Community collaboration to build an Open Tree of Life

Emily Jane McTavish

University of California, Merced
ejmctavish@ucmerced.edu, Twitter:@snacktavish

August 19, 2018

To make science a better endeavor in the future
than the past, we need
open access to research and data,
to develop
a **diverse and inclusive** community.

“Open science offers the promise of democratizing access and engagement in research,

"Open science offers the promise of democratizing access and engagement in research, but it doesn't guarantee that it will occur." - Nosek

(Nosek, 2017, How can we improve diversity and inclusion in the open science movement? Center for Open Science)

KLINT FINLEY BUSINESS 06.02.17 07:00 AM

DIVERSITY IN OPEN SOURCE IS EVEN WORSE THAN IN TECH OVERALL

KLINT FINLEY BUSINESS 06.02.17 07:00 AM

DIVERSITY IN OPEN SOURCE IS EVEN WORSE THAN IN TECH OVERALL

According to a 2016 survey:
16% of open source users and developers belong to an ethnic or national minority.
34% of professional computer programmers in the US belong to minority groups.

KLINT FINLEY BUSINESS 06.02.17 07:00 AM

DIVERSITY IN OPEN SOURCE IS EVEN WORSE THAN IN TECH OVERALL

According to a 2016 survey:
16% of open source users and developers belong to an ethnic or national minority.
34% of professional computer programmers in the US belong to minority groups.

2% of open source users and developers were women.
22.6% of professional computer programmers are women.
(Finley, Wired, June 2, 2016)

Some tenets of the Open Science movement run counter to efforts to eliminate barriers to success in the sciences.

Some tenets of the Open Science movement run counter to efforts to eliminate barriers to success in the sciences.

> Preprints and open source code can prevent double blind review

Some tenets of the Open Science movement run counter to efforts to eliminate barriers to success in the sciences.

- Preprints and open source code can prevent double blind review
- Open access fees can be a burden for early career researchers

Some tenets of the Open Science movement run counter to efforts to eliminate barriers to success in the sciences.

- Preprints and open source code can prevent double blind review
- Open access fees can be a burden for early career researchers
- Dogmatic attitudes are particularly unwelcoming to researchers from groups underrepresented in the sciences

If Open Science isn't effectively open to everyone, it is not achieving its aims.

If Open Science isn't effectively open to everyone, it is not achieving its aims.

And it is not as good science as it could be.

