Terrestrial Parasite Tracker Data and Collections-Based Research: Updates from the TPT Research Advisory Board and Other Collaborators

TPT Webinar March 31st
2 pm - 3 pm EST
Terrestrial Parasite Tracker TCN

- Transcribe and georeference label data from 1.2+ million arthropod parasite specimens from 22 collections across North America (U.S. and territories)
- Document 500,000+ parasite-host associations via GloBI
TPT Research Advisory Board

- RAB is comprised of researchers from academic and non-academic institutions
- Working to define project goals and conditions for collaboration (e.g., how to handle student data, data requested for proposals, publications, etc.)
- Developing digitization priorities for research publications (currently ticks, flies, lice)
CRITICAL COMMENT . . . . .

A MISIDENTIFICATION CRISIS PLAGUES SPECIMEN-BASED RESEARCH: A CASE FOR GUIDELINES WITH A RECENT EXAMPLE (ALI ET AL., 2020)

Sarah E. Bush¹, Daniel R. Gustafsson², Vasyl V. Tkach³, and Dale H. Clayton¹
Better Understanding of Parasite Geographic Distributions

• Often assume that lice are **host specific** and **not geographically specific** - parasite distributions aren’t that simple

• We **know very little** about louse geographic distributions

• Localities often poorly resolved on louse slide labels but maybe be can get **more detailed localities** from associated host specimens

• Digitization will help us to construct a picture of parasite geographic specificity with respect to hosts

Guimaraesiella cicchinoi
A widespread New World Trogon parasite
Using Artificial Intelligence to Develop a Taxonomy Guide
Creating an AI model with 14 louse genera found on poultry

**Ischnocera:**

1. Campanulotes
2. Chelopistes
3. Coloceras
4. Columbicola
5. Cuclotogaster
6. Goniocotes
7. Goniodes
8. Lagopoecus
9. Lipeurus
10. Oxylipeurus

**Amblycera:**

11. Colpocephalum
12. Hohorstiella
13. Menacanthus
14. Menopon

Photos: Ricardo Palma
• Ticks are vectors of:
  - Rocky Mountain Spotted Fever
  - Tularemia
  - Ehrlichiosis
  - Spotted fever rickettsiosis
  - Babesiosis
  - Lyme disease
NSF funded Digi-Leap project
Goal: to increase the rate of transcription and georeferencing

- Adding in machine learning steps to Notes from Nature
- Working collaboratively with volunteers to vet and quality control the data coming from OCR readers
- Information Extraction ML process to parse and identify
NSF funded Digi-Leap project
Goal: to increase the rate of transcription and **georeferencing**

- Millions of specimens have been georeferenced
- Many collectors visit the same areas - producing the same localities
- Building a gazetteer of all georeferenced from GBIF, iDigBio
- New localities can be compared with the gazetteer - reduce the number that need to be georeferenced
Leveraging Big Data to Improve Prediction of Tick-borne Disease Patterns and Dynamics

Mike Teglas, University of Nevada Reno

Guide to the Surveillance of Metastriate Ticks (Acari: Ixodidae) and their Pathogens in the United States, CDC
Tick Crawler
- Tick data acquisition from the Internet/Scientific Literature

Field Collections
- Project personnel, county and state agencies

Historical tick data
- Museums and Collections

Forecasting Tick-borne Diseases in the Western U.S.

TickBase News
Species Distribution Modeling of North American Tabanids

Sophia Zaslow
Collaborators: Emily Sandall and Maureen Turcatel

Map of Tabanid distribution recorded via GBIF [https://www.gbif.org/species/6919](https://www.gbif.org/species/6919)

Map of Ecuador depicting the potential distribution of Chrysops varians var. tardus

Image of Tabanus rubidus

Standard set up for Tabanid photographing
Ingest

Name lists for the various groups have been received in the form of spreadsheets, csv files, and text documents.

Clean

Develop a reproducible, non-destructive process in R which:

1) Combines datasets relating to the same taxonomic groups

2) Cleans the resulting lists to flag duplication, misspellings and other possible errors

Review

Cleaned data will be reviewed by our team for easily resolved issues

Issues which cannot be easily resolved will be sent to the appropriate expert(s) for review

Publish

Combine taxonomy into a single “TPT Taxonomy” dataset

Publish to the Global Names resource

Maintain

As taxonomy is not static, there will need to be periodic maintenance of the dataset

We are investigating tools to facilitate the long-term maintenance and continued updating of the dataset by group expert(s)
<table>
<thead>
<tr>
<th>Step</th>
<th>Task</th>
</tr>
</thead>
</table>
| 01   | **Standardize**  
  - Fix formatting of names and author strings (e.g., capitalization)  
  - Remove unresolvable names (sp, unidentified, probably, nov sp, etc.)  
  - Cast column headers into Darwin Core format |
| 02   | **De-Duplicate**  
  - Remove duplicate names  
  - Flag names with very similar and/or short epithets |
| 03   | **Complete**  
  - Add missing higher taxa  
  - Add author citation  
  - Add synonyms |
| 04   | **Review**  
  - TPT Team review of flagged names  
  - Taxonomic expert review of remaining flagged names |
| 05   | **Finalize**  
  - Combine resolved names into working taxonomy file  
  - Track remaining rejected names in a separate list  
  - Provide TPT taxonomy to Global Names |
How do their contents differ?

Have we captured everything?

Derived from GBIF backbone snapshot taken on March 26, 2021
Digitized Specimen Data = Entomological Intelligence

Information about what specimens are available for study,.vouched distribution records and host associations

Type locality:
"Carolina" (exact site and host unknown)

United States National Tick Collection (USNTC) Specimens:
Males: 1673
Females: 1055
Nymphs: 1940
Larvae: 4103

Specimens documented from the following states:
Alabama, Arizona, Arkansas, District of Columbia, Florida, Georgia, Indiana, Kansas, Louisiana, Mississippi, Montana, New Mexico, North Carolina, Oklahoma, South Carolina, Texas, Texas, Virginia

Geo-referenced collection sites can drive ecological niche models, providing situational awareness

Making data available via EMU allows for integration with other DB
https://globalbioticinteractions.org/parasitetracker

Indexing and Reviewing Research Datasets

<table>
<thead>
<tr>
<th>TPT Collections Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Click on badges to browse/download indexed records or inspect automated reviews.</td>
</tr>
<tr>
<td><strong>edit collection list</strong></td>
</tr>
<tr>
<td><strong>status</strong></td>
</tr>
</tbody>
</table>
Creating Research Datasets: Directed Research in Data Science for Undergraduates

Understandable by people and computers

Link to taxonomic treatments

Data review with GloBI

Bee Interaction & Tick Interaction Databases
A Practical Exploration of Biotic Interaction Data Management and Information Retrieval through TPT and GloBI

Register at iDigBio: https://www.idigbio.org/content/practical-exploration-biotic-interaction-data-management-and-information-retrieval-through