Retrieve the geographic coordinates associated with DNA sequences

Luis Amador

2023-08-29

Overview

This is a small tutorial to retrieve the geographic information of published DNA sequences in the form of geographic coordinates: Latitude and Longitude (e.g., Latitude = 35.0775, Longitude = -106.6625).

This work is part of the NSF project awarded to professor Lisa Barrow: *Determinants of amphibian genomic diversity: Integrating traits, phylogeny and geography*. Specifically, we are trying to retrieve geographic coordinates of mitochondrial DNA (mtDNA) sequences of the Cytochrome B gene belonging to amphibian species.

We recommend some steps to follow:

Steps

1. We have a directory with sequences of amphibian species in **FASTA** format. In the figure 1 we can see a typical alignment, in this case we see an alignment of nine sequences of the Tonchek spiny-chest frog (*Alsodes gargola*). Each sequence is coded with the scientific name and the accession number of **GenBank** (e.g., Alsodes.gargola.AY843787). This is our input, since we can extract from the alignment the information that we will need to find the geographic coordinates associated with a certain DNA sequence.

The alignment can also be opened in any program that reads text files, for example Microsoft Notepad or Microsofot WordPad, and BBEdit if you are using macOS. In addition, there are several free software to edit and visualize alignments, such as BioEdit, MEGA, Aliview or Jalview.

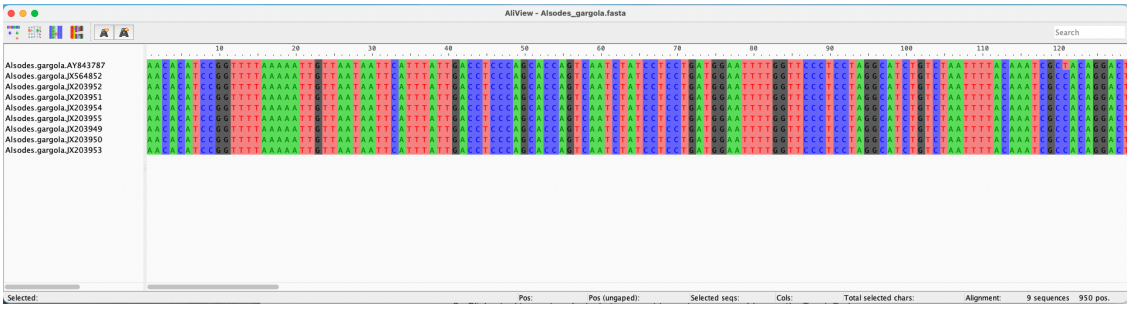


Figure 1: View of a typical alignment in the software Aliview.

2. We extract the accession number of each sequence (e.g., AY843787) of the alignment and go to the website to The National Center for Biotechnology Information (NCBI) https://www.ncbi.nlm.nih.gov/nucleotide/. Then we can search for the information of the sequence in the NCBI portal by write the accession number in the browser (Figure 2).

3. We can see the information of the GenBank submission that consist primarily of the nucleotide sequence data, source organism information, and sequence features. If the geographic coordinates are not part of the features of the GenBank record, two key parameters to start the search of the geographic

coordinates are: 1) “TITLE” that is the name of the publication where the sequenced was released, and 2) feature source such as the locality “country” and which are shown as shaded text (Figure 3).

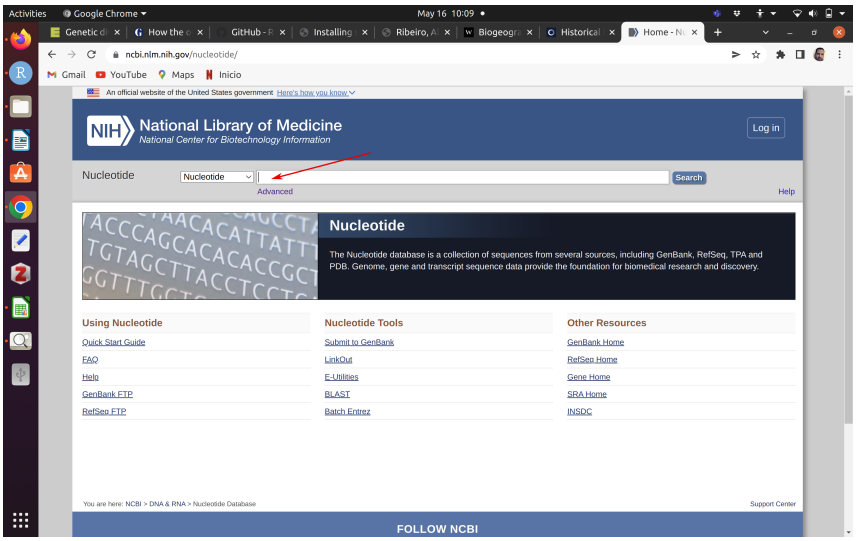


Figure 2: Window of NCBI portal to search in the nucleotide database.

