Detecting clusters of high nontuberculous mycobacteria infection risk for persons with cystic fibrosis – An analysis of U.S. counties.

Rachel A Mercaldo [1], Julia E Marshall [1], D Rebecca Prevots [1], Ettie M Lipner [1], Joshua P French [2]

[1]Division of Intramural Research, Epidemiology and Population Studies Unit, NIAID, NIH, Rockville, MD, USA.

[2]Department of Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO, USA.

Keywords

nontuberculous mycobacteria; NTM; cystic fibrosis; clusters; clustering

Abstract

Nontuberculous mycobacteria are ubiquitous environmental bacteria that frequently cause disease in persons with cystic fibrosis (pwCF). The risks for NTM infection vary geographically. Detection of high-risk areas is important for focusing prevention efforts. In this study, we apply five cluster detection methods to identify counties with high NTM infection risk. Four clusters were detected by at least three of the five methods, including twenty-five counties in five states. The geographic area and number of counties in each cluster depended upon the detection method used. Identifying these clusters supports future studies of environmental predictors of infection and will inform control and prevention efforts.

Introduction

Nontuberculous mycobacteria (NTM) include a broad range of ubiquitous environmental bacteria species that cause chronic lung disease. Underling host susceptibilities include genetic, structural, and immunologic conditions that can put some persons at increased risk of disease. These conditions include cystic fibrosis, with a 5-year prevalence of 20% for NTM infections [1-3]. Treatment of NTM disease is challenging; prolonged antibiotic courses are required, and poor treatment responses are common [4].

The risk of NTM infection varies by geographical area of the United States. The state of Florida, for example, has the highest risk of NTM infection in the continental United States, both in CF and non-CF populations [2, 5]. This risk has been associated with environmental variables, such as evapotranspiration and percent coverage by surface water [5, 6]. In Colorado, Oregon, and Hawaii, we have shown intra-state variability associated with water quality factors, namely the concentrations of the trace metals vanadium and molybdenum in groundwater aquifers and surface water [7-10]. Studies in Australia have described both geographic and temporal trends associated with temperature and precipitation [11, 12]. In addition to climatic factors, high population density is associated with increased risk, possibly because persons with CF (pwCF), or other susceptible individuals, tend to be referred to tertiary care centers with specialized care teams [13]. This association is complicated by confounding environmental factors, such as water distribution systems that are very different in urban versus rural settings. To accurately identify geographic areas of high NTM infection risk, analytic approaches must control for the underlying population structure.

Studies among persons with CF offer invaluable insight regarding NTM epidemiology. The Cystic Fibrosis Foundation's patient registry (CFFPR) has served as a valuable resource for better understanding the epidemiology of NTM [14]. The registry acts as a repository of data for approximately 90% of pwCF in the United States, who are approached for participation upon CF diagnosis and continue to contribute data to the registry throughout their lifetimes, through their care at CF care clinics. Annually, the CF Foundation also releases a report describing the population of registrants. Since 2010, the registry has included data on NTM mycobacterial cultures and results, allowing researchers to identify trends in screening or infection. Data on the patient's geographic location of residence, at the zip code level, has been used for more precise estimation of geographic risk [14].

In analyses of geographic patterns, researchers are often also interested in identifying clusters of disease or infection. Clusters are collections of regions where incidence rates are higher—or, sometimes, lower—than those of surrounding regions [15]. Identifying the location of clusters offers researchers the opportunity to analyze environmental predictors at a broad level and is a valuable first step in identifying predictors of infection or disease.

In this study, we apply five tests for geographic clustering to data provided by the CFFPR, to identify high-risk areas of NTM infection in pwCF. These tests are the spatial scan method originally proposed by Kulldorff and Nagarwalla in 1995 [16], and four extensions of this method: the elliptic, flexibly-shaped, restricted flexible, and double connection scanning methods. These cluster detection methods were identified as having desirable combinations of sensitivity and positive predictive value [15]. In applying these methods, we describe the geography of NTM incidence in pwCF in US counties.

Methods

The study population comprised persons with cystic fibrosis (pwCF) represented in the Cystic Fibrosis Foundation Patient Registry (CFFPR) [14]. Approximately 90% of U.S. cystic fibrosis patients (or their guardians) consent to enrollment in the CFFPR upon CF diagnosis. The CFFPR offers the most complete and comprehensive data for cystic fibrosis and associated conditions in the United States. Since 2010, this dataset has included variables representing mycobacterial cultures and results for nontuberculous mycobacteria. We obtained a limited dataset for the study period of 2010 through 2019. We extracted zip code and nontuberculous mycobacteria isolation data for 29220 CFFPR patients in the United States aged $\geq$ 12 years.

Zip codes of patient residence were converted into county FIPS code using the zip code midpoint latitude and longitude, as provided by the United States Postal Services zip code database [17]. If a patient's zip codes contained apparent typographic errors but were 1) similar to their other listed zip codes and 2) were in the same state, that patient's zip codes were corrected using that patient's accurately formatted zip codes. Of the 29220 patients initially included in the dataset, 653 were excluded for missing zip codes or irreconcilable zip code errors. Zip code longitude and latitude were geocoded to county FIPS code and county mid-point longitude and latitude using R [18].