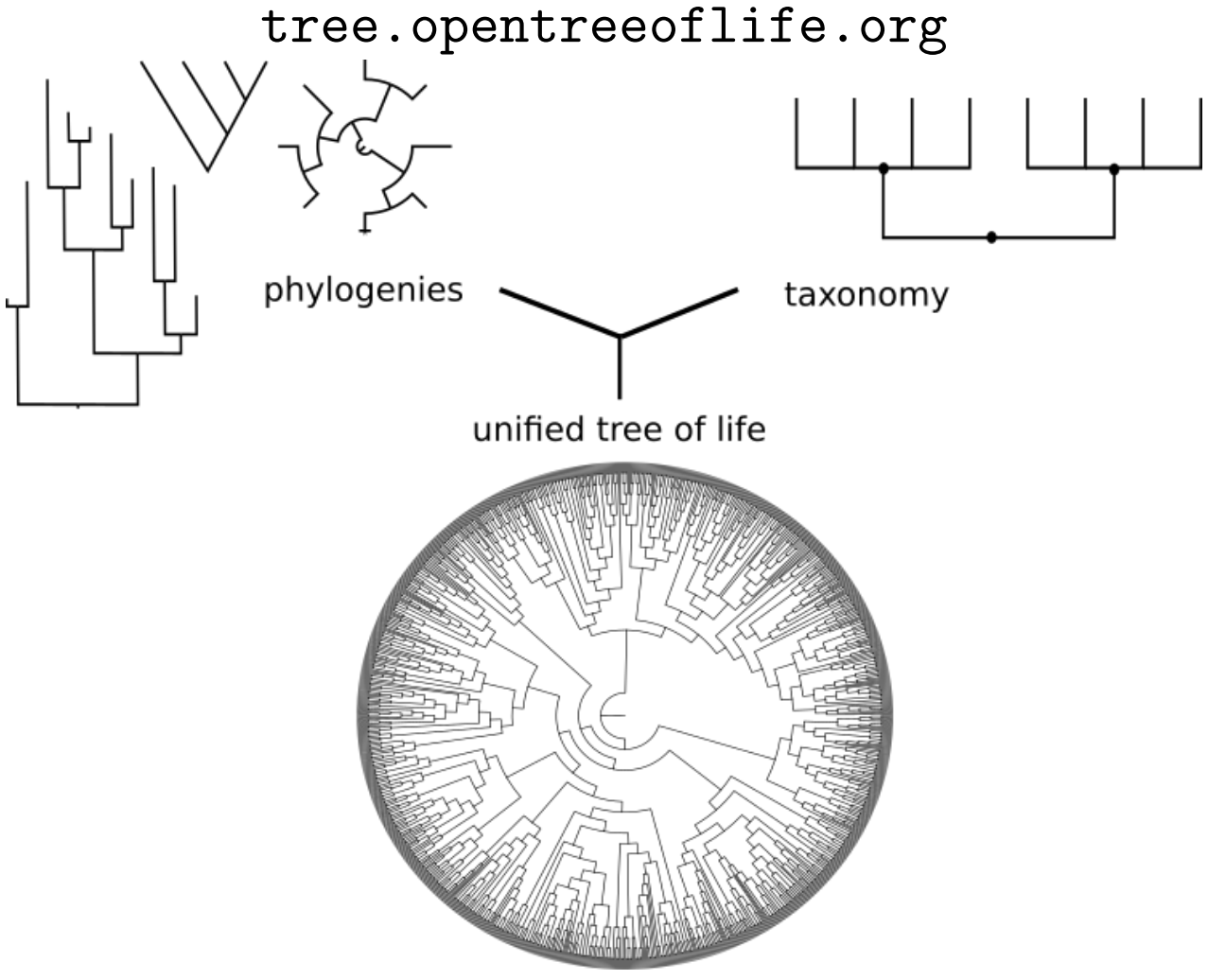
(Editorial, Science benefits from diversity. Nature. June 2018)

Are we achieving our aims of
open access to research and data,
and developing
a **diverse and inclusive** community
in the context of the Open Tree of Life Project?

Goal: Build a tree of all life.

