

iDigBio Research Tools

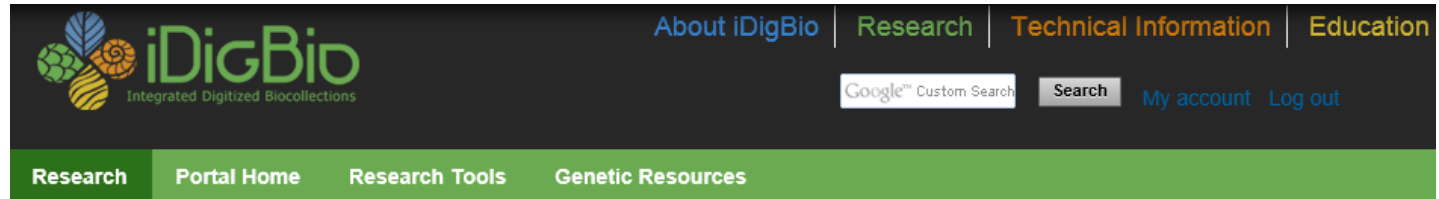
Andréa Matsunaga (ammatsun@ufl.edu)

Cyberinfrastructure Working Group (CYWG) meeting
October 10, 2014



iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. All images used with permission or are free from copyright.

iDigBio Research Section



Research

Looking for research ideas?

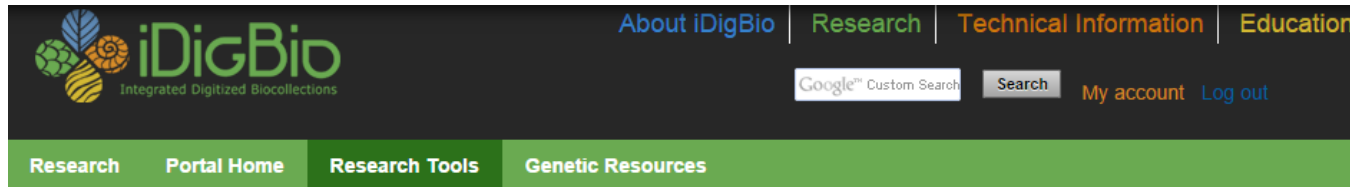
Read about the research questions proposed by each TCN:

2011	<ul style="list-style-type: none">• InvertNet• Tri-trophic• Lichens & Bryophytes (LBCC)
2012	<ul style="list-style-type: none">• New England Vascular Plants (NEVP)• PaleoNICHES• Macrofungi Collection Consortium (MaCC)• Southwest Collections of Arthropods Network (SCAN)
2013	<ul style="list-style-type: none">• Fossil Insect Collaborative (FIC)• Vouchered Animal Communication Signals (VACS)• Macroalgal Herbarium Consortium (MHC)
2014	<ul style="list-style-type: none">• Great Lakes Invasives• InvertEBase• SouthEast Regional Network of Expertise and Collections (SERNEC)

Links to TCN research
List of iDigBio publications

- Expanding: <https://www.idigbio.org/research>

iDigBio Research Tools



Community Research Tools

To facilitate the study of biodiversity, a number of research tools are being developed to take advantage of the data being digitized at US institutions and made available by iDigBio through **web services**. You can find below some of these online tools developed by the community. If you would like your tool to be included in this list, please use the **feedback form** to tell us about your work.