The baseline population for each county comprised all CF patients aged ≥ 12 years. Cases of incident NTM infection were defined as a pwCF with a positive NTM culture result after two consecutive negative results, who had lived in the same county for at least two years. The definition required two consecutive negative results to correct for possible false negative results and to reduce misclassification. For all other pwCF, the county in which they spent most of their time during the study period was selected. We excluded pwCF who were persistently NTM culture-positive or who had positive culture results after only a single negative culture. In applying these inclusion criteria, we excluded 3262 pwCF from analysis, with data for 25305 pwCF remaining.

Spatial scan methods to detect clusters of NTM cases among pwCF at the county level were performed in SaTScan [19] or in R, using the *smerc* or *rflexscan* packages [20, 21]. Spatial scan methods "scan" the regions in the study area to identify collections of regions (candidate zones) that have elevated incidence of disease relative to what is expected when the risk of outbreak is identical everywhere (possibly after adjusting for relevant explanatory variables). A suitable test statistic is computed for every candidate zone considered in the observed data set. If two candidate zones overlap, then only the candidate zone with the largest test statistic is retained. Many data sets are then simulated under the null hypothesis of no disease outbreak, for each simulated data set, the largest test statistic across all candidate zones is determined. The test statistics from the observed data set are compared to the test statistics from the simulated data sets to compute Monte Carlo p-values. The p-values are used to determine the significance of each region. The most likely cluster is the candidate zone observed with the largest test statistic while secondary clusters are candidate zones observed with smaller test statistics. French et al. provide an overview of many popular scan methods [15].

Applying scan methods to all potential candidate zones is computationally infeasible, so in practice, scan methods are applied on a much smaller but flexible number of candidate zones. In general, scan methods differ in the approach used to construct the set of candidate zones. Five spatial scan methods were applied: the original, circular, spatial scan method, proposed by Kulldorff and Nagarwalla, which detects circular clusters [16], and the elliptic [22], flexibly-shaped [23], restricted flexibly-shaped [24, 25], and double connection (DC) [26] extensions, which are better at detecting non-circular clusters. For all methods, the population upper-bound was set to 0.01 (1%), to ensure that clusters did not include more than 1% of the overall pwCF population. Default parameter values were otherwise selected for the circular, elliptic and DC scan tests. For the flexibly-shaped method, which considers all possible sets of connected counties within a given county's nearest neighbors, we set the limit of nearest neighbors to fifteen (k = 15). Finally, for the restricted flexibly-shaped scan test, we additionally filtered potential clusters by their middle p-value using α_1=0.2 [24, 25], to identify those clusters with the greatest risk.

County longitude and latitude were used directly for the circular, flexibly-shaped, restricted flexibly-shaped, and DC scan tests, and all tests were conducted in R using the *smerc* and *rflexscan* packages. As the elliptic scan test uses cartesian coordinates rather than longitude and latitude, we converted longitude and latitude to cartesian coordinates. These transformed coordinates were used within the SaTScan software for the elliptic scan method.

The results from the five cluster detection methods were compiled, and counties that were included in high-risk clusters by at least three of the five testing methods were identified. All high-risk clusters were mapped. These maps, and a table of all counties included in a high-risk cluster, are reported as Supplementary Material.

Results

Of the 25305 pwCF included in our analysis, 13239 (52.3%) were male, and the mean age was 30.22 years (sd: 13.5 years) at the beginning of 2019.

There were 3626 (14.3%) pwCF who met our definition of an incident NTM infection case. While the overall population of pwCF lived across 2359 continental US counties, only 1099 (47%) had cases. Twenty-five counties within five states were identified as high-risk by at least three of the five employed methods (Table 1, Figure 1). Areas in southern Florida, New York City, and Kansas City were included in clusters using all five methods. The size of these clusters, and the number of counties included in each, depended on the scanning method used (see Supplementary Material).

The first of the spatial clustering methods employed, the circular spatial scan statistic, returned four clusters of high NTM infection risk. Notably, these clusters were similar in size to the clusters detected in the same regions by the elliptic scan method, but included different counties. For example, the Kansas City cluster included 18 counties in Kansas and 14 in Missouri when the circular method was used. The elliptic scan returned 11 Kansas counties and 22 Missouri counties (Table S1).

The elliptic scan results included one additional cluster compared with the circular method. This fifth cluster included the San Francisco Peninsula region, a collection of five counties including San Francisco, Santa Clara, San Mateo, and Santa Cruz counties, as well as Marin County across the Golden Gate strait. While both the southern Florida and Kansas City clusters were also found to be significant using the elliptic scan statistic, the counties included differed. The only cluster in which the same counties were identified in both the circular and elliptic methods was the New York City region, including the counties of Kings, New York, Queens, and Richmond.

The flexibly-shaped, restricted flexibly-shaped, and double connection scanning methods were more specific than either the circular or elliptic scanning tests, identifying smaller clusters where the elliptic or circular methods would include more counties and a larger overall area. In our example of Kansas City, both the flexibly-shaped and restricted flexible scan methods included only eighteen counties while the double connection method included only nine. The flexibly-shaped and restricted flexible scanning methods also detected an additional cluster in California and Arizona that was not significant in the circular, elliptic, or double-connection tests (Table S1).

