The Report Committee for Kumar Prasad Mainali
Certifies that this is the approved version of the following report:

Areas of endemism for rare fauna in karst regions of Hays county, Texas

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Daniel A. Powers

Camille Parmesan

Areas of endemism for rare fauna in karst regions of Hays county, Texas

by

Kumar Prasad Mainali, M.S.; M.S.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

August 2014

Dedication

This report is dedicated to my parents for a number of reasons that are clear to me and for many other reasons that are vague to me. My dad infected me with insatiable hunger of knowledge. My mom, however, deserves a special note because she was the one to love me as soon as I turned to a zygote. She clearly has a bigger share on me than my dad does. My mom gave me her mitochondrial genome; even her X chromosome delivered to me an order of magnitude more genes than its counterpart Y chromosome offered by dad. The vague reasons are beautiful; I do not want to spoil them by mentioning them.

Acknowledgements

Blake Sissel created maps for Figs. 1 and 4 in ArcMap after the analysis was completed. His assistance is thankfully acknowledged.

Abstract

Areas of endemism for rare fauna in karst regions of Hays county, Texas

Kumar Prasad Mainali, M.S. Stat

The University of Texas at Austin, 2014

Supervisor: Daniel A. Powers

An area of endemism contains many species restricted to the area and therefore it is rich in species diversity. Consequently, an area of endemism is an area of high conservation priority. An area of endemism is always determined with reference to a bigger landscape using various algorithms and mathematical approaches. Using parsimony analysis of endemicity (PAE) and endemism (NDM), this study analyzed distribution of 45 rare fauna – aquatic and terrestrial salamanders and arthropods – in karst regions of Hays county, Texas. PAE sought for the most parsimonious solutions heuristically by creating 97,216 trees. The method stored 16 best solutions from which a consensus was generated. NDM analyzed 285 potential areas of endemism. The area of endemism with highest endemicity score determined by NDM and the consensus tree generated by PAE select the identical geographic range as the best area of endemism. The two methods have many differences in the specifications of determining endemicity but have a common fundamental principle: determining geographic ranges with many species

largely confined to it. The two methods select 12% of the karst region with species records as area of endemism, which has 64% of the total species, with 38-40% species being endemic to the area.

Table of Contents

CHAPTER 1

INTRODUCTION

An area of endemism in a landscape remains within a boundary where non-random distributional congruence is observed among various taxa (Morrone 1994). Different from an area of distribution, an area of endemism is relatively small compared to the whole region and contains many endemic species (Szumik and Goloboff 2004) with disproportionately high species richness (Orme et al. 2005). The species pattern in area of endemism is considered by some to be determined by the region's history (Rosen 1978); others argue that the distributional pattern is a result of contemporary environment (Francis and Currie 2003). Irrespective of causality behind distributional pattern, an area of endemism is a region of high conservation priority (Stattersfield et al. 1998) because many narrow range species are confined in the area at present.

The distribution pattern of a handful species can be visually inspected to determine approximate area of endemism. When dealing with many species, especially when their distribution has complex overlap, a quantitative approach is essential to determine area with high richness of species that are restricted to the area. Various approaches have been

proposed to determine areas of endemism, of which parsimony analysis of endemicity and NDM are commonly used.

Parsimony analysis, widely used in phylogenetic studies, can be applied to delimit the areas of endemism within a region. The parsimony analysis of endemicity (PAE) was originally described by Rosen (1988), and an algorithm and analytical framework was implemented by Morrone (1994). Morrone's method classifies a region into a grid and follows a cladistic approach to unite the operational geographic units (grid cells) according to their shared species. The cells are clustered based on synapomorphies (shared taxa among grid cells). The most parsimonious tree has the minimum tree length or minimum steps required to explain all data. A cluster has grid cells with some taxa that are found in multiple cells and some taxa unique to the cells. The most parsimonious solution contains clusters of grid cells which are areas of endemism.