Researchers

Browse our specimen portal



Collections Staff

Learn how your collection can benefit from our work



Teachers & Students

Learning resources & opportunities to engage



List of Tools Integrating iDigBio Web Services

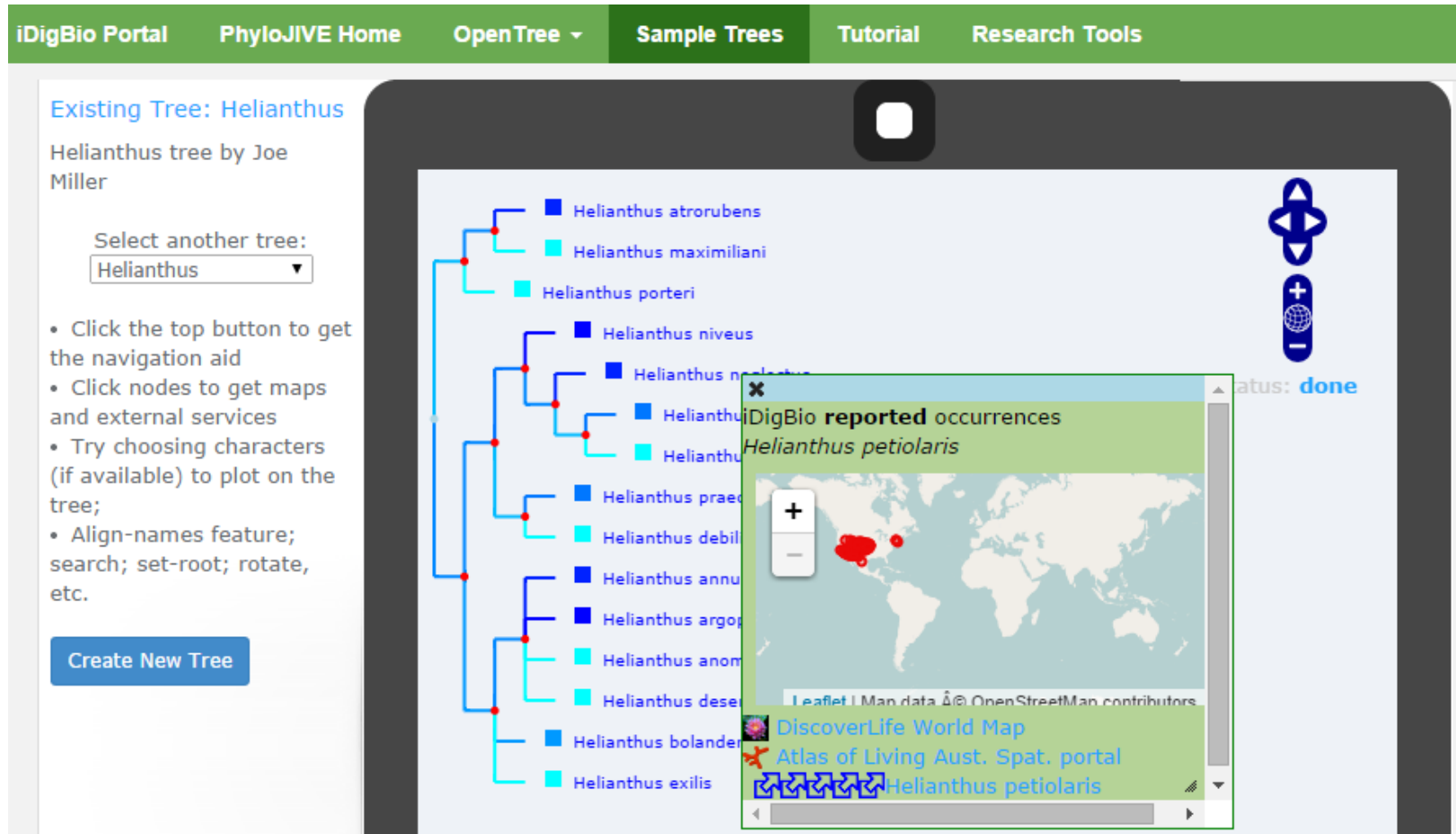
PhyloJIVE (Phylogeny Javascript Information Visualiser and Explorer)



Solutions to fundamental questions about biodiversity require a new approach that integrates across phylogeny, biogeography, geology, and paleobiology. **PhyloJIVE**, developed by Garry Jolley-Rogers, Joe Miller, and Temi Varghese, integrates biodiversity data with phylogeny. Through **PhyloJIVE**, occurrence records can be viewed in a phylogenetic context, and user-supplied character data can be visualized on the phylogeny. Exploration of the linkages between phylogeny, distributions, and character states can lead to new understanding of biodiversity.