Discussion

A number of scanning methods can be used to detect clusters of an event of interest. In this study, we employed five such methods based on a Poisson model, to identify clusters of US counties with a higher than expected risk of NTM infection. Using the five methods concurrently, we identified twenty-five US counties, within five states, with higher than expected NTM infection risk among pwCF. NTM infection prevalence and incidence are increasing, both in pwCF and the general population [5, 27-30]. As NTM are environmental organisms, predicting the environmental conditions associated with infection will benefit prevention efforts.

The clusters of US counties described in this study may represent regions with optimal environmental conditions for NTM. Future studies could leverage these insights for discovery of significant environmental predictors of infection.

Previous studies have reported clusters of high-risk counties for NTM. California, Florida, Hawaii, Louisiana, New York, Oklahoma, Pennsylvania, and Wisconsin contain such counties, as reported by a study of Medicare Part B beneficiaries [5]. For pwCF, analysis of CFFPR data spanning 2010-2011 detected high risk counties centered in Wisconsin, Arizona, South Florida, and Maryland [3]. Our results are based on a longer time span, from 2010 to 2019, which likely explains the different clusters detected in this study. Several prior studies have also focused on prevalent infections, rather than incident, with greater sample sizes that could allow for greater power to detect clusters. The different results also suggest a need for analyses including a temporal component. The result of spatiotemporal clustering analyses may highlight trends in NTM risk geography that are relevant to the study of environmental determinants in a changing climate.

Our study also highlights the wisdom of using more than one method to detect relevant clustering. Though still widely applied, the circular spatial scan statistic originally proposed by Kulldorff and Nagarwalla detected fewer clusters than several of the extensions used in our study. Of the clusters identified, the circular scan method tended to include a broad area to maintain the circular shape required by the method, while the extensions were capable of more specific selection.

Our study does have several limitations. We used patients' reported zip codes to aggregate data by county and may have misclassified patients due to zip code errors even though we made efforts to rectify erroneous zip codes in our analysis (see Methods). Additionally, screening for NTM is not consistent across the US, and our clustering analysis is limited in that the likelihood of identifying incident NTM infections may vary by region. When screening rates are low, NTM cases may be overrepresented in the data, as only symptomatic individuals may be screened. Nonetheless, the population of persons with cystic fibrosis represent a high risk group, and annual screening for NTM is recommended by the American Thoracic Society.

By using data from this well-described population of high-risk individuals, we have described four significant clusters of counties with higher-than-expected risk of NTM infection.

As NTM are environmental organisms, spatial clustering may indicate areas of optimal environmental conditions for the bacteria. Further study of environments in these regions will add to what is known of NTM biogeography and benefit prevention efforts.

The 5 scan methods used in this study have been shown to perform better than competing scan methods [15]. The circular scan method [16] is the "original" spatial scan method. It searches for clusters with a circular shape. It is fast to apply and powerful but can struggle to identify irregularly shaped clusters. The elliptical scan method [16] adds elliptical candidate zones to the circular candidate zones of the circular scan method. It retains many of the positives of the circular scan method while being able to detect slightly more irregular clusters. However, the elliptical scan method does take slightly longer to apply than the circular scan method and still may not be able to detect highly irregular cluster shapes. The flexibly-shaped scan method [23] is able to detect highly irregular clusters by considering as candidate zones all possible sets of connected regions within a certain distance of each region. It takes longer to apply than the previous two methods. For a single data set, this is typically not an issue but can become problematic when applying the method to many data sets. The restricted flexibly-shaped scan method [24, 25] seeks to improve the computational speed of the flexibly-shaped scan method by pre-filtering certain regions from candidate zones. The clusters detected by the restricted flexibly-shaped scan method are typically smaller than the other methods, and it has reduced power to detect a cluster. The double connection scan method performs similarly to the restricted flexibly-shaped scan method but uses a greedy algorithm to search for candidate zones that maximize the test statistic. However, it too has less power than the circular, elliptical, and flexibly-shaped scan methods.

It is unlikely that all 5 scan methods considered will find the same clusters. There is no singular recommended approach for resolving this inconsistency; it is a result of the fact that the different methods use different sets of candidate zones. In principle, the candidate zones from all methods could simultaneously be considered, but this has never been done in practice and would take considerably longer. We suggest using the clusters returned by these competing approaches for hypothesis generation of possible causative factors explaining the why clusters appear in certain parts of the study area. The information returned by the different spatial scan methods is complementary rather than competitive.

Acknowledgements

RAM, JEM, DRP, and EML were supported by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. JPF was partially supported by NSF award 1915277 and NIH NIAID subcontract 75N93021P00818.