4. In some cases, we can find in the GenBank submission page a link (Figure 4A) to the specimen voucher information deposited in databases (e.g., museum databases like arctos). Luckily, the geographic coordinates in latitude and longitude format can be directly extracted (copied) from this database (Figure 4B).

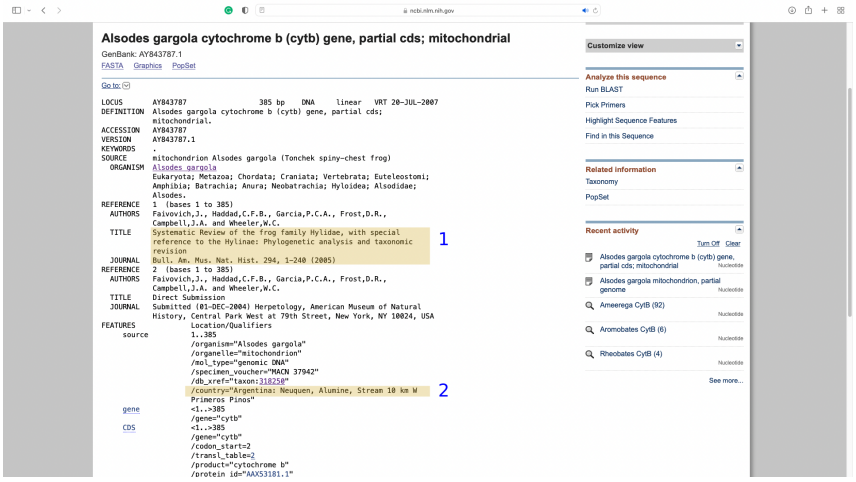


Figure 3: Window of NCBI portal with the information of the GenBank submission.

5. There are several ways to retrieve geographic coordinates from the information of the GenBank submission. Here we are showing two ways using as example the GenBank accession number KJ418850. In this example, the article associated to the sequence appears as unpublished (Figure 5) in “TITLE”, in addition, we can found information about the locality of the specimen voucher (Region de O’Higgins, Rinconada de Idahue).

First we can do a Google search with the title of the article (maybe it is published already!) and check within the article, for the information (i.e., geographic coordinates) related to the accession number/specimen, that could be presented wrote in the text of the article, in a table in the main text of the article, or as part of

Figure 4: Geographic coordinates found in linked database associated to voucher specimen - Genbank submission.