In a PAE, grid cells are united in a cluster based only on shared taxa; it does not incorporate spatial relationship among cells in the analysis. When a cluster in the most parsimonious solution contains grid cells that are not contingent, a single area of endemism is split into geographically isolated areas. This is not congruent with biogeographic explanation of endemism. Such a situation also presents problems in conservation planning because non-adjacent cells are more difficult to manage than a single contiguous area. To address these issues, Szumik et al (2002) designed an optimality criterion protocol to determine areas of endemism. This protocol, called NDM (shorthand of eNDeMicity), was updated in 2004 (Szumik and Goloboff 2004). NDM creates many sets of grid cells and calculates endemicity score for each set based on

uniqueness of species and contiguity of cells in the set. NDM differs from PAE in the following ways (Szumik et al. 2002):

1. Higher scores are given for continuous areas than discontinuous ones.
2. A continuous range of scores, as opposed to zero for absent and 1 for present, given by NDM provides more information about a species' endemicity in a region, and this approach could be more useful for conservation. For instance, a species with fewer records outside the area is more endemic to the area than a species with more records outside.
3. Two sets of grid cells can partially overlap and still represent two areas of endemism if different sets of species contribute to their endemicity score.

Because PAE and NDM have different criteria and algorithm to determine areas of endemism, it is not necessary that both methods pick same geographic area as the area of endemism. However, both methods are guided by the same principle: determine geographic space within a landscape with many species restricted to the area. An overlap of the area of endemism determined by the two methods is, therefore, expected.

This study determines the areas of endemism for the rare karst fauna of Hays county, Texas with both PAE and NDM. The analysis will use distribution record of 45 aquatic and terrestrial salamander and arthropod species. Hays county lies within "Edwards Plateau", an ecoregion characterized by karst topography and underground drainage, supporting many endemic and unique aquatic and terrestrial biota (Bowles and Arsuffi 1992). Various karst invertebrates of the ecoregion are federally listed as threatened and

endangered species (Bowles and Arsuffi 1992, Campbell 2003). Urban development is reported as the primary threat to such karst species (Campbell 2003). As the urban expansion continues, it is important to recognize the areas of endemism in the karst regions for appropriate conservation planning.

CHAPTER 2

METHODS

Study area, datasets and operational geographical units

Occurrence records for the species were provided by Zara Environmental LLC. The data included a total of 136 occurrence records for 46 species. The study area, most of Hays county, was divided into operational geographical units of three sizes: 2.5 km × 2.5 km, 5 km × 5 km, and 10 km × 10 km. I performed a preliminary NDM analysis and found that a grid of a 5 km gave much higher endemicity score than that of 2.5 km or 10 km. So, I selected 5 km resolution for the analysis. The entire grid consisted of 90 cells; karst region was found in 52 of them. Species occurrences were recorded in 26 of the 52 cells. Each operational geographical unit (cell) in the grid was given a unique number (Figure 1, Table 1). This pattern of numbering is maintained for both PAE and NDM analyses.

