.F., 2015. Areas of endemism as a conservation criterion for Iberian gypsophilous flora: a multi-scale test using the NDM/VNDM program. Plant Biosyst. 149, 483–493.

Matzke, N.J., 2013. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2.1.

Matzke, N.J., 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Syst. Biol. 63, 951–970.

Michonneau, F., Collins, M., Chamberlain, S.A. 2015. An interface to iDigBio's search API that allows downloading specimen records. In: R package version 0.3.1. R Foundation for Statistical Computing, Vienna. Available at: http://www.R-project.org/ (Accessed 1 August 2015).

Morrone, J.J., 1994. On the identification of areas of endemism. Syst. Biol. 43, 438–441.

Morrone, J.J., 2014a. Cladistic biogeography of the Neotropical region: identifying the main events in the diversification of the terrestrial biota. Cladistics 30, 202–214.

Morrone, J.J., 2014b. Biogeographical regionalisation of the Neotropical region. Zootaxa 3782, 1–110.

Nelson, G., Platnick, N.I., 1981. Systematics and Biogeography: Cladistics and Vicariance. Columbia University Press, New York, N.Y. 567 pp.

Nylander, J.A.A., Olsson, U., Alström, P., Sanmartín, I., 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: *Turdus*). Syst. Biol. 57, 257–268.

Platnick, N.I., 1991. On areas of endemism. Austral. Syst. Bot. 4, 11–12.

Prado, J.R., Brennand, P.G., Godoy, L.P., Libardi, G.S., Abreu-Júnior, E.F., Roth, P.R., Chiquito, E.A., Percequillo, A.R., 2015. Species richness and areas of endemism of oryzomyine rodents (Cricetidae, Sigmodontinae) in South America: an ndm/vndm approach. J. Biogeogr. 42, 540–551.

Ribeiro, G.C., Santos, C.M., Olivieri, L.T., Santos, D., Berbert, J.M., Eterovic, A., 2014. The world's biogeographical regions revisited: global patterns of endemism in Tipulidae (Diptera). Zootaxa 3847, 241–258.

Ronquist, F., 2001. DIVA version 1.2. Computer program for MacOS and Win32. Evolutionary Biology Centre, Uppsala

University. Available at: http://www. ebc. uu. se/systzoo/research/diva/diva.html (Accessed 30 June 2015).

Sanmartin, I., Enghoff, H., Ronquist, F., 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biol. J. Linn. Soc. 73, 345–390.

Schaffner, J.C., Schwartz, M.D., 2008. Revision of the Mexican genera *Ficinus* distant and *Jornandes* distant, with the description of 21 new species (Heteroptera, Miridae, Orthotylinae, Orthotylini). Bull. Am. Mus. Nat.Hist. 309, 87 pp.

Schuh, R.T., 2000a. Revision of *Oligotylus* Van Duzee with descriptions of ten new species from western North America and comments on *Lepidargyrus* in the Nearctic (Heteroptera: Miridae: Phylinae: Phylini). Am. Mus. Nov. 3300, 1–44.

Schuh, R.T., 2000b. Revision of the North American plant bug genus *Megalopsallus* Knight, with the description of eight new species from the West (Heteroptera: Miridae: Phylinae). Am. Mus. Nov. 3305, 1–69.

Schuh, R.T., 2001. Revision of New World *Plagiognathus* Fieber, with comments on the Palearctic fauna and the description of a new genus (Heteroptera: Miridae: Phylinae). Bull. Am. Mus. Nat. Hist. 266, 267pp.

Schuh, R.T., 2002–2013. On-line Systematic Catalog of Plant Bugs (Insecta: Heteroptera: Miridae). Available at: http://research.amnh.org/pbi/catalog/ (Accessed 15 August 2015).

Schuh, R.T., 2004a. Revision of *Tuxedo* Schuh (Hemiptera: Miridae: Phylinae). Am. Mus. Nov. 3435, 1–26.

Schuh, R.T., 2004b. Revision of *Europiella* Reuter in North America, with the description of a new genus (Heteroptera: Miridae: Phylinae). Am. Mus. Nov. 3463, 1–58.