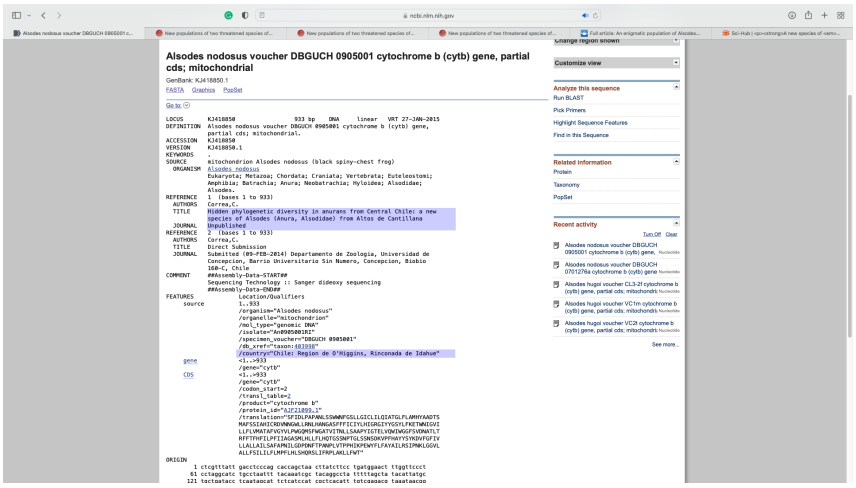
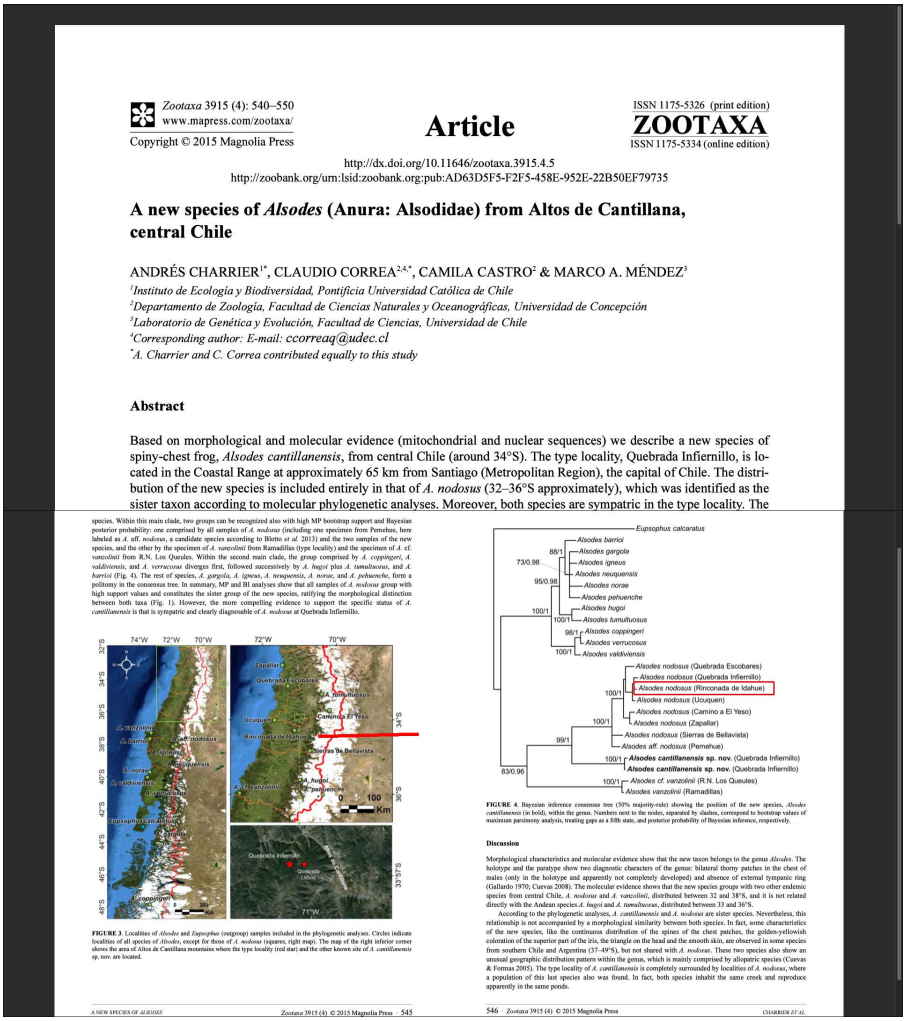


Figure 5: Window of NCBI portal with the information of the GenBank submission.

the supplementary material of the article. In this example, we do not find the geographic coordinates within the article, but we can use other information that could help to get the geographic coordinates (Figure 6).



Zootaxa 3915 (4): 540–550
www.mapress.com/zootaxa/
Copyright © 2015 Magnolia Press

Article

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

http://dx.doi.org/10.11646/zootaxa.3915.4.5
http://zoobank.org/urn:lsid:zoobank.org:pub:AD63D5F5-F2F5-458E-952E-22B50EF79735

A new species of *Alsodes* (Anura: Alsodidae) from Altos de Cantillana, central Chile

ANDRÉS CHARRIER[1*], CLAUDIO CORREA[2,4*], CAMILA CASTRO[2] & MARCO A. MÉNDEZ[3]

[1]*Instituto de Ecología y Biodiversidad, Pontificia Universidad Católica de Chile*
[2]*Departamento de Zoología, Facultad de Ciencias Naturales y Oceanográficas, Universidad de Concepción*
[3]*Laboratorio de Genética y Evolución, Facultad de Ciencias, Universidad de Chile*
[4]*Corresponding author: E-mail: ccorreaq@udec.cl*
[*]*A. Charrier and C. Correa contributed equally to this study*