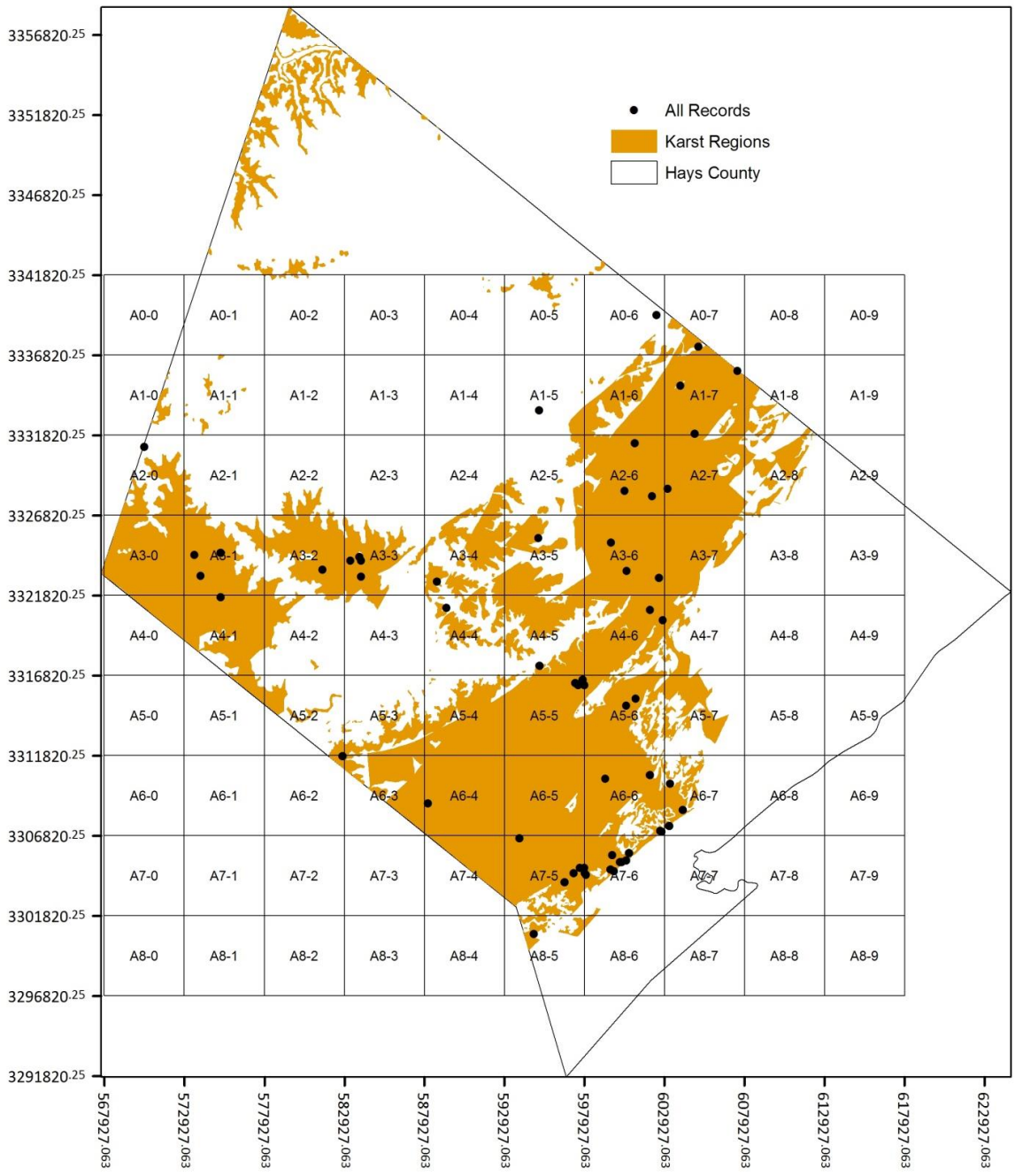

Figure 1. Study area with all occurrence points and operational geographical units (grid cells). Occurrence points in black dots were collected from Karst regions (in orange background) of Hays county, Texas. In several instances, a single black dot represents several occurrence points that overlay on each other. The axes carry UTM coordinates.

Table 1. All occurrence records and their placement in operational geographic units.

Cell ID	Site Name	Name of Species
A6-6	Artesian Well	Allotexiweckelia hirsuta
A6-6	Artesian Well	Artesia subterranea
A7-6	Ezell's Cave	Artesia subterranea
A6-6	Artesian Well	Calathaemon holthuisi
A7-6	Ezell's Cave	Calathaemon holthuisi
A7-6	Ezell's Cave	Cicurina ezelli
A6-6	McGlothin Sink	Cicurina ubicki
A7-6	Ezell's Cave	Eidmanella n. sp.
A6-6	McGlothin Sink	Eidmanella n. sp.
A6-7	San Marcos Springs	Eurycea nana
A6-6	Artesian Well	Eurycea rathbuni
A7-6	Ezell's Cave	Eurycea rathbuni
A7-6	Johnson's Well	Eurycea rathbuni
A7-6	Primer's Well	Eurycea rathbuni
A6-7	Rattlesnake Cave	Eurycea rathbuni
A6-7	San Marcos Springs	Eurycea rathbuni
A6-6	Seep on Sessoms Creek	Eurycea rathbuni
A7-6	Wonder Cave	Eurycea rathbuni
A7-6	Johnson's Well	Eurycea robusta
A7-6	Primer's Well	Eurycea sp. federally listed (nana/sosorum)
A6-6	Artesian Well	Haideoporus texanus
A6-7	San Marcos Springs	Heterelmis comalensis
A6-6	Artesian Well	Holsingerius samacos
A6-6	Artesian Well	Lirceolus smithii
A6-7	San Marcos Springs	Lirceolus smithii
A6-6	Artesian Well	Mooreobdella n.sp.
A7-6	Ezell's Cave	Mooreobdella n.sp.
A6-7	San Marcos Springs	Mooreobdella n.sp.
A6-6	Artesian Well	Palaemonetes antrorum
A7-6	Ezell's Cave	Palaemonetes antrorum
A7-6	Johnson's Well	Palaemonetes antrorum
A7-6	Wonder Cave	Palaemonetes antrorum
A6-6	Artesian Well	Phreatodrobia micra
A6-7	San Marcos Springs	Phreatodrobia micra

Table 1 (continued).