References

1. Floto, R.A., et al., *US Cystic Fibrosis Foundation and European Cystic Fibrosis Society consensus recommendations for the management of non-tuberculous mycobacteria in individuals with cystic fibrosis.* Thorax, 2016. **71 Suppl 1**: p. i1-22.
2. Adjemian, J., K.N. Olivier, and D.R. Prevots, *Epidemiology of Pulmonary Nontuberculous Mycobacterial Sputum Positivity in Patients with Cystic Fibrosis in the United States, 2010-2014.* Ann Am Thorac Soc, 2018. **15**(7): p. 817-826.
3. Adjemian, J., K.N. Olivier, and D.R. Prevots, *Nontuberculous mycobacteria among patients with cystic fibrosis in the United States: screening practices and environmental risk.* Am J Respir Crit Care Med, 2014. **190**(5): p. 581-6.
4. Daley, C.L., et al., *Treatment of Nontuberculous Mycobacterial Pulmonary Disease: An Official ATS/ERS/ESCMID/IDSA Clinical Practice Guideline.* Clin Infect Dis, 2020. **71**(4): p. 905-913.
5. Adjemian, J., et al., *Spatial clusters of nontuberculous mycobacterial lung disease in the United States.* Am J Respir Crit Care Med, 2012. **186**(6): p. 553-8.
6. Foote, S.L., et al., *Environmental predictors of pulmonary nontuberculous mycobacteria (NTM) sputum positivity among persons with cystic fibrosis in the state of Florida.* Plos One, 2021. **16**(12).
7. Lipner, E.M., et al., *Nontuberculous mycobacterial infection and environmental molybdenum in persons with cystic fibrosis: a case-control study in Colorado.* J Expo Sci Environ Epidemiol, 2022. **32**(2): p. 289-294.
8. Lipner, E.M., et al., *Nontuberculous Mycobacterial Disease and Molybdenum in Colorado Watersheds.* Int J Environ Res Public Health, 2020. **17**(11).
9. Lipner, E.M., et al., *Vanadium in groundwater aquifers increases the risk of MAC pulmonary infection in O'ahu, Hawaii.* Environmental Epidemiology, 2022.
10. Lipner, E.M., et al., *Nontuberculous Mycobacteria Infection Risk and Trace Metals in Surface Water: A Population-based Ecologic Epidemiologic Study in Oregon.* Ann Am Thorac Soc, 2022. **19**(4): p. 543-550.
11. Thomson, R.M., et al., *Factors associated with the isolation of Nontuberculous mycobacteria (NTM) from a large municipal water system in Brisbane, Australia.* Bmc Microbiology, 2013. **13**.
12. Thomson, R.M., et al., *Influence of climate variables on the rising incidence of nontuberculous mycobacterial (NTM) infections in Queensland, Australia 2001-2016.* Sci Total Environ, 2020. **740**: p. 139796.

13. McBennett, K.A., P.B. Davis, and M.W. Konstan, *Increasing life expectancy in cystic fibrosis: Advances and challenges.* Pediatr Pulmonol, 2022. **57 Suppl 1**: p. S5-S12.
14. Cystic Fibrosis Foundation, *Patient Registry*. 2022.
15. French, J.P., et al., *A comparison of spatial scan methods for cluster detection.* Journal of Statistical Computation and Simulation, 2022.
16. Kulldorff, M., *A spatial scan statistic.* Communications in Statistics-Theory and Methods, 1997. **26**(6): p. 1481-1496.
17. Codes, U.S.Z., *ZIP Code Database*. https://www.unitedstateszipcodes.org/zip-code-database/.
18. Team, R.C., *R: A Language and Environment for Statistical Computing*. 2022, R Foundation for Statistical Computing: Vienna, Austria.
19. Kulldorff, M. and I. Information Management Services, *SaTScan v9.7: Software for the spatial and space-time scan statistics*. 2021.
20. French, J. and M. Meysami, *smerc: Statistical Methods for Regional Counts*. 2022.
21. Otani, T. and K. Takahashi, *rflexscan: The Flexible Spatial Scan Statistic*. 2021.
22. Kulldorff, M., et al., *An elliptic spatial scan statistic.* Statistics in Medicine, 2006. **25**(22): p. 3929-3943.
23. Tango, T. and K. Takahashi, *A flexibly shaped spatial scan statistic for detecting clusters.* Int J Health Geogr, 2005. **4**: p. 11.
24. Tango, T. and K. Takahashi, *A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters.* Statistics in Medicine, 2012. **31**(30): p. 4207-4218.
25. Tango, T., *A Spatial Scan Statistic with a Restricted Likelihood Ratio.* Japanese Journal of Biometrics, 2008. **29(2)**: p. 75-95.
26. Costa, M.A., R.M. Assuncao, and M. Kulldorff, *Constrained spanning tree algorithms for irregularly-shaped spatial clustering.* Computational Statistics and Data Analysis, 2012. **56**(6): p. 1771-1783.
27. Adjemian, J., et al., *Prevalence of nontuberculous mycobacterial lung disease in U.S. Medicare beneficiaries.* Am J Respir Crit Care Med, 2012. **185**(8): p. 881-6.
28. Strollo, S.E., et al., *The Burden of Pulmonary Nontuberculous Mycobacterial Disease in the United States.* Ann Am Thorac Soc, 2015. **12**(10): p. 1458-64.
29. Thomson, R.M., N.T.M.w.g.a.Q.T.C. Centre, and L. Queensland Mycobacterial Reference, *Changing epidemiology of pulmonary nontuberculous mycobacteria infections.* Emerg Infect Dis, 2010. **16**(10): p. 1576-83.

30. Winthrop, K.L., et al., *Incidence and Prevalence of Nontuberculous Mycobacterial Lung Disease in a Large U.S. Managed Care Health Plan, 2008-2015.* Ann Am Thorac Soc, 2020. **17**(2): p. 178-185.

Declarations of interest

None.