Abstract

Based on morphological and molecular evidence (mitochondrial and nuclear sequences) we describe a new species of spiny-chest frog, *Alsodes cantillanensis*, from central Chile (around 34°S). The type locality, Quebrada Infiernillo, is located in the Coastal Range at approximately 65 km from Santiago (Metropolitan Region), the capital of Chile. The distribution of the new species is included entirely in that of *A. nodosus* (32–36°S approximately), which was identified as the sister taxon according to molecular phylogenetic analyses. Moreover, both species are sympatric in the type locality. The

FIGURE 4. Bayesian inference consensus tree (50% majority-rule) showing the position of the new species, *Alsodes cantillanensis* (in bold), within the genus. Numbers next to the nodes, separated by slashes, correspond to bootstrap values of maximum parsimony analysis, treating gaps as a fifth state, and posterior probability of Bayesian inference, respectively.

Discussion

Morphological characteristics and molecular evidence show that the new taxon belongs to the genus *Alsodes*. The holotype and the paratype show two diagnostic characters of the genus: bilateral thorny patches in the chest of males (only in the holotype and apparently not completely developed) and absence of external tympanic ring (Gallardo 1970; Cuevas 2008). The molecular evidence shows that the new species groups with two other endemic species from central Chile, *A. nodosus* and *A. vanzolinii*, distributed between 32 and 38°S, and it is not related directly with the Andean species *A. hugoi* and *A. tumultuosus*, distributed between 33 and 36°S.

According to the phylogenetic analyses, *A. cantillanensis* and *A. nodosus* are sister species. Nevertheless, this relationship is not accompanied by a morphological similarity between both species. In fact, some characteristics of the new species, like the continuous distribution of the spines of the chest patches, the golden-yellowish coloration of the superior part of the iris, the triangle on the head and the smooth skin, are observed in some species from southern Chile and Argentina (37–49°S), but not shared with *A. nodosus*. These two species also show an unusual geographic distribution pattern within the genus, which is mainly comprised by allopatric species (Cuevas & Formas 2003). The type locality of *A. cantillanensis* is completely surrounded by localities of *A. nodosus*, where a population of this last species also was found. In fact, both species inhabit the same creek and reproduce apparently in the same ponds.

546 · *Zootaxa* 3915 (4) © 2015 Magnolia Press · CHARRIER *ET AL.*

Figure 6: Article associated to the Genbank submission (note that the title of the article is different to the indicated in NCBI).

6. Another way to retrieve the geographic coordinates is to search for the name of the locality in **Google Earth** (e.g., Rinconada de Idahue), and obtain the geographic coordinates directly from there (Figure 7). Google Earth gives the coordinates in DMS (degrees, minutes and seconds) but for our analyses we are using DD (decimal degrees) latitude and longitude format for geographic coordinates (Table 1).

Table 1. Geographic coordinates of a locality in DMS and DD.

Locality	DMS coordinates	Latitude	Longitude
Rinconada de Idahue	34°17'28"S; 71°08'42"W	-71.145	-34.291

Fortunately, there are several websites to change the format of coordinates, we are using https://www.gps-coordinates.net/ which is easy to use and with an acceptable interface (Figure 8).

Finally, if all we have is the name of the locality, or even only the name of a County in the case of US species, we can get approximate coordinates using free geographical databases such as GeoNames

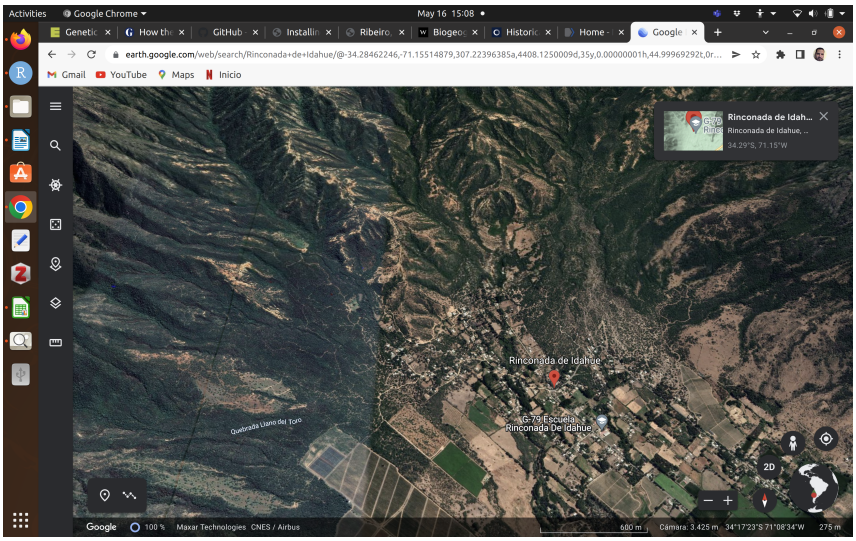


Figure 7: Google Earth interface, geographic coordinates can be found in the bottom right of the screen.