Cell ID	Site Name	Name of Species
A6-6	Artesian Well	Phreatodrobia plana
A6-7	San Marcos Springs	Phreatodrobia plana
A6-7	San Marcos Springs	Phreatodrobia punctata
A6-6	Artesian Well	Phreatodrobia rotunda
A6-7	San Marcos Springs	Phreatodrobia rotunda
A7-6	Ezell's Cave	Rhadine n. sp. 2 (subterranea group)
A6-7	Finger Cave	Rhadine n. sp. 2 (subterranea group)
A6-6	Artesian Well	Seborgia relicta
A7-6	Ezell's Cave	Seborgia relicta
A6-6	Artesian Well	Sphalloplana mohri
A7-6	Ezell's Cave	Sphalloplana mohri
A6-6	Artesian Well	Stygobromus flagellatus
A7-6	Ezell's Cave	Stygobromus flagellatus
A6-7	Rattlesnake Cave	Stygobromus flagellatus
A6-7	San Marcos Springs	Stygobromus flagellatus
A6-6	Artesian Well	Tethysbaena texana
A7-6	Ezell's Cave	Tethysbaena texana
A6-7	San Marcos Springs	Tethysbaena texana
A7-6	Electrical Cord Cave	Texella mulaiki
A7-6	Ezell's Cave	Texella mulaiki
A6-6	McGlothin Sink	Texella mulaiki
A7-6	Slip Cave	Texella mulaiki
A7-6	Tricopherous Cave	Texella mulaiki
A7-6	WWD-72T	Texella mulaiki
A7-6	Ezell's Cave	Texella renkesae
A6-6	Artesian Well	Texiweckelia texensis
A7-6	Ezell's Cave	Texiweckelia texensis
A6-7	San Marcos Springs	Texiweckelia texensis
A6-6	Artesian Well	Texiweckeliopsis insolita
A6-7	San Marcos Springs	Texiweckeliopsis insolita

Parsimony Analysis of Endemicity

The PAE was performed with TnT v. 1.1 (Goloboff et al. 2003). The exact solution can be determined with TnT within a reasonable amount of time for a dataset with 15–30 grid cells (Goloboff et al. 2008). For bigger datasets, heuristic search is more appropriate because of the amount of computation involved. However, heuristic methods can yield a sub-optimal solution because of local optima. I analyzed a subset of the data (only terrestrial species) with heuristic algorithm and found that the solution was identical to exact solution. This indicates that heuristic solutions are close to exact solution. Based on this, heuristic solutions for the complete dataset were generated.

A heuristic search method was employed with branch swapping between random trees. The best (most parsimonious) trees were saved and used to generate consensus tree. Finally, a consensus tree was generated with both strict and majority rule. The strict consensus tree contains only those clusters found in all most parsimonious trees whereas the majority rule consensus tree (at a cutoff of 50) contains all the clusters found in at least half of such trees (Goloboff et al. 2008). Both consensus trees and a map of synapomorphies were saved.

NDM

The NDM program (NDM/VNDM version 2.5, Goloboff 2004) computes endemicity scores for various sets of cells with the following formula (Szumik and Goloboff 2004). The score, *E*, for area, *A*, with a fixed number of grid cells, *n*, is given by:

$$E = \sum_{j=0}^{n} V_j$$

where V_j is the endemicity score of individual species *j*, which is given by:

$$V_j = \frac{p + (iF_i) + (aF_a)}{S + \frac{o}{F_o} + \frac{d}{F_d} + \frac{n}{F_n}}$$

"where *p*, number of cells in *A* in which species *j* is present; *i*, number of cells in which species is not present but is inferred as present because it is present in the surrounding cells; *a*, number of cells in *A* in which species is assumed to be present; *a* is equal to zero for our purposes; *o*, number of cells adjacent to *A* in which species has been observed; *d*, number of cells adjacent to *A* in which species has been assumed; *n*, number of cells outside of *A* and non-adjacent to *A* and in which species has been assumed (Szumik and Goloboff 2004). Both *d* and *n* are equal to zero in this analysis.

The influence of inferred and assumed presence is made more or less influential by giving a score between 0 and 1 for the factors F_i and F_a. The following defaults provided by the program were used: $F_i = 0.5$, $F_a = 0.75$, $F_o = 0.5$, $F_d = 2$, and $F_n = 0.5$ (Szumik and Goloboff 2004). However, the only multipliers that apply to this analysis are F_i and F_o.

Because there are no "assumed" occurrences, F_a, F_d and F_n do not contribute to the endemicity score.