- <https://www.idigbio.org/content/community-research-tools>
- Welcome your contributions!

PhyloJIVE instance in iDigBio

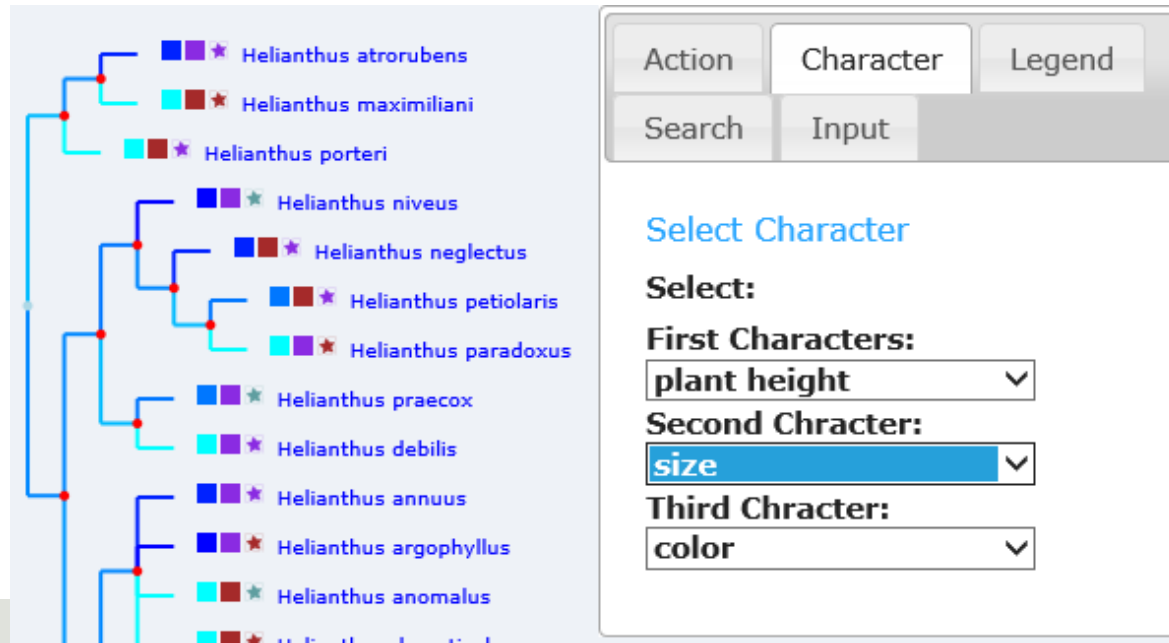


The screenshot displays the iDigBio Portal interface for the PhyloJIVE instance. The navigation bar includes links for iDigBio Portal, PhyloJIVE Home, OpenTree, Sample Trees, Tutorial, and Research Tools. The main content area shows an existing tree titled "Existing Tree: Helianthus" by Joe Miller. A dropdown menu allows selecting another tree from the "Helianthus" genus. A list of species is shown on the right, including Helianthus atrorubens, Helianthus maximiliani, Helianthus porteri, Helianthus niveus, Helianthus mollis, Helianthus scaberrimus, Helianthus debilis, Helianthus annuus, Helianthus argophyllus, Helianthus anomus, Helianthus desertorum, Helianthus bolanderi, and Helianthus exilis. A map window is open, showing "iDigBio reported occurrences" for Helianthus petiolaris, with red dots indicating locations in the western United States. The map includes a zoom control and a status bar at the bottom with links to DiscoverLife World Map, Atlas of Living Aust. Spat. portal, and Helianthus petiolaris.