Schuh, R.T., 2006. Revision, phylogenetic, biogeographic, and host analysis of the endemic western North American *Phymatopsallus* group, with the description of 9 new genera and 15 new species (Insecta, Hemiptera, Miridae, Phylinae). Bull. Am. Mus. Nat. Hist. 301, 115pp.

Schuh, R.T., 2008. *Beckocoris* Knight, *Nevadocoris* Knight, and *Lattinocoris artemisii*, new genus and new species, Asteraceae feeding Phylini (Hemiptera: Heteroptera: Miridae) from the interior of western North America. Proc. Entomol. Soc. Washington 110, 1068–1110.

Schuh, R.T., 2012. Integrating specimen databases and revisionary systematics. ZooKeys 209, 255–267.

Schuh, R.T., Schwartz, M.D., 1985. Revision of the plant bug genus *Rhinacloa* Reuter with a phylogenetic analysis (Hemiptera, Miridae). Bull. Am. Mus. Nat. Hist. 179, 379–470.

Schuh, R.T., Schwartz, M.D., 1988. Revision of the New World Pilophorini (Heteroptera, Miridae, Phylinae). Bull. Am. Mus. Nat. Hist. 187, 101–201.

Schuh, R.T., Schwartz, M.D., 2004. New genera, new species, new synonyms, and new combinations in North America and Caribbean Phylinae (Heteroptera: Miridae). Am. Mus. Nov. 3436, 1–36.

Schuh, R.T., Schwartz, M.D., 2005. Review of North American *Chlamydatus* Curtis species, with new synonomy and the description of two new species (Heteroptera: Miridae: Phylinae). Am. Mus. Nov. 3471, 1–55.

Schuh, R.T., Hewson-Smith, S., Ascher, J.S., 2010. Specimen databases: a case study in entomology using Web-based Software. Am. Entomol. 56, 206–216.

Schwartz, M.D., 1984. A revision of the black grass bug genus *Irbisia* Reuter (Heteroptera: Miridae). J. N. Y. Entomol. Soc. 92, 193–306.

Schwartz, M.D., 2004. *Melymacra*, a new plant bug genus (Heteroptera: Miridae: Orthotylinae) inhabiting manzanita, *Arctostaphylos* Adanson (Ericaceae). Pan-Pacific Entomol. 80, 29–41.

Schwartz, M.D., 2005. Redescription of *Strophopoda aprica* Van Duzee and the description of two new genera and five new species from the southwestern United States and northern Mexico (Heteroptera, Miridae). Am. Mus. Nov. 3489, 1–23.

Schwartz, M.D., 2006. Review of *Plesiodema* Reuter and a description of a new genus to accommodate *Psallus sericeus* Heidemann (Heteroptera: Miridae: Phylinae). Russian Entomol. J. 15, 211–220.

Schwartz, M.D., 2011. Revision and phylogenetic analysis of the North American genus *Slaterocoris* Wagner with new synonymy, the description of five new species and a new genus from Mexico, and a review of the genus *Scalponotatus* Kelton (Heteroptera, Miridae, Orthotylinae). Bull. Am. Mus. Nat. Hist. 354, 290pp.

Schwartz, M.D., Foottit, R.G., 1998. Revision of the Nearctic species of the genus *Lygus* Hahn, with a review of the Palaearctic species (Heteroptera: Miridae). Memoirs on Entomology 10. International Associated Publishers, Gainesville, FL, 428pp.

Schwartz, M.D., Schuh, R.T., 1999. New genera and species of conifer-inhabiting phyline plant bugs from North America (Heteroptera: Miridae). J. N.Y. Entomol. Soc. 107, 204–237.

Schwartz, M.D., Stonedahl, G.M., 1986. Revision of the plant bug genus *Noctuocoris* Knight (Heteroptera: Miridae: Orthotylinae). Pan-Pacific Entomologist. 62, 237–247.

Schwartz, M.D., Stonedahl, G.M., 1987. *Oaxacacoris*, a new plant bug genus and three new species of Orthotylini from Mexico (Heteroptera: Miridae). Proc. Entomol. Soc. Washington 89, 15–23.

Schwartz, M.D., Stonedahl, G.M., 2004. Revision of *Phoenicocoris* Reuter with descriptions of three new species from North America and a new genus from Japan (Heteroptera: Miridae: Phylinae). Am. Museum Novitates. 3464, 1–55.