Two partially overlapping areas can be considered as separate areas of endemism if different species contribute to the endemicity score, i.e., different species are endemic to the two overlapping areas according to the criteria of NDM. Because of relatively small number of species in the analysis, a relaxed rule was used in overlapping areas of endemism; two sets of cells were considered separate areas of endemism if at least 50% of the species were unique.

CHAPTER 3

RESULTS

TnT saved 16 best trees after 97,216 rearrangements. The consensus trees generated from the 16 best trees are presented in Figures 2 and 3. The area of endemism (red box in Figure 3) carries 18 species in three cells. NDM determined 285 "potential" areas of endemism with various endemicity scores. A default cutoff endemicity score of 2 was used by NDM to determine a group of cells as an area of endemism. This yielded only one area of endemism with a score of 14.33; the selected area had 17 species endemic to the cells (Figure 4). Different areas of endemism can have different sets of species. So, more than one areas of endemism together will have more endemic species than any one of them. In this case, NDM identified only one area with an endemicity score greater than 2, and other areas of endemism with a score of less than 2. Given the large difference in endemicity score between the top two areas of endemism (14.33 vs less than 2), only the first group of cells was selected as the area of endemism.

The consensus tree generated from the 16 best trees produced by PAE method (Figure 2, 3) and the area of endemism determined by NDM (Figure 4) choose the same set of three

grid cells as the area of highest endemicity. Table 2 lists the endemic species selected by the two methods in the area of endemism.

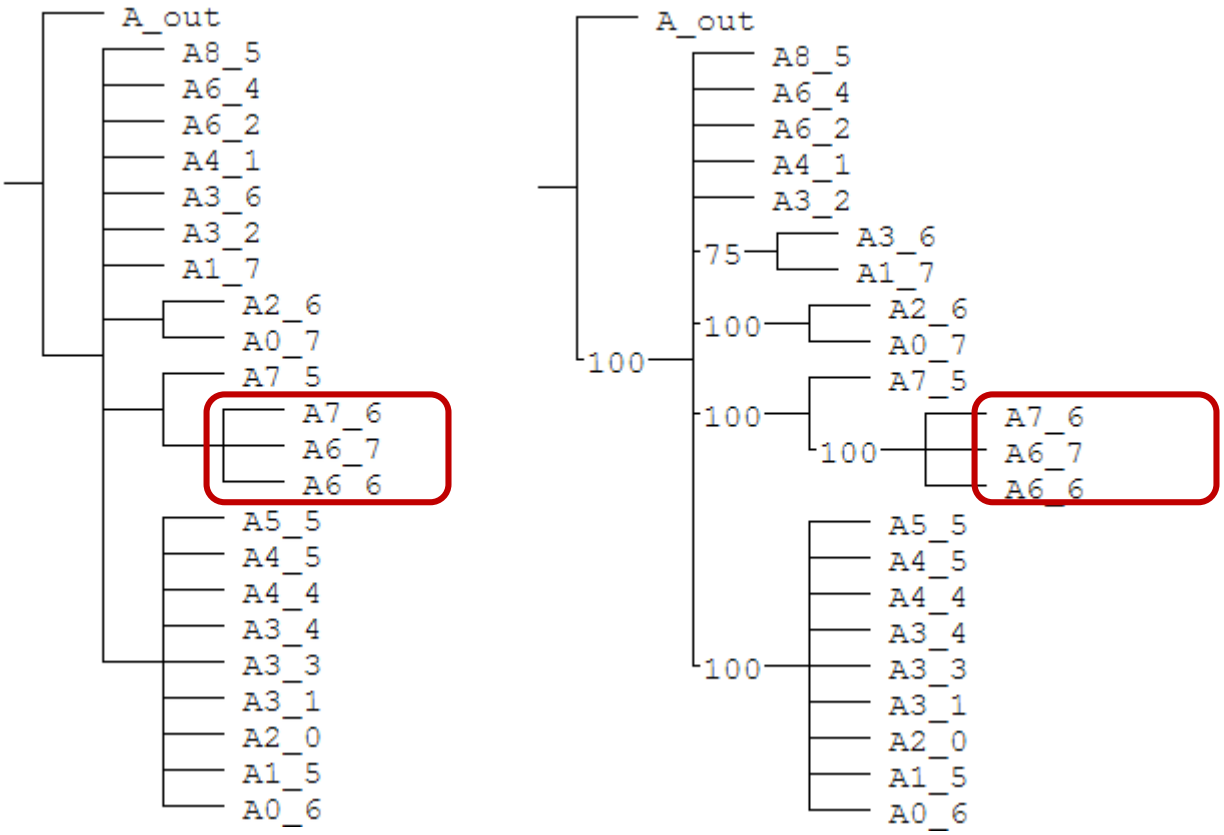

Figure 2. Parsimony Analysis of Endemicity of the study region performed in TnT. Strict consenssu tree (left) and majority rule (50%) consensus tree derived from 16 best trees generated after 97,216 rearrangements. Red box represents the "potential" area of endemism.