Table 1: Counties included in clusters of high NTM infection risk areas. Asterisk (*) indicates the county was identified as part of a cluster by the given scan test. Shaded cells indicate the county was not included in a cluster by the given scan test.

	Spatial scan method				
State/County	Circular	Elliptic	Flexibly-shaped	Restricted Flexibly-shaped	Double Connection
CA					
Marin		*	*	*	
San Francisco		*	*	*	
Santa Clara		*	*	*	
Santa Cruz		*	*	*	
FL					
Charlotte	*	*	*	*	*
Collier	*	*	*	*	
Hendry	*	*	*	*	*
Martin	*	*	*	*	*
Okeechobee	*	*	*	*	*
Palm Beach	*	*	*	*	*
Sarasota		*	*	*	
St Lucie	*	*	*	*	*
KS					
Douglas	*	*	*	*	
Johnson	*	*	*	*	*
Wyandotte	*	*	*	*	*
MO					
Buchanan	*	*	*	*	
Clay	*	*	*	*	*
Clinton	*	*	*	*	*
Jackson		*	*	*	*
Johnson		*	*	*	*
Lafayette		*	*	*	*
NY					
Kings	*	*	*	*	*
New York	*	*	*	*	*
Queens	*	*	*	*	*
Richmond	*	*	*	*	*

308

Figures

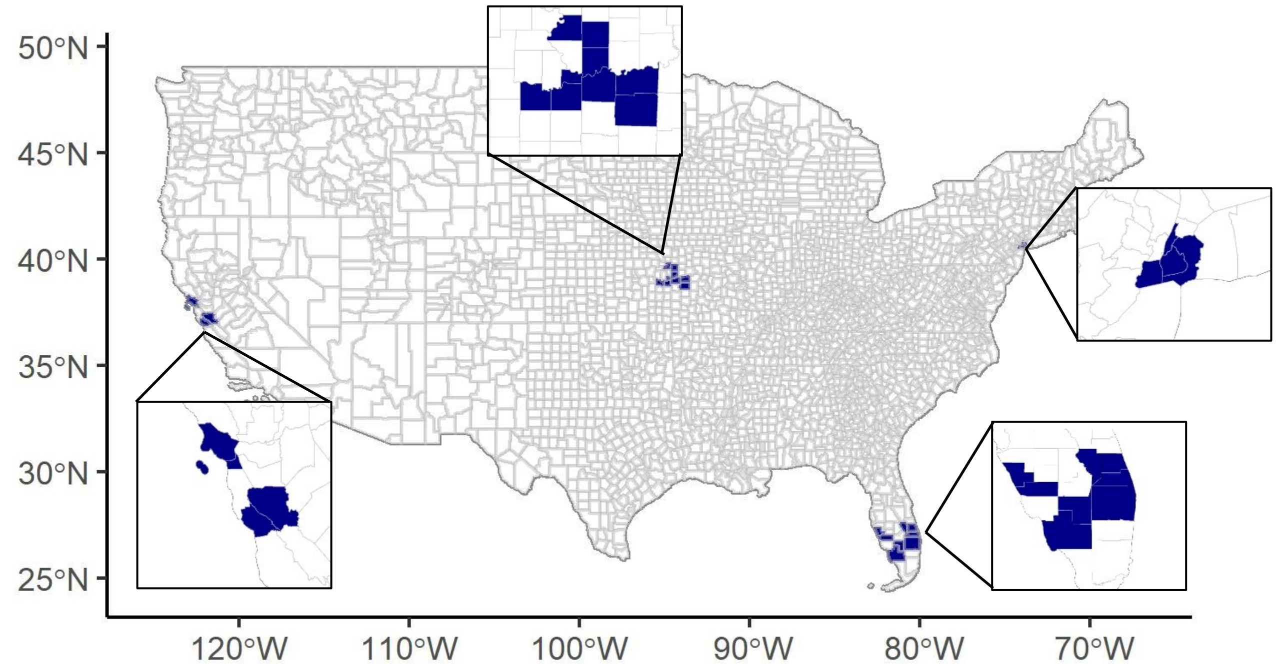

Figure 1: Counties found in significant clusters by at least three of the five scanning methods.

Supplementary Material

Table S1:

	Spatial scan method				
State/County	Circular	Elliptic	Flexibly-shaped	Restricted Flexibly-shaped	Double Connection
AZ					
Coconino			*		
Gila			*		
La Paz			*	*	
Mohave			*		
Pinal			*		
CA					
Alameda			*	*	
Contra Costa			*	*	
Marin		*	*	*	
Monterey			*		
Riverside			*	*	
San Bernardino				*	
San Diego			*	*	
San Francisco		*	*	*	
San Mateo		*			
Santa Clara		*	*	*	
Santa Cruz		*	*	*	
CO					
Adams			*	*	
Arapahoe			*	*	
Broomfield			*	*	
Denver			*	*	
Douglas			*	*	
El Paso			*	*	
Elbert			*	*	
Jefferson			*	*	
Morgan			*	*	
Pueblo			*	*	
Washington			*	*	
FL					
Charlotte	*	*	*	*	*