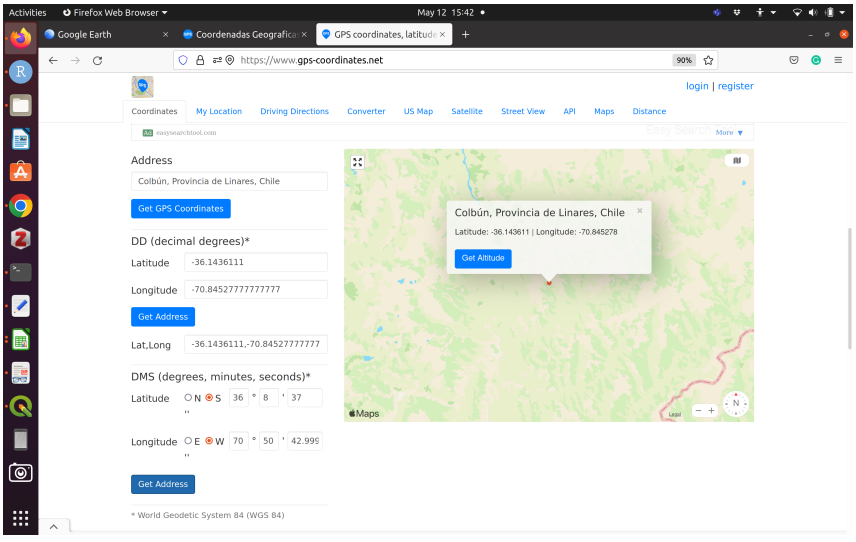


Figure 8: Example of tool where we can format the geographic coordinates.

(https://www.geonames.org) or GEOLocate (https://www.geo-locate.org/web/WebGeoref.aspx). For instance, GEOLocate (Figure 9; A Platform for Georeferencing Natural History Collections Data) can be used clicking in 1) 'Web Application', 2) 'Standard Client' to open the GEOLocate Web Application.

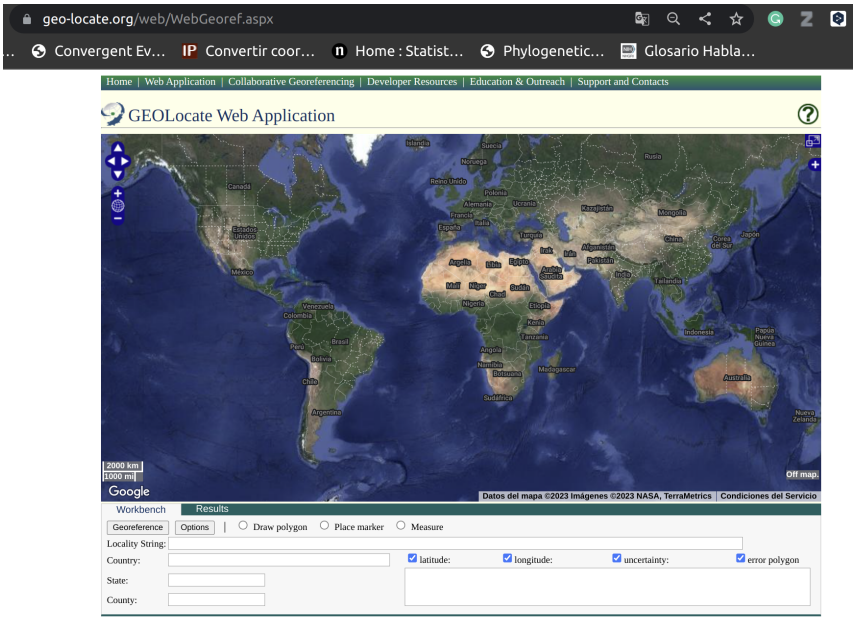


Figure 9: GEOLocate Web Application.

IN GEOLocate we need write the name of the locality in 'Locality String' and the name of the Country, in the example the name is "Loma Alta" which is a natural reserve in western Ecuador. After that we can see a green point that indicates where the locality was recovered by GEOLocate and the coordinates in latitude and longitude format that we can extract directly (Figure 10). The red point could be representing another locality in Ecuador named 'Loma Alta' as well, but the red color indicates that it may be wrong.

There is also the option of move the green point with the mouse if we need or want change the locality to a another more precise position based on a priori information.

In the case we have just the name of the county for US species, we have to write the name of the county how is showed in the figure 11, the Country and the name of the State. We have the option of edit the uncertainty which we can include in another column next to our coordinates.

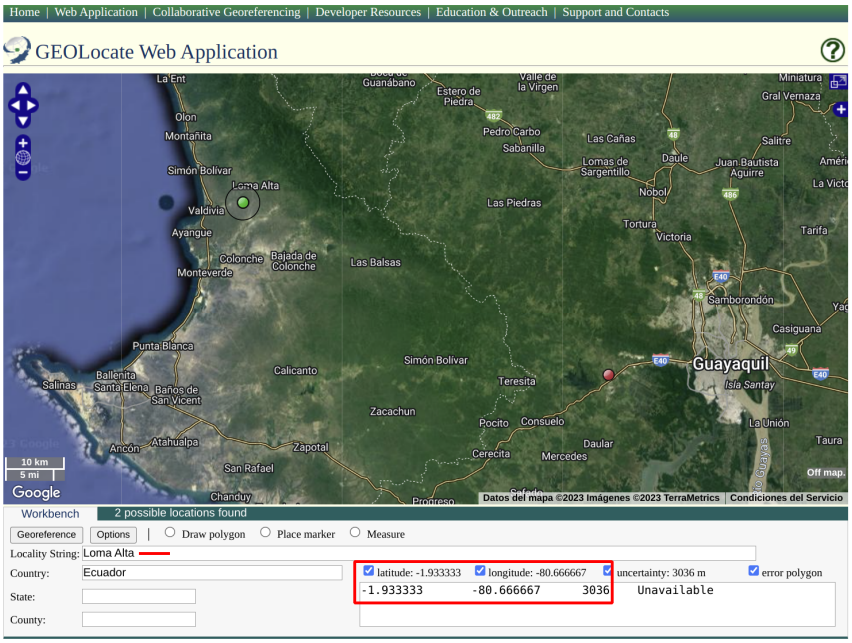


Figure 10: Visualization of a search in GEOLocate indicating the name of the locality and the coordinates retrieved.

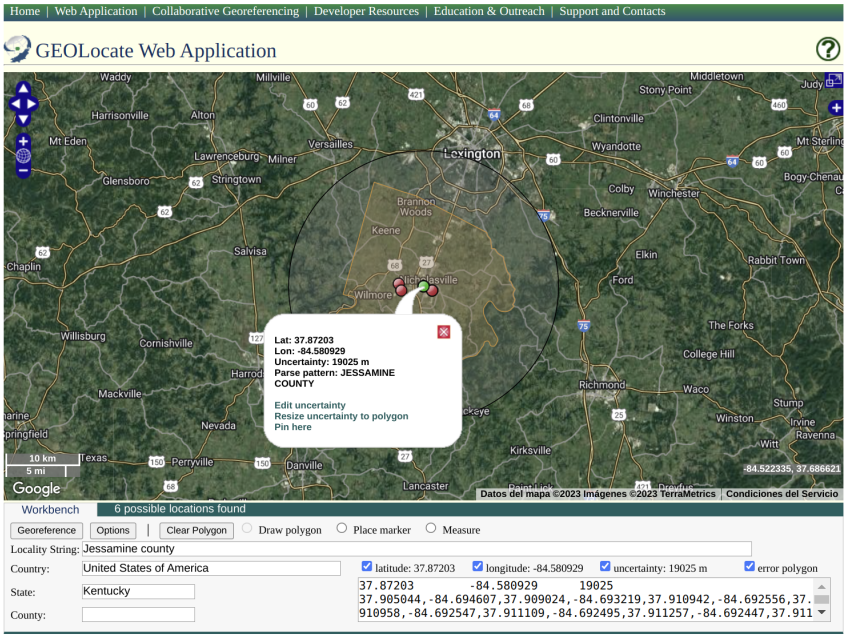


Figure 11: Visualization of a search in GEOLocate with an example of a species with the locality in an county from USA.