- Developed by Garry Jolley-Rogers, Joe Miller, and Temi Varghese
- Integrates biodiversity data with phylogeny.
- <http://phylojive.acis.ufl.edu/>

PhyloJIVE Features

- Displays phylogenetic trees in Newick format
- Displays up to 10 characters (traits) in charJSON format using color scale to indicate numerical intensity or categories
- Tree branches are colored according to the predicted value for the first character, as calculated using reverse parsimony

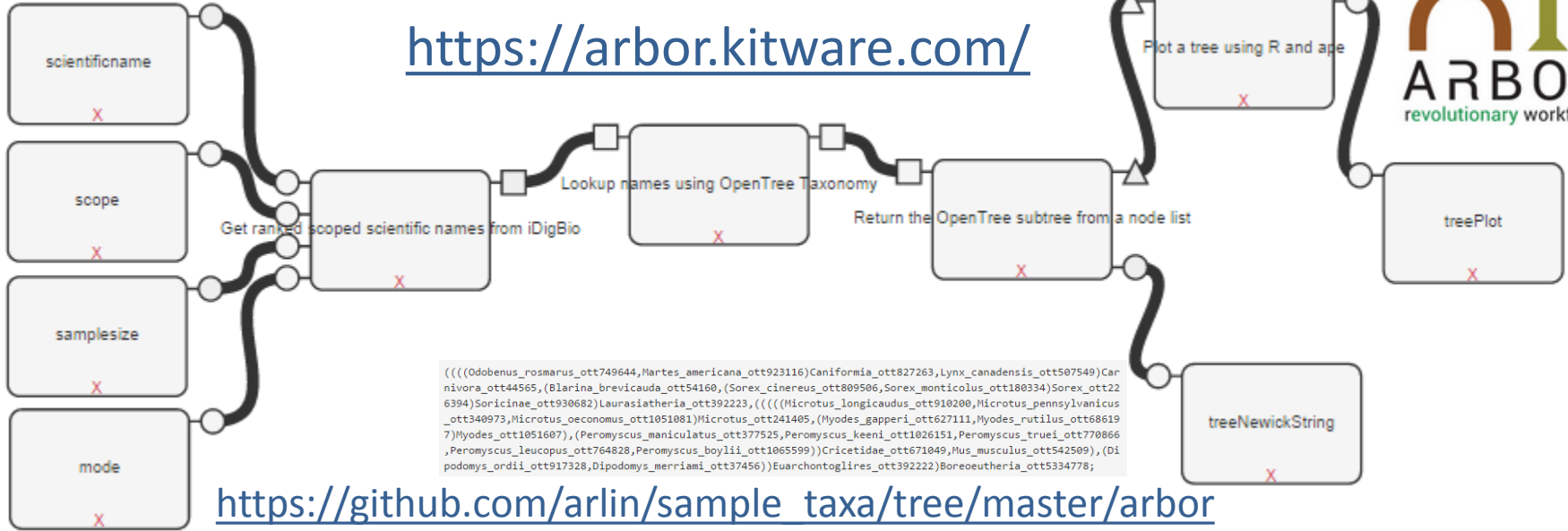


Unique features in iDigBio instance

- Integrated with iDigBio portal search and map per species
 - https://www.idigbio.org/wiki/index.php/IDigBio_API
- Linkage to other third party sites (ALA taxonomy, EOL, DiscoverLife, and ALA spatial portal)
- Select number of plant studies from OpenTree can be dynamically accessed via APIs
 - <https://github.com/OpenTreeOfLife/opentree/wiki/Open-Tree-of-Life-APIs>
- Canned searches on iDigBio (search API) for top or random species names that are used to produce an induced tree using OpenTree APIs
- Sample trees: from Joe Miller, from TTD TCN, from Tree-for-all hackathon
- Create your own tree and character set
- Tutorial

Arbor, OpenTree, and iDigBio

<https://arbor.kitware.com/>



https://github.com/arlin/sample_taxa/tree/master/arbor

Workflow to get an induced tree from a configurable iDigBio query

scientificname:

scope:

samplesize:

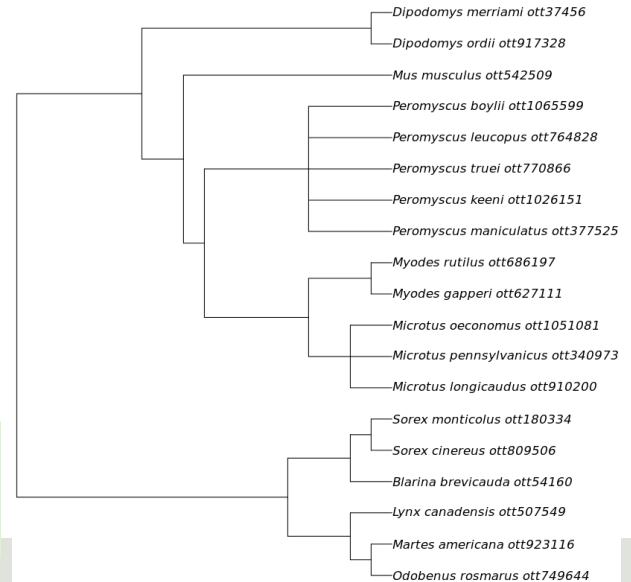
mode:

```

1 {
2   "query": {
3     "query_string": {
4       "default_field": "order",
5       "query": "rodentia"
6     }
7   },
8   "aggregations": {
9     "my_agg": {
10      "terms": {
11        "field": "scientificname",
12        "size": 100
13      }
14    }
15  }
16 }
    
```

Success! Produced the following outputs:

- Workflow to get an induced tree from a configurable iDigBio query treeNewickString [string]
- Workflow to get an induced tree from a configurable iDigBio query treePlot [image]



PhyloJIVE, OpenTree, and iDigBio

OpenTree ▾ Sample Trees

OpenTree Studies

iDigBio Frequency

Query

```

1 {
2   "query": {
3     "query_string": {
4       "query": "mammalia"
5     }
6   },
7   "size": 0,
8   "aggregations": {
9     "my_agg": {
10      "terms": {
11        "field": "scientificname",
12        "size": 100
13      }
14    }
15  },
16  "filter": {
17    "exists": { "field": "geopoint" }
18  }
19 }

```

Response

```

"aggregations": {
  "my_agg": {
    "buckets": [
      {
        "key": "peromyscus maniculatus",
        "doc_count": 36597
      },
      {
        "key": "peromyscus leucopus",
        "doc_count": 20699
      },
      {
        "key": "myodes rutilus",
        "doc_count": 19017
      }
    ]
  }
}

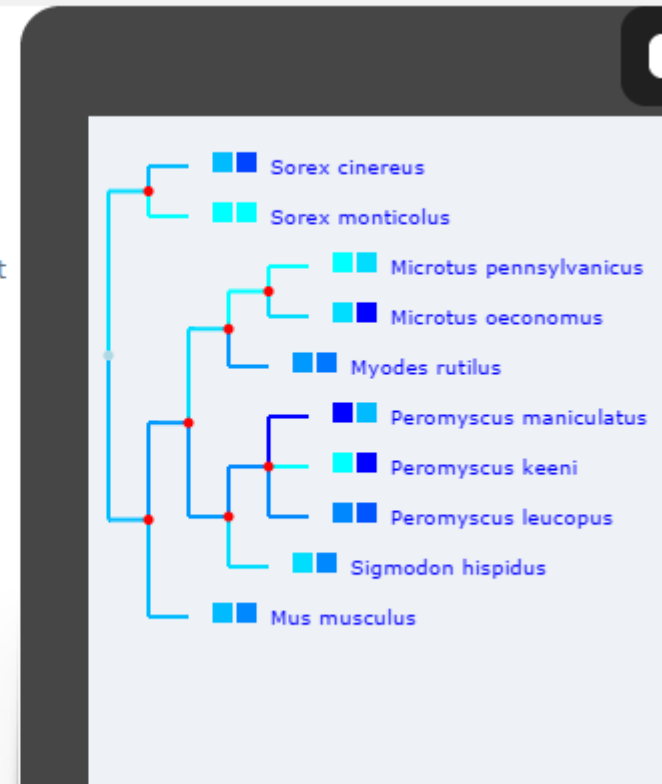
```