Collier	*	*	*	*	
Desoto	*	*			
Glades	*				*
Hardee	*	*			
Hendry	*	*	*	*	*
Highlands	*	*			
Lee	*	*			
Martin	*	*	*	*	*
Miami Dade			*	*	
Okeechobee	*	*	*	*	*
Palm Beach	*	*	*	*	*
Sarasota		*	*	*	
St Lucie	*	*	*	*	*
GA					
Appling	*	*			
Bryan	*	*			
Bulloch	*	*			
Burke	*	*			
Candler	*	*			
Chatham	*	*			
Columbia	*	*			
Effingham	*	*			
Emanuel	*	*			
Evans	*	*			
Glascock	*				
Glynn		*			
Hancock	*				
Jefferson	*	*			
Jenkins	*				
Johnson	*				
Laurens	*				
Liberty	*	*			
Lincoln	*				
Long	*				
Mcduffie	*	*			
Montgomery	*	*			
Richmond	*	*			
Screven	*	*			
Taliaferro	*				
Tattnall	*	*			
Toombs	*	*			

Treutlen	*				
Warren	*	*			
Washington	*	*			
Wayne		*			
Wheeler	*				
Wilkes	*				
IA					
Fremont	*				
Page	*				
Taylor	*				
KS					
Atchison	*	*			
Brown	*				
Doniphan	*				
Douglas	*	*	*	*	
Franklin		*			
Jackson	*				
Jefferson	*	*			
Johnson	*	*	*	*	*
Leavenworth	*	*			
Marshall	*				
Miami		*			
Nemaha	*				
Pottawatomie	*				
Riley	*				
Shawnee	*				
Wabaunsee	*				
Wyandotte	*	*	*	*	*
MO					
Andrew	*	*			
Atchison	*				
Benton		*			
Buchanan	*	*	*	*	
Caldwell		*			
Cass		*			
Cedar		*			
Clay	*	*	*	*	*
Clinton	*	*	*	*	*
Dekalb	*				
Gentry	*				
Henry		*			

Holt	*				
Jackson		*	*	*	*
Johnson		*	*	*	*
Lafayette		*	*	*	*
Nodaway	*	*			
Pettis					*
Platte	*	*			
Ray		*			*
St Clair		*			
Worth	*				
NE					
Gage	*				
Johnson	*				
Nemaha	*				
Otoe	*				
Pawnee	*				
Richardson	*				
NJ					
Hudson	*	*			
Mercer			*	*	
Middlesex			*	*	
Somerset			*	*	
Union			*	*	
NV					
Clark			*	*	
NY					
Kings	*	*	*	*	*
New York	*	*	*	*	*
Queens	*	*	*	*	*
Richmond	*	*	*	*	*
SC					
Aiken	*	*			
Allendale	*				
Bamberg	*				
Barnwell	*				
Beaufort	*	*			
Berkeley	*	*			
Calhoun	*				
Charleston	*	*			
Clarendon	*	*			
Colleton	*	*			

Dorchester	*	*			
Edgefield	*	*			
Fairfield	*				
Greenwood	*	*			
Hampton	*	*			
Jasper	*				
Laurens		*			
Lexington	*	*			
Mccormick	*				
Newberry	*	*			
Orangeburg	*	*			
Richland	*	*			
Saluda	*	*			
Sumter	*	*			