Seltmann, K.C. 2016a. AOE analysis subset of the Arthropod Easy Capture (AEC) database. figshare. Available at: https://dx.doi.org/10.6084/m9.figshare.2060979 (Accessed 6 January 2016).

Seltmann, K.C. 2016b. AOE analysis software. zendo. Available at: https://dx.doi.org/10.5281/zenodo.44387 (Accessed 6 January 2016).

Seltmann, K.C., Schwartz, M.D., Schuh, R.T., Weirauch, C. 2016a. Host Plant Distributions of Nearctic Miridae. figshare. Available at: https://dx.doi.org/10.6084/m9.figshare.2059503 (Accessed 6 January 2016).

Seltmann, K.C., Schwartz, M.D., Schuh, R.T., Weirauch, C. 2016b. Miridae Distributions. figshare. Available at: https://dx.doi.org/10.6084/m9.figshare.2059506 (Accessed 6 January 2016).

Stonedahl, G.M. 1988. Revision of the mirine genus *Phytocoris* Fallén (Heteroptera, Miridae) for western North America. Bulletin of the American Museum of Natural History. 188, 257pp.

Stonedahl, G.M. 1990. Revision and cladistic analysis of the Holarctic genus *Atractotomus* Fieber (Heteroptera, Miridae, Phylinae). Bulletin of the American Museum of Natural History. 198, 88pp.

Stonedahl, G.M., Henry, T.J., 1991. A new genus of mirine plant bugs, *Gracilimiris*, with three new species from North America (Heteroptera: Miridae). J. N. Y. Entomol. Soc. 99, 224–234.

Stonedahl, G.M., Schuh, R.T., 1986. Squamocoris Knight and Ramentomiris, new nenus (Heteroptera: Miridae: Orthotylinae). A cladistic analysis and description of seven new species from Mexico and the western United States. Am. Museum Novitates 2825, 1–26.

Stonedahl, G.M., Schwartz, M.D., 1986. Revision of the plant bug genus *Pseudopsallus* Van Duzee (Heteroptera: Miridae). Am. Museum Novitates 2842, 1–58.

Stonedahl, G.M., Schwartz, M.D., 1988. New species of *Oaxacacoris* Schwartz & Stonedahl and *Pseudopsallus* Van Duzee, and a new genus, *Presidiomiris*, from Texas (Heteroptera, Miridae, Orthotylini). Am. Museum Novitates. 2928, 1–18.

Stonedahl, G.M., Schwartz, M.D., 1996. Two new genera for pine-inhabiting species of Phylini North America (Heteroptera: Miridae: Phylinae). Am. Museum Novitates 3166, 1–15.

Szumik, C.A., Goloboff, P.A., 2004. Areas of endemism: an improved optimality criterion. Syst. Biol. 53, 968–977.

Szumik, C.A., Goloboff, P.A., 2015. Higher taxa and the identification of areas of endemism. Cladistics 31, 568–572.

Szumik, C.A., Cuezzo, F., Goloboff, P.A., Chalup, A.E., 2002. An optimality criterion to determine areas of endemism. Syst. Biol. 51, 806–816.

Szumik, C., Aagesen, L., Casagranda, D., Arzamendia, V., Baldo, D., Claps, L.E., Cuezzo, F., Díaz Gómez, J.M., Di Giacomo, A., Giraudo, A., Goloboff, P., Gramajo, C., Kopuchian, C., Kretzschmar, S., Lizarralde, M., Molina, A., Mollerach, M., Navarro, F., Nomdedeu, S., Panizza, A., Pereyra, V.V., Sandoval, M., Scrocch, G., Zuloaga, F.O. 2012. Detecting areas of endemism with a taxonomically diverse data set: plants, mammals, reptiles, amphibians, birds, and insects from Argentina. Cladistics 28, 317–329.

The Taxonomic Name Resolution Service [Internet], 2015. iPlant Collaborative. Version 4.0. Available at: http://tnrs.iplantcollaborative.org (Accessed 1 December 2015).

