Building a Global Consortium of Bryophytes and Lichens: Keystones of Cryptobiotic Communities (GLOBAL)

Jessica Budke & Matt von Konrat

- Collaborators: 25 US Herbaria
- Project Start Date: 15 September 2021
- Project Duration: 3 years
GLOBAL Goals

I. Establish a novel cryptobiotic consortium integrating 6M records.
GLOBAL Goals

I. Establish a novel cryptobiotic consortium integrating 6M records.

II. Digitize label data and specimens for 1.2M bryophytes/lichens focusing on non-North American specimens from 25 US herbaria.
GLOBAL Goals

I. Establish a novel cryptobiotic consortium integrating 6M records.

II. Digitize label data and specimens for 1.2M bryophytes/lichens focusing on non-North American specimens from 25 US herbaria.

III. Create a connected world: Innovative automation, integration, image tagging, and machine learning.
GLOBAL Goals

I. Establish a novel cryptobiotic consortium integrating 6M records.

II. Digitize label data and specimens for 1.2M bryophytes/lichens focusing on non-North American specimens from 25 US herbaria.

III. Create a connected world: Innovative automation, integration, image tagging, and machine learning.

IV. Focus on public engagement and education.
GLOBAL Digitization Progress

- # Barcodes Added
- # Labels Imaged
- # Specimens Imaged
- # Uploaded to Portal
- # Skeletal Records Created
- # Fully Transcribed
- # Georeferenced

Legend:
- 2021-Q2
- 2021-Q1
- 2020-Q4
- 2020-Q3

Types of Specimens:
- Bryophytes
- Lichens
Acknowledge the Traditional Custodians of the land on which we virtually gather today, and pay my respects to their Elders, past and present. I extend that respect to Indigenous and First Nation people all around the world that are joining us today.
Building a Global Consortium of Bryophytes and Lichens: Keystones of Cryptobiotic Communities

I. Establish a novel cryptobiotic consortium integrating 6M records.

II. Digitize label data and specimens for 1.2M bryophytes/lichens focusing on non-North American specimens from 25 US herbaria.

III. Create a connected world: Innovative automation, integration, image tagging, and machine learning.
Added to CNABH or CNALH

- 21 new international herbaria
- 9 new countries
- 972,743 new bryophyte specimen records
- 998,083 new lichen specimen records
Duplicate Matching

Duplicate Coordinate R Script
- Run for 13 collections
- Julie and Katie cleaned up lists of possible duplicates
- 8,000 imported to portals (30,000+ additional awaiting review)

International Collections Data
- 4 more international collections added to the portals in 2021-Q2
- 21 total collections, 970,000+ Bryophyte and 980,000+ Lichen records

Exsiccatae List and Field Updates
- ASU working on protocols to add exsiccati identifier for Snapshots
- Blanka cleaned 1/5 of the exsiccatae library on the Bryophyte Portal
GloBaL: Deep-learning approaches

- taxonomy
- natural history collection management
- species identification, co-occurrences

Accuracy
Training: 100%
Validation: 86%
Convolutional Neural Networks (CNN)

Step 1: Train NN
Step 2: Test & use NN

Epochs - rounds of training (5 epochs = each training image is used 5 times)
Results

Frullania rostrata

undescribed species

CNN: ~91% accuracy in differentiating species
Unlocking handwritten text using Machine learning solutions

Existing tools can get us started...
CNN-RNN-CTC + handwritten word images

Raw Google Cloud images

Manually cleaned

Thresholding

Cropped whitespace

Padded all words to the same image size
### CNN-RNN-CTC initial models

626 training images, 70 testing images

<table>
<thead>
<tr>
<th>Random seed</th>
<th>Run #</th>
<th>prediction</th>
<th>prediction</th>
<th>prediction</th>
<th>prediction</th>
<th>prediction</th>
<th>prediction</th>
<th>prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>Howell</td>
<td>Soaldt</td>
<td>Nol</td>
<td>Hoel</td>
<td>Howell</td>
<td>Howell</td>
<td>Duriers</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>T53</td>
<td>TSW</td>
<td>T23</td>
<td>T13</td>
<td>T53</td>
<td>T53</td>
<td>T2,</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Stone</td>
<td>Sian</td>
<td>Shon</td>
<td>Sre</td>
<td>Stone</td>
<td>Stone</td>
<td>Proek</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Camptosorus</td>
<td>Cart</td>
<td>Patu</td>
<td>Cop</td>
<td>Camptorus</td>
<td>Camptosorus</td>
<td>leigaCaolu</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Hollow</td>
<td>Giro</td>
<td>bl</td>
<td>Slo</td>
<td>Hollow</td>
<td>Hollow</td>
<td>blufan</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Creek</td>
<td>Creek</td>
<td>Creek</td>
<td>Creek</td>
<td>Creek</td>
<td>Creek</td>
<td>Creek</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>northwest</td>
<td>northest</td>
<td>at</td>
<td>torhet</td>
<td>northwest</td>
<td>northwest</td>
<td>arguats</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>at</td>
<td>t</td>
<td>of</td>
<td>at</td>
<td>at</td>
<td>S.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>miles</td>
<td>sess</td>
<td>mily</td>
<td>mile</td>
<td>miles</td>
<td>miles</td>
<td>nirth</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>Shaded</td>
<td>Sec</td>
<td>Sec</td>
<td>Shded</td>
<td>Shded</td>
<td>Shded</td>
<td>limelus</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>margin</td>
<td>magy</td>
<td>mrne</td>
<td>ay</td>
<td>margin</td>
<td>margin</td>
<td>margvgc,</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>658</td>
<td>T5</td>
<td>Aune</td>
<td>368</td>
<td>658</td>
<td>658</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>Willow</td>
<td>Can</td>
<td>bl</td>
<td>tar</td>
<td>Willow</td>
<td>Willow</td>
<td>bats</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>Watt</td>
<td>Wat</td>
<td>Mot</td>
<td>Mect</td>
<td>Mear</td>
<td>Wat</td>
<td>Oect.</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>mi.</td>
<td>mi.</td>
<td>mi.</td>
<td>mi.</td>
<td>mi.</td>
<td>mi.</td>
<td>we.</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>May.6,1939</td>
<td>No,619</td>
<td>Mar,9</td>
<td>Mar,13</td>
<td>May,6,19</td>
<td>May,6,19</td>
<td>narpne</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>mi.</td>
<td>mi.</td>
<td>mid</td>
<td>mi.</td>
<td>mi.</td>
<td>mi.</td>
<td>nia</td>
</tr>
</tbody>
</table>
Python 3 with Tensorflow 2
Google Colab (~Jupyter Notebook with virtual GPU)
Building a Global Consortium of Bryophytes and Lichens: Keystones of Cryptobiotic Communities