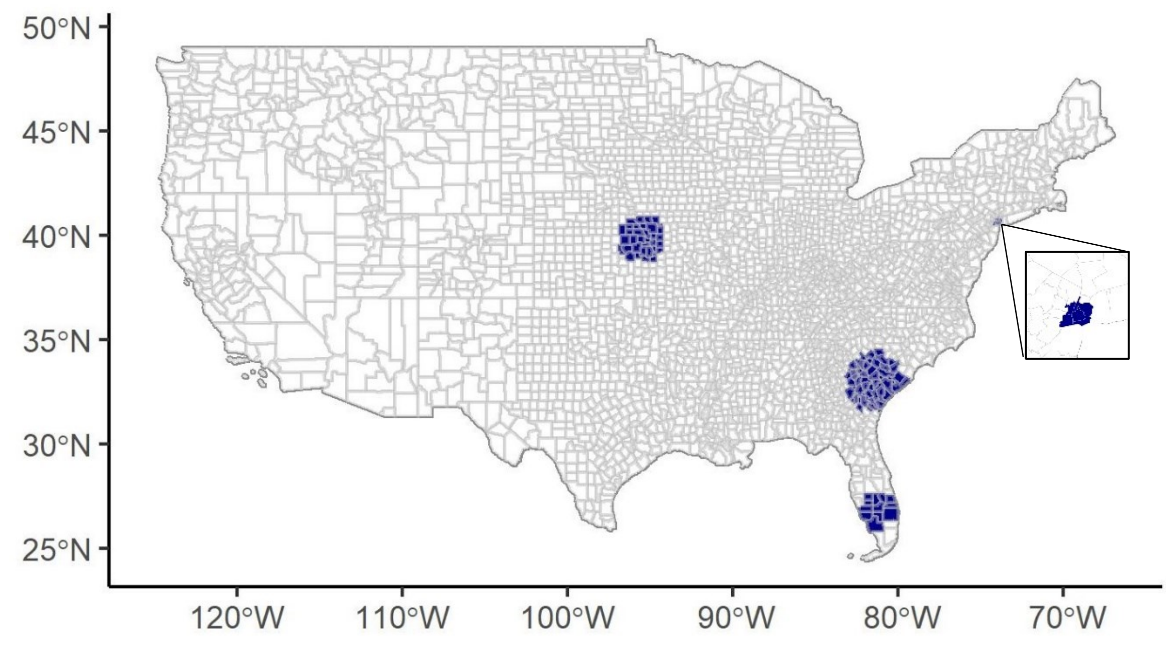

Figure S1: Circular scan results

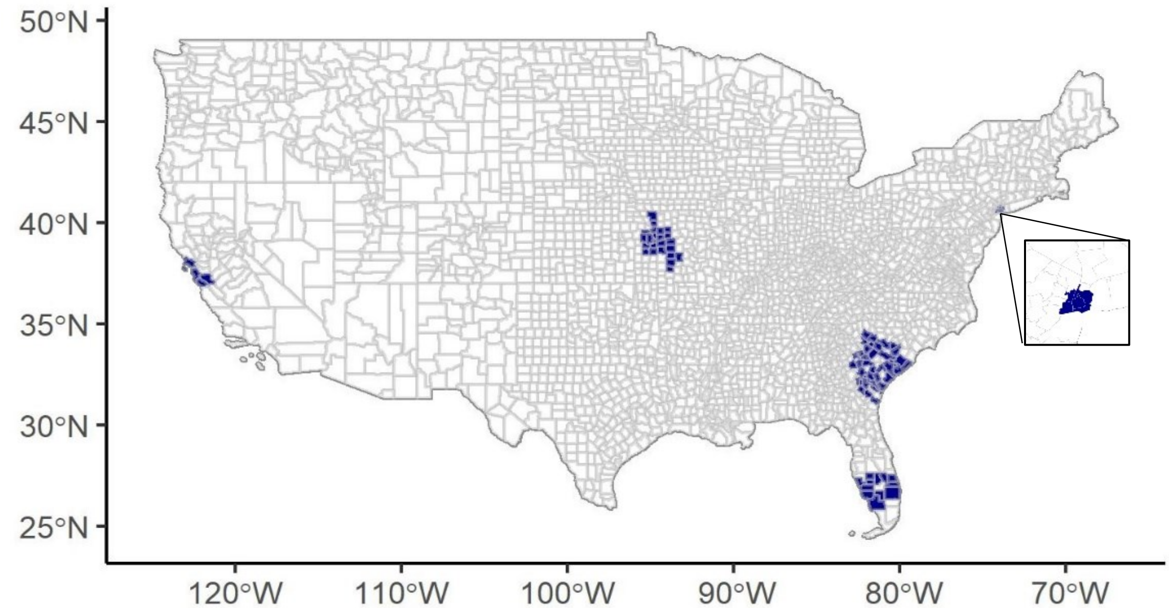

Figure S2: Elliptic scan results

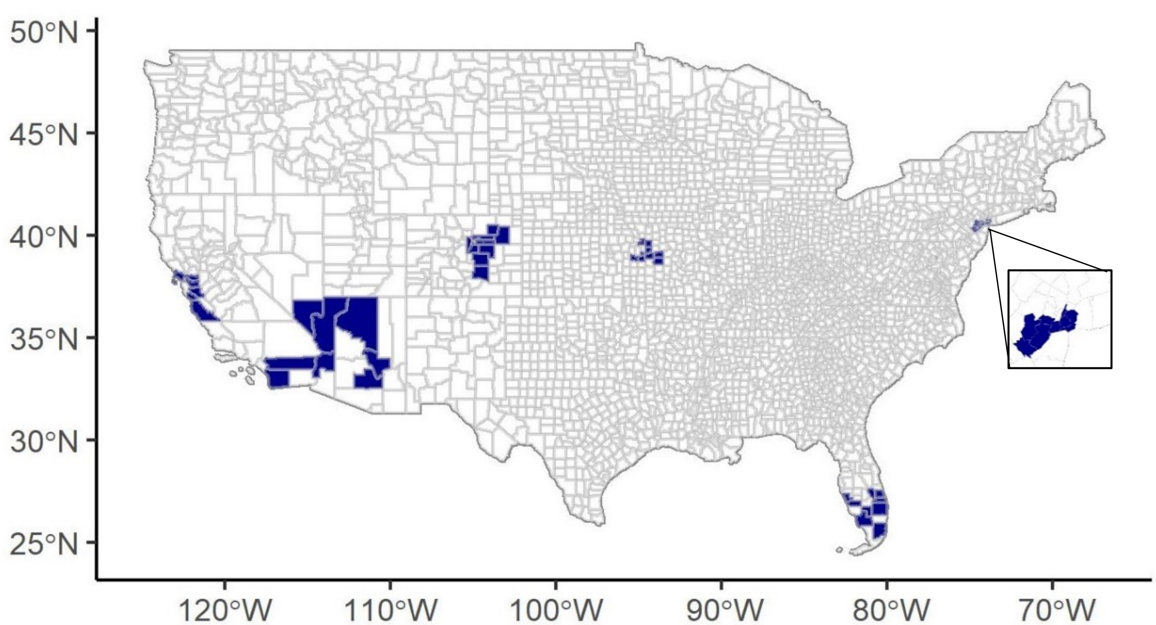