Torres-Miranda, A., Luna-Vega, I., Oyama, K., 2013. New approaches to the biogeography and areas of endemism of Red Oaks (*Quercus* L., Section Lobatae). Syst. Biol. 62, 555–573.

Van Duzee, E.P., 1916. Monograph of the North American species of *Orthotylus* (Hemiptera). Proc. Calif. Acad. Sci. 6, 87–128.

Van Duzee, E.P., 1918. New species of Hemiptera chiefly from California. Proc. Calif. Acad. Sci. 8, 271–308.

Weirauch, C., 2006a. New genera, new species, and new combinations in western Nearctic Phylini (Heteroptera: Miridae: Phylinae). Am. Museum Novitates. 3521, 1–41.

Weirauch, C., 2006b. New genera and species of oak-associated Phylini (Heteroptera: Miridae: Phylinae) from western North America. Am. Museum Novitates. 3522, 1–54.

Weirauch, C., 2009. Two new genera of Phylini, *Roburocoris* and *Viscacoris*, from Mexico and the southwestern United States (Heteroptera: Miridae: Phylinae). Entomol. Am. 115, 1–35.

Wilson, D.E., Reeder, D.M. 2005. Mammal Species of the World: A Taxonomic and Geographic Reference. JHU Press, Baltimore. 2142 pp.

Wyniger, D., 2010. Resurrection of the Pronotocrepini Knight, with revisions of the Nearctic genera *Orectoderus* Uhler, *Pronotocrepis* Knight, and *Teleorhinus* Uhler, and comments on the Palearctic *Ethelastia* Reuter (Heteroptera: Miridae: Phylinae). Am. Museum Novitates. 3703, 1–67.

Wyniger, D., 2011. Revision of the Nearctic genus *Coquillettia* Uhler with a transfer to the tribe Phylini, the description of 14 new species, a new synonymy, and the description of two new Nearctic genera *Leutiola* and *Ticua* and two new species (Heteroptera: Miridae: Phylinae). Entomologica Americana. 117, 134–211.

Yu, Y., Harris, A.J., Blair, C., He, X., 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol. Phylogenet. Evol. 87, 46–49.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. The 45 consensus areas of endemism derived from the default Miridae data set and analysis. Numbers reported as headers are (in that order): area identifier number, number of candidate areas included, number of scoring species, minimum endemicity score and maximum endemicity score.

Fig. S2. The ten consensus areas of endemism derived from the plant data set and analysis. Numbers reported as headers are (in that order): area identifier number, number of candidate areas included, number of scoring species, minimum endemicity score and maximum endemicity score.

Fig. S3. The 55 consensus areas of endemism derived from the 1566 spp. Miridae data set and analysis (“smaller min. # records [3]”). Numbers reported as headers are (in that order): area identifier number, number of candidate areas included, minimum endemicity score and maximum endemicity score.

Fig. S4. The 33 consensus areas of endemism derived from the 1004 spp. Miridae data set and analysis (“larger min. # records [10]”). Numbers reported as headers are (in that order): area identifier number, number of candidate areas included, minimum endemicity score and maximum endemicity score.

Fig. S5. The 21 consensus areas of endemism derived from the default Miridae data set and analysis with more restrictive criteria for an area of endemism (“larger # of spp./area [5]”). Numbers reported as headers are (in that order): area identifier number, number of candidate areas included, minimum endemicity score and maximum endemicity score.

Fig. S6. The 12 consensus areas of endemism derived from the default Miridae data set and analysis with very restrictive criteria for an area of endemism (“larger # of spp./area [10]”). Numbers reported as headers are (in that order): area identifier number, number of candidate areas included, minimum endemicity score and maximum endemicity score.

Fig. S7. The 19 consensus areas of endemism derived from the default Miridae data set and analysis with smaller grid cell size (1 degree). Numbers reported as headers are (in that order): area identifier number, number of candidate areas included, minimum endemicity score and maximum endemicity score.

Table S1. Table of scoring species for the 45 consensus areas derived from the default Miridae data set. The numbers of scoring species are reported for each area and species names provided.

Table S2. Table of scoring species for the ten consensus areas derived from the plant data set. The numbers of scoring species are reported for each area and species names provided.

Table S3. Comparison of area size (as defined by number of grid cells) between the default Miridae and plant data sets and analyses.