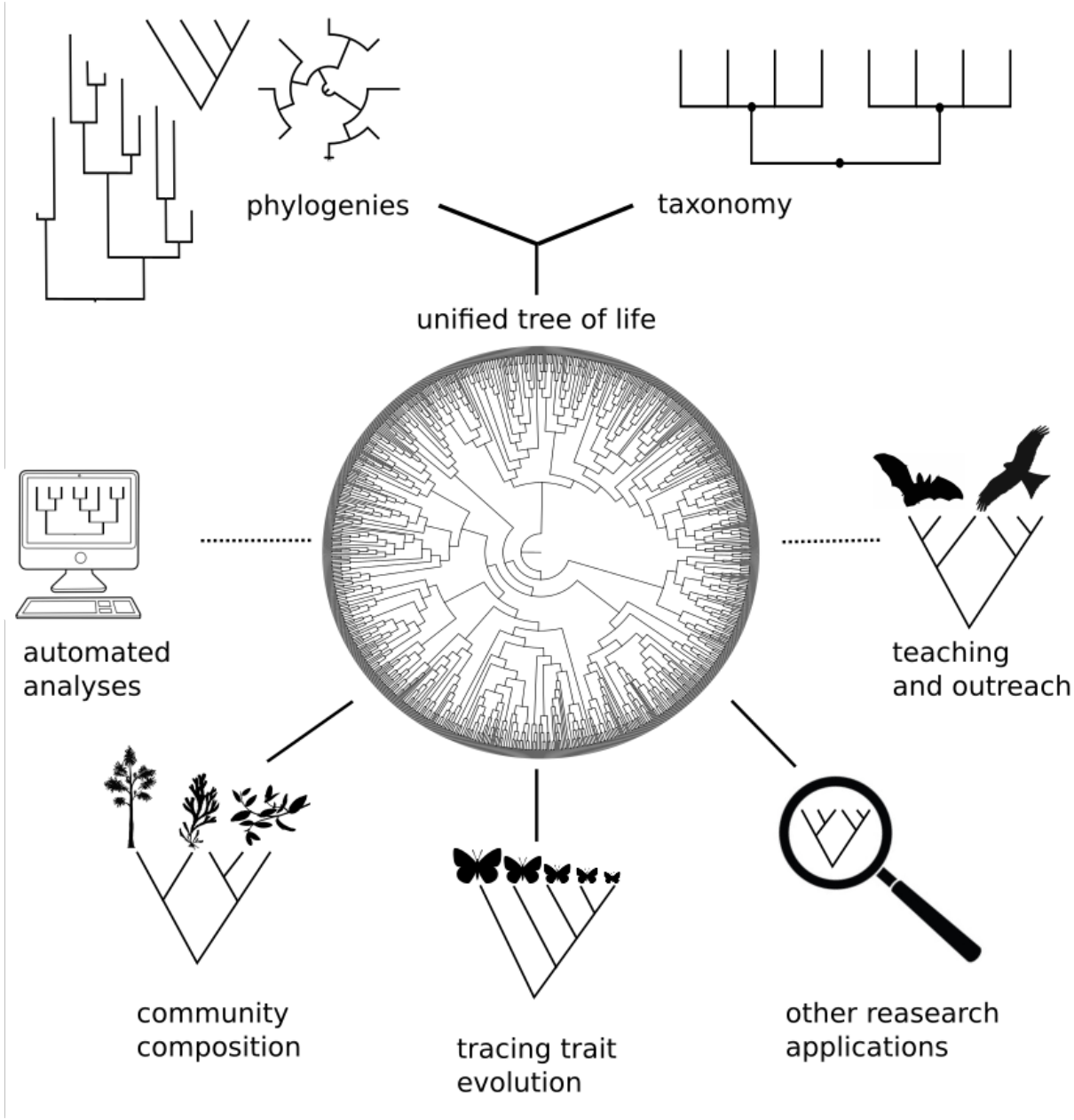
Goal: Build a tree of all life.

- Every named species
- Updated as new data becomes available
- Freely and easily accessible



(McTavish et al. Bioessays 2017)

tree.opentreeoflife.org

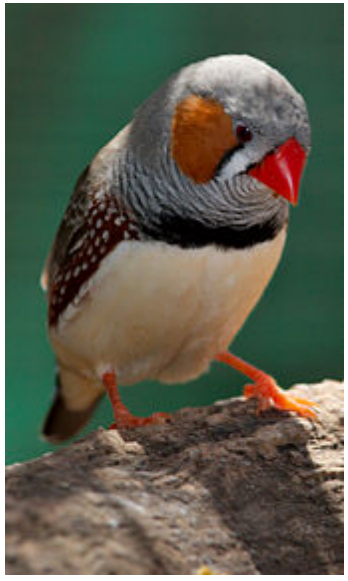


(McTavish et al. Bioessays 2017)

Which existing reference genome should I use to assemble the transcriptome of the Society finch?

Lonchura striata domestica

Zebra finch
(*Taeniopygia guttata*)



Ground finch
(*Geospiza fortis*)

Access 

through the `rotl` package from ROpenSci

(Michonneau et al. Methods in Ecology and Evolution 2016)