Figure S3: Flexibly-shaped scan results

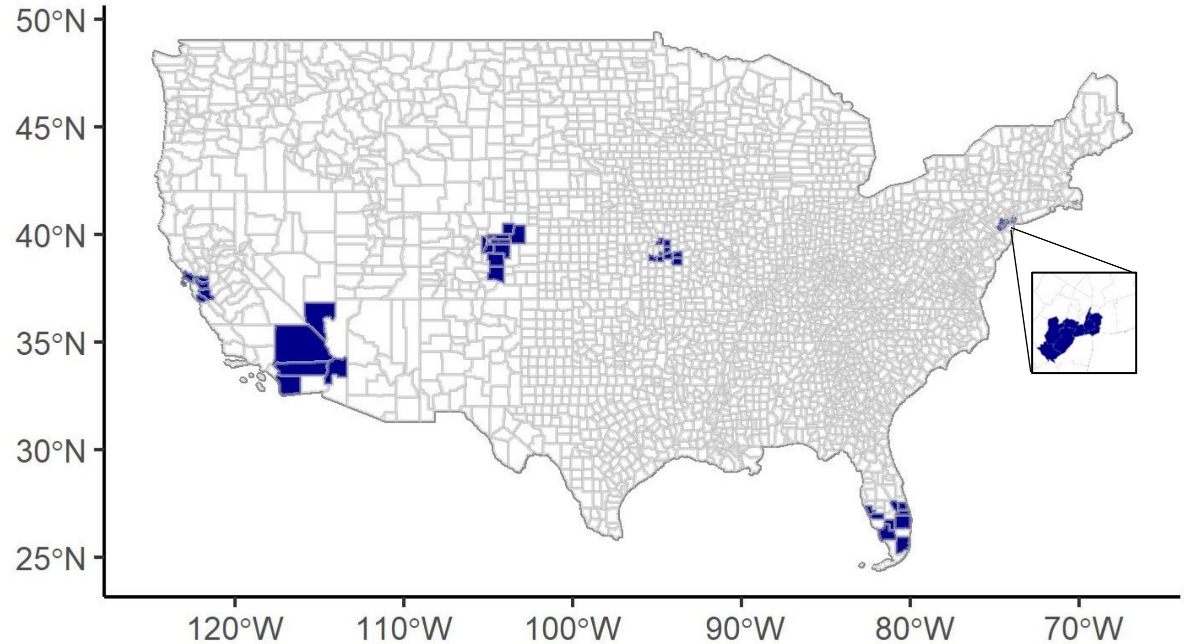

Figure S4: Restricted flexible scan results

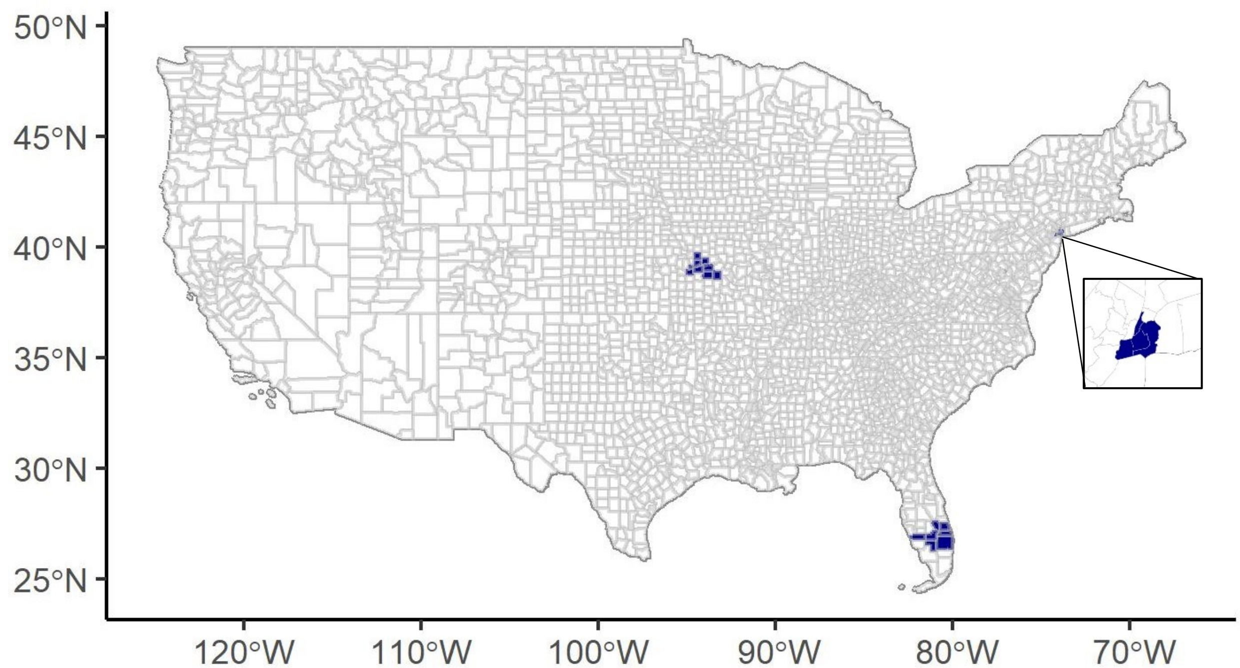

Figure S5: Double connection scan results