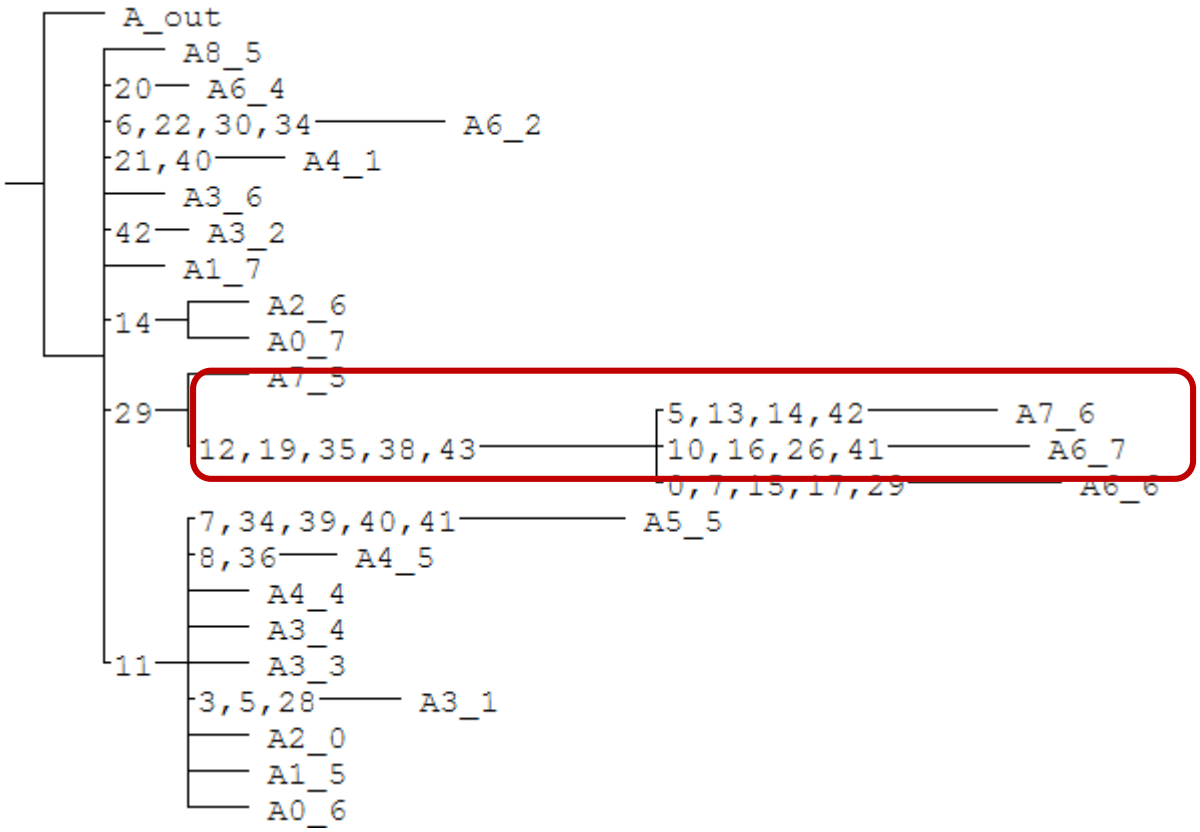

Figure 3. Synapomorphies common to all most parsimonious trees identified by PAE in TnT in Fig 2. Red box represents the area of endemism, for it carries a cluster of cells with many species of the entire study region that are endemic to the cluster of cells.