Access  through the `rotl` package from ROpenSci

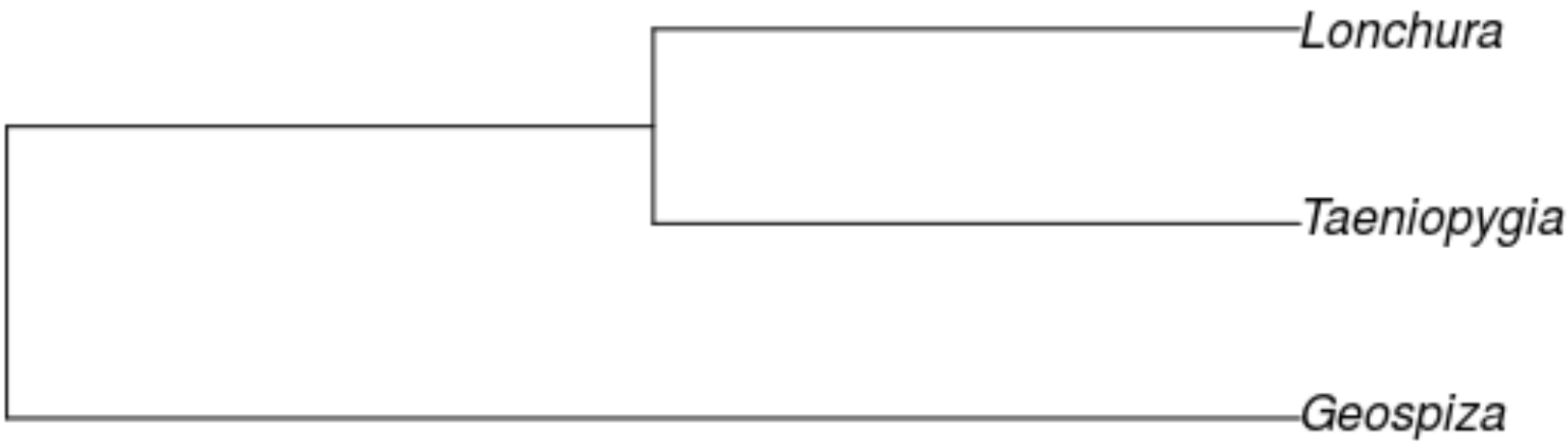
(Michonneau et al. Methods in Ecology and Evolution 2016)

```
install.packages("rotl")
library(rotl)
spp <- c("Geospiza", "Taeniopygia", "Lonchura")
taxa <- tnrs_match_names(spp, context="Animals")
tr <- tol_induced_subtree(ott_id(taxa), label="name")
plot(tr)
```

Access OPEN Tree of Life through the `rotl` package from ROpenSci

(Michonneau et al. Methods in Ecology and Evolution 2016)

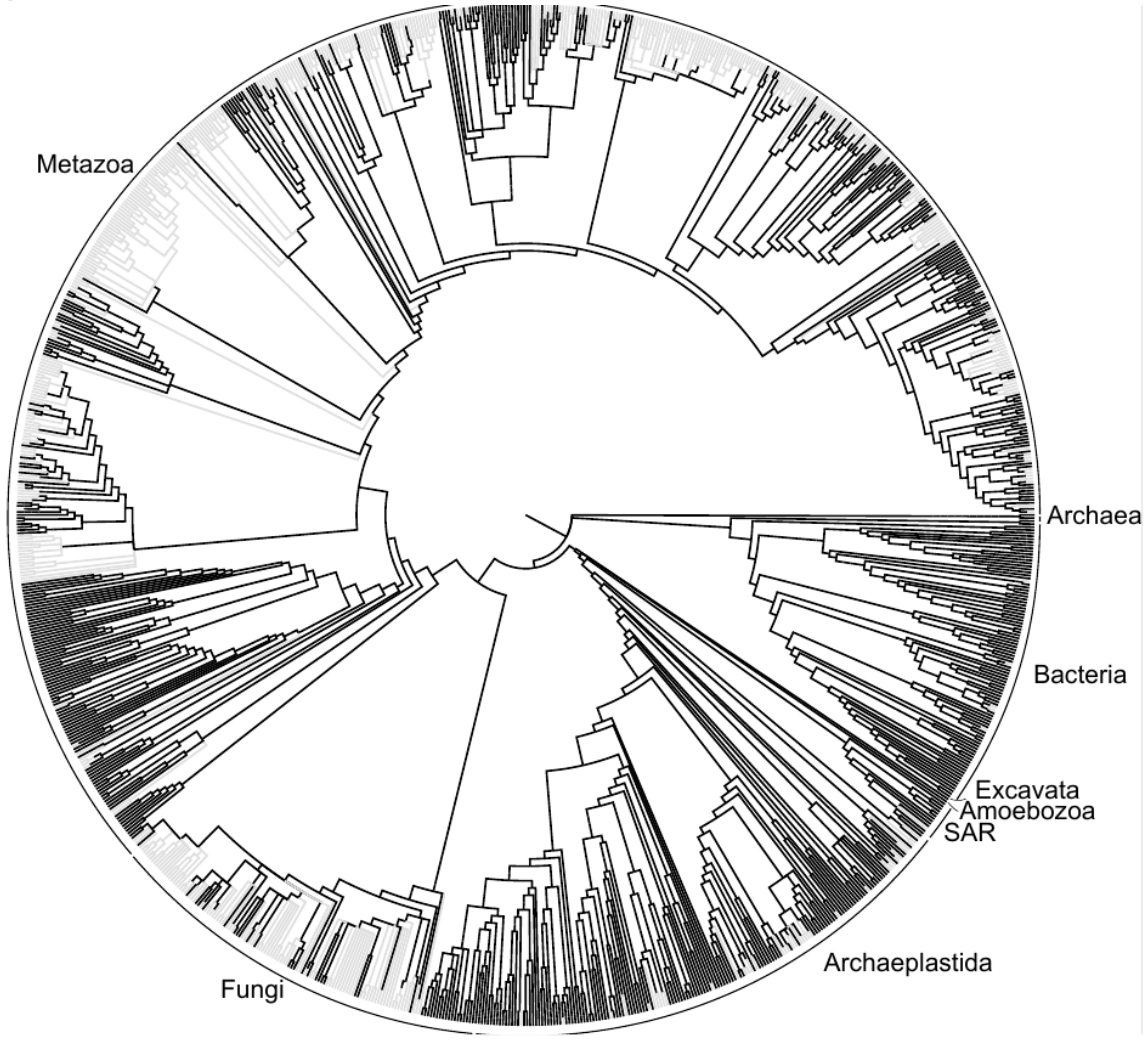
```
install.packages("rotl")
library(rotl)
spp <- c("Geospiza", "Taeniopygia", "Lonchura")
taxa <- tnrs_match_names(spp, context="Animals")
tr <- tol_induced_subtree(ott_id(taxa), label="name")
plot(tr)
```