I. Establish a novel cryptobiotic consortium integrating 6M records.

II. Digitize label data and specimens for 1.2M bryophytes/lichens focusing on non-North American specimens from 25 US herbaria.

III. Create a connected world: Innovative automation, integration, image tagging, and machine learning.

IV. Focus on public engagement and education.
Unfolding of Microplant Mysteries

Welcome! Bienvenidos! Bem-vinda! Namaste! Marhaba! Unlock mysteries, become a scientific detective!

Help discover the microscopic world of tiny plants

Learn more
Microplants, scientifically named bryophytes, are a group of green land plants. They include liverworts, mosses, and hornworts. There are an estimated 25,000 species in the world. They prefer damp habitats although some species can survive in drier environments. Bryophytes are non-vascular land plants, i.e., do not have a complex transport system. They are seedless plants and do not have flowers, seeds, or fruits. They have simple reproductive systems. Gametangia and spore cases are adapted for water movement.
Please identify if the image of the microplant shown best corresponds to a female, male, sterile, or both a female and a male structure.

Female
Male
Both Female and Male
Sterile

NEED SOME HELP WITH THIS TASK?

Done & Talk  Done

Thanks for volunteering your time and your efforts to help us collect data to advance the identification and classification of these microplants.

Please don't worry about a wrong
Acknowledgements

National Science Foundation
• Advancing Digitization of Biodiversity Collections

Collectors & Researchers of the past and present

2011 Bryophyte/Lichen TCN: Leadership & Participants
Especially, Tom Nash & Corinna Gries

Collaborating Herbaria: ALA, ASU, BRU, BRY, CINC, COLO, DUKE, F, FLAS, ILL/ILLS, LSU, MICH, MIN, MO, MSC, MU, NEB, NY, OSC, PH, TENN, UC, WIS, WTU, YU

CONTACT US: globaltcn@utk.edu    http://globaltcn.utk.edu    #GlobalTCN
GLOBAL Team

Aaron Liston
Aimee Davis
Alan Fryday
Alan Prather
Alison Staelgraeve
Amanda Chandler
Amy Harmon
Andrew Gallardo
Andy Miller
Ardath Dixon
Arthur J. Shaw
Barbara Thiers
Blanka Aguero
Brent Mishler
Brian Heitz
Brian Schuh
Cameron Lala
Campbell Webb
Carson Parham
Christina Mozzicato
Cristian Cartuche
Daniel Le
Daniel Stanton
Diego Inclan
Dina Clark
Edward Gilbert
Elana Benamy
Elizabeth H. Teresi
Emily Clark
Emma Kruse
Eric J. Tepe
Erin Tripp
Francois Lutzoni
Frank Bungartz
Gaby Mata
Garth Holman
George Weiblen
Ginger Apolo
J. Ryan Allen
Jack Haynes
Jaclyn Jones
Jake Burgoyne
Jake Henrie
James Lendemer
James Mickley
Jason Alexander
Jeffrey Clancy
Jennifer Kluse
Jessica Budke
Jessie Uehling
Jesus Hernandez
John Brinda
Jordan Teisher
Joshua Jacobs
Julie Smith
Katie Pearson
Kayla M. Wheatley
Kenneth Cameron
Kent Perkins
Kim Kersh
Kim Watson
Klára Scharnagl
Kyelle Corcoran
Laura Briscoe
Laura Lagomarsino
Lauren Keeling
Laurie Heldman
Leanna McMillin
Leanne Sheffer
Lilli Young
Lucas C. Majure
Lucas Ferguson
Madison Dyer
Madison Latiolais
Margaret Oliver
Maria L. Montero
Matt Chansler
Matt von Konrat
Maja Mahoney
Miranda Zwingelberg
Natalie Detwiler
Nico Franz
Nicolle Tineo
Nidhi Thirthamattur
Nkosi Evans
Olivia Leek
Olivia Taylor
Patrick Sweeney
Phillip Anders
Rosa Batallas
Scott LaGreca
Stefanie Ickert-Bond
Stephanie Zaborac-Reed
Steve Leavitt
Stuart McDaniel
Thomas Barlow
Thomas Nash
Thorsten Lumbsch
Timothy Hogan
Timothy James
Timothy Whitfield
Todd Widhelm
Tyson J. Thorpe
Wendy Quinlan
William A. Rivero
William Seely
Wyatt Gaswick
Yarency Rodriguez
Zoe Ryans
Zoey Molsberry
Questions?