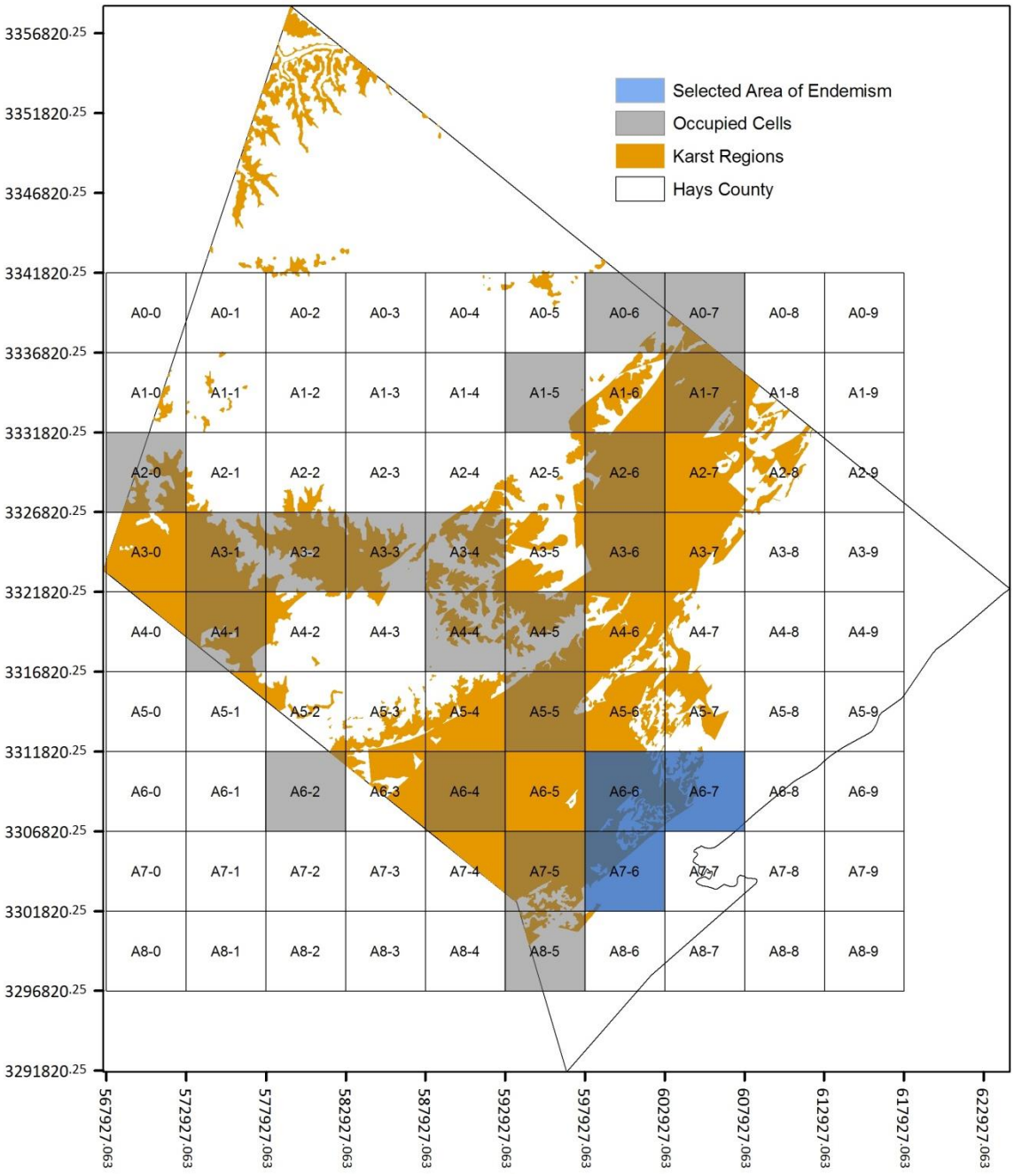

Figure 4. The only area of endemism with score >2 determined by NDM after examining 285 areas. Of the 29 species that exist in the selected area of endemism, only 17 contributed to the score (please see Table 2 for the list of species).

Table 2. Endemic species found in the area of endemism (Figure 3, 4) as determined by both parsimony analysis and NDM.

Species # used in TnT and NDM	Name of Species	Endemic species in the area of endemism	
		Parsimony analysis	NDM
0	Allotexiweckelia hirsuta	●	
1	Arrhopilites texensis		
2	Artesia subterranea		●
3	Batrisodes grubbsi		
4	Calathaemon holthuisi		●
5	Cicurina ezelli	●	
6	Cicurina russelli		
7	Cicurina ubicki	●	
8	Comaldessus stygius		
9	Eidmanella n. sp.		●
10	Eurycea nana	●	
11	Eurycea pterophila		
12	Eurycea rathbuni	●	●
13	Eurycea robusta	●	
14	Eurycea sp. federally listed (nana/sosorum)	●	
15	Haideoporus texanus	●	
16	Heterelmis comalensis	●	
17	Holsingerius samacos	●	
18	Lirceolus smithii		●
19	Mooreobdella n.sp.	●	●
20	Neoleptoneta eyeless n. sp.?		
21	Neoleptoneta n. sp. 1		
22	Neoleptoneta n. sp.2		
23	Palaemonetes antrorum		●
24	Phreatodrobia micra		●
25	Phreatodrobia plana		●
26	Phreatodrobia punctata	●	
27	Phreatodrobia rotunda		●
28	Rhadine insolita		
29	Rhadine n. sp. 2 (subterranea group)	●	●
30	Rhadine sp. [subterranea group] eyed		

Table 2 (continued).