Use the zebra finch as reference!

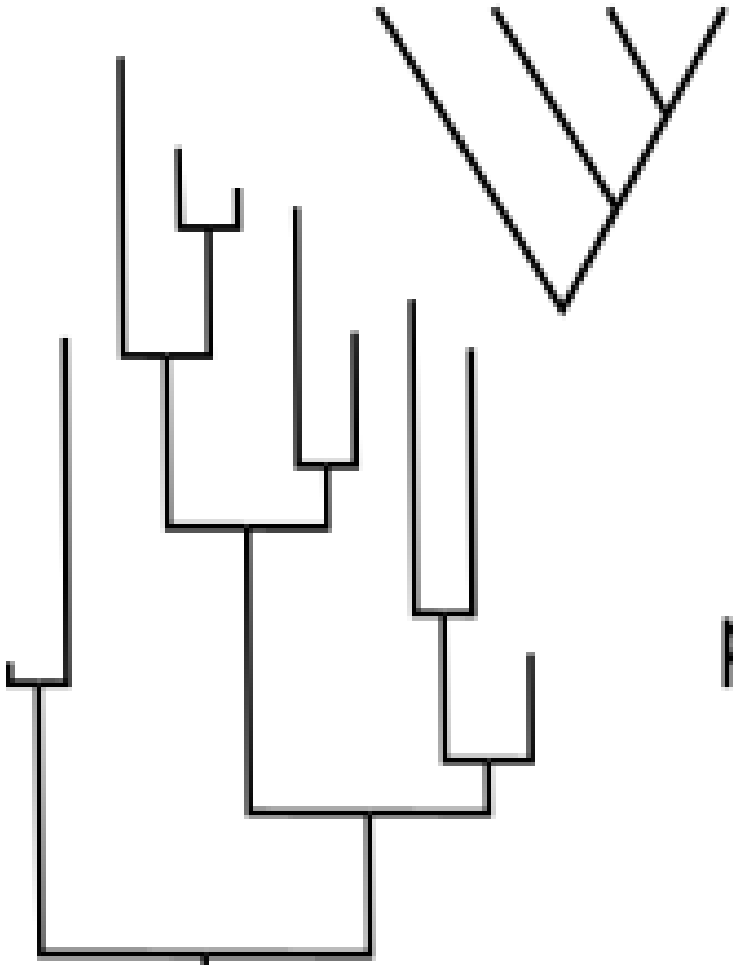
The synthetic tree (Hinchliff et al., PNAS 2015)

