

VectorBase and VEuPathDB needs for mapping geographical locations

ESIP SemTech Meeting 2020-03-24

Bob MacCallum

Sam Rund

Chris Stoeckert

VEuPathDB is an NIAID Bioinformatics Resource Center covering eukaryotic pathogens and invertebrate vectors of disease

VectorBase is part of VEuPathDB

VectorBase provides population biology datasets that include geographical locations [visualized on a map](#) (MapVEu).

Other VEuPathDB sites will be using MapVEu for pathogen isolate datasets.

ClinEpiDB is associated with VEuPathDB providing clinical epidemiological datasets that will also use MapVEu.

Standardized geographical locations are needed for MapVEU.

Details for selected samples

Anopheles walkeri

Sample VBS0455963 (pool of 12 individuals)

Col. Date(s) Location
 Mon, May 19, 2014 Lucas (Ohio (United States))

Project(s) Col. Protocol(s)
 VBPO000316 new jersey trap catch

Attractant(s)
 light

Abundance data

Abundance: 12 individual(s) per day

12 individual(s) Protocol(s)
 1 day(s) morphological examination

Sex Development stages
 female adult

Aedes vexans sensu lato

Sample VBS0455964 (pool of 1 individuals)

Col. Date(s) Location
 Mon, May 19, 2014 Lucas (Ohio (United States))

Project(s) Col. Protocol(s)
 VBPO000316 new jersey trap catch

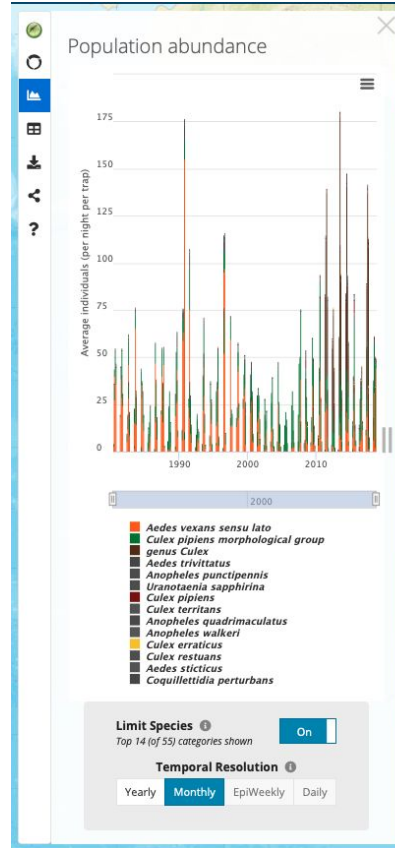
Attractant(s)
 light

Abundance data

Abundance: 1 individual(s) per day

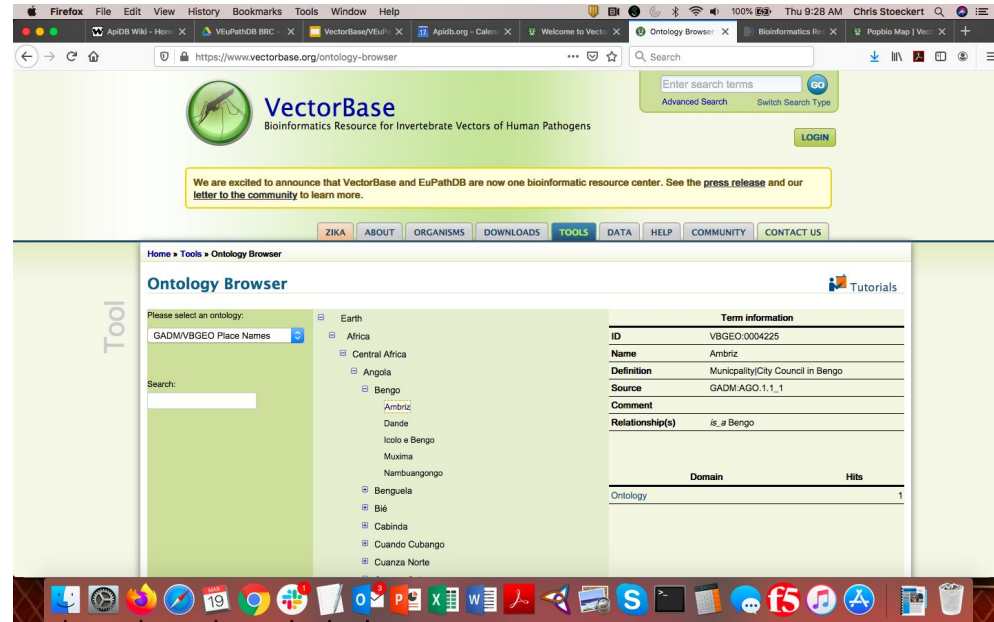
1 individual(s) Protocol(s)
 1 day(s) morphological examination

Sex Development stages
 female adult



Current and desired representation of locations

- Currently use a VectorBase in-house representation of GADM.
- GADM provides a shape-based representation of locations facilitating mapping of GPS coordinates.
- Need a hierarchy to provide generalization - de-identification.
- Want to be interoperable with VEuPathDB ontology terms used for representing samples, collections, study design, and other details.
- VEuPathDB ontology is based on OBO Foundry ontology terms



The screenshot displays the VectorBase Ontology Browser interface. The top navigation bar includes links for ZIKA, ABOUT, ORGANISMS, DOWNLOADS, TOOLS, DATA, HELP, COMMUNITY, and CONTACT US. The main content area is titled "Ontology Browser" and features a search bar and a hierarchical tree of geographical locations. The tree is expanded to show the following structure:

- Earth
 - Africa
 - Central Africa
 - Angola
 - Bengo
 - Ambriz
 - Dande
 - Icolo e Bengo
 - Muxima
 - Nambugangongo
 - Benguela
 - Bie
 - Cabinda
 - Cuando Cubango
 - Cuanza Norte

On the right side, the "Term information" table provides details for the selected term:

Term information	
ID	VBGEO:0004225
Name	Ambriz
Definition	Municipality/City Council in Bengo
Source	GADM-AGO.1.1.1
Comment	
Relationship(s)	is_a Bengo
Domain	Hits
Ontology	1

VectorBase's journey from GAZ to GADM

- Manually assigned GAZ terms to collection site records until 1 y ago
 - Error-prone: synonymous place names lead to curation errors... (which Santa Maria?)
 - No GPS info in GAZ assist look-ups
 - Benefits: high level search queries (e.g. "West Africa") find records annotated with lower-level terms (e.g. Ougadougou)
- GADM
 - comprehensive coverage of polygons down to ADM2 for everywhere except small islands
 - available in consistent files/format
 - open access
- Converted GADM country->ADM1->ADM2 to ontology-like file
- Added manually curated top level terms (e.g. "West Africa")

New VectorBase/VEuPath curation process

- Curators ensure accurate lat/long information only
- Scripted assignment of “GADM” ADM2 terms by polygon lookup
- That’s it!

(A very tiny number of collection sites are unassignable due to being too far out to sea - there’s a X km fuzzy factor used in the lookup.)

Future uses of GADM/geo information

PopBio map aka MapVEU

Currently display aggregated point data as circular markers

Polygon markers (e.g. choropleth) would be nice to add.

Not wedded to GADM but so far no problems with it.

Other

?

Discussion topics

What are ENVO, an OBO Foundry ontology, plans for shape-based representation of geographical locations?

GAZ, an OBO Foundry ontology representing places through their names, appears to be under ENVO development. What are the plans for it?

Use of GeoJSON-LD?