Species # used in TnT and NDM	Name of Species	Endemic species in the area of endemism	
		Parsimony analysis	NDM
31	Rhadine sp. cf. austinica		
32	Seborgia relicta		●
33	Sphalloplana mohri		●
34	Stygobromus balconis		
35	Stygobromus flagellatus	●	●
36	Stygoparnus comalensis		
37	Tartarocreagris grubbsi		
38	Tethysbaena texana	●	●
39	Texella diplospina		
40	Texella grubbsi		
41	Texella mulaiki	●	
42	Texella renkesae	●	
43	Texiweckelia texensis	●	●
44	Texiweckeliopsis insolita		●

CHAPTER 4

DISCUSSION

NDM and PAE determined that different set of species present in the same set of cells were endemic to the area because most of the species that contributed to the endemicity score are different for the two methods (Table 2). This happened because of different mathematical approaches for computing endemicity score of a species in a selected region. However, the two methods selected the same set of cells as the area of endemism. That is because the two methods, in general, respond to the same biogeographic process: species confined to a region are likely to be endemic to the region. The area of highest endemism across the datasets, as determined by both methods, include three grid cells: A6-6, A6-7 and A7-6. The three cells comprise 12% of the karst region with species recorded for this study; the three cells, however, have 64% of the species in the region with 38-40% contributing to endemicity score (17 out of 45 species in NDM and 18 out of 45 species in PAE).

REFERENCES

Bowles DE, Arsuffi TL. 1992. Karst aquatic ecosystems of the Edwards Plateau region of central Texas, USA: A consideration of their importance, threats to their existence, and efforts for their conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* **3**(4): 317–329

Campbell L. 2003. *Endangered and Threatened Animals of Texas: Their Life History and Management.* US Fish and Wildlife Services.

Francis AP and Currie DJ. 2003. A globally consistent richness climate relationship for angiosperms. *American Naturalist* **161:** 523–536

Goloboff P. 2004. *NDM/VNDM, Programs for identification of areas of endemism*. Program and documentation, available at www.zmuc.dk/public/phylogeny/endemism

Goloboff P, J Farris, and K Nixon. 2003. *TNT: Tree Analysis Using New Technology*. Program and documentation, available from the authors, and at www.zmuc.dk/public/phylogeny

Goloboff PA, JS Farris, and KC Nixon. 2008. TNT, a free program for phylogenetic analysis. *Cladistics* **24**: 774–786

Morrone JJ. 1994. On the identification of areas of endemism. *Systematic Biology* **43**(3): 438–441

Orme CDL, Davies RG, Burgess M, Eigenbrod F, Pickup N, Olson VA, Webster AJ, Ding TS, Rasmussen PC, Ridgely RS, Stattersfield AJ, Bennett PM, Blackburn TM, Gaston KJ, Owens IPF. 2005. Global hotspots of species richness are not congruent with endemism or threat. *Nature* **436**: 1016–1019

Rosen BR. 1988. From fossils to earth history: applied historical biogeography. In: *Analytical biogeography: an integrated approach to the study of animal and plant distributions* (ed. by Myers AA and PS Giller), pp. 437–481. Chapman and Hall, London

Rosen DE. 1978. Vicariant patterns and historical explanation in biogeography. *Systematic Zoology* **27:** 159–188

Stattersfield AJ, MJ Crosby, AJ Long, and DC Wege. 1998. *Endemic Bird Areas of the World: Priorities for Biodiversity Conservation*. Bird Life International, Cambridge, UK

Szumik C, F Cuezzo, P Goloboff, and A Chalup. 2002. An optimality criterion to determine areas of endemism. *Systematic Biology* **51**: 806–816

Szumik CA, and PA Goloboff. 2004. Areas of Endemism: An Improved Optimality Criterion. *Systematic Biology* **53**(6): 968–977