Criteria:
top_10_Mammalia

Select another criteria:

top_10_Mammalia ▾

- top_10_Diptera
- top_50_Diptera
- top_10_Felidae
- top_50_Felidae
- top_10_Insecta
- top_50_Insecta
- top_10_Mammalia
- top_50_Mammalia
- top_10_Nematoda
- top_50_Nematoda
- top_10_Rodentia
- top_50_Rodentia
- top_10_Rupchand
- rand_50_Diptera
- rand_50_Felidae
- rand_50_Insecta
- rand_50_Mammalia
- rand_50_Nematoda
- rand_50_Rupchand
- rand_50_Rodentia



- Searching for any occurrence of the word “Mammalia” with lat/long
- Selects the top or random scientific names occurring in iDigBio
- Keep only binomials, ignore names with certain special characters
- <http://search.idigbio.org/idigbio/records/search>

OpenRefine, OpenTree and iDigBio

Google refine *A power tool for working with messy data.*

Create Project
Open Project
Import Project

Create a project by importing data. What kinds of data files can I import?
TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with Google Refine extensions.

Get data from

This Computer
Web Addresses (URLs)
Clipboard
Google Data

Enter one or more web addresses (URLs) pointing to data to download:

Add Another URL Next »

1. Create a new project from "Web addresses"

http://search.idigbio.org/idigbio/records/_search?source={%22size%22:0,%22query%22:{%22query_string%22:{%22query%22:%22mammalia%22}},%22aggregations%22:{%22my_agg%22:{%22terms%22:{%22field%22:%22scientificname%22,%22size%22:10}}}}

2. Parse as JSON and select the my_agg node

OpenRefine, OpenTree and iDigBio

Google refine *A power tool for working with messy data.*

« Start Over Configure Parsing Options

Create Project

Open Project

Import Project

	my_agg - buckets - __anonymous__ - key	my_agg - buckets - __anonymous__ - doc_count
1.	peromyscus maniculatus	36597
2.	peromyscus leucopus	20699
3.	myodes rutilus	19017
4.	sorex cinereus	16331
5.	mus musculus	14354
6.	peromyscus maniculatus sonoriensis	13185
7.	mammalia	12675
8.	microtus oeconomus	12230
9.	sigmodon hispidus	12047
10.	peromyscus leucopus noveboracensis	11909
11.	peromyscus maniculatus nebrascensis	10753
12.	dipodomys merriami merriami	10192
13.	peromyscus maniculatus rufinus	10185
14.	peromyscus maniculatus gambelii	10130
15.	microtus pennsylvanicus	9855
16.	peromyscus leucopus	8623

Parse data as

- CSV / TSV / separator-based files
- Line-based text files
- Fixed-width field text files
- PC-Axis text files
- JSON files

Load at most 0

Store file source (file names, URLs) in each row

3. Create new column to hold OTT IDs

```
import tnrs
return tnrs.getOttId(value)
```

4. Filter out monomials

5. Generate a tree visualization

Opentree induced sub-tree from Google Refine data



Next steps

- List your research tools
- Get your phylogenetic tree displayed
- Additional uses of iDigBio APIs

Questions?

Feedback: https://www.idigbio.org/contact/PhyloJIVE_feedback



www.idigbio.org



facebook.com/iDigBio



twitter.com/iDigBio



vimeo.com/idigbio



idigbio.org/rss-feed.xml



<webcal://www.idigbio.org/events-calendar/export.ics>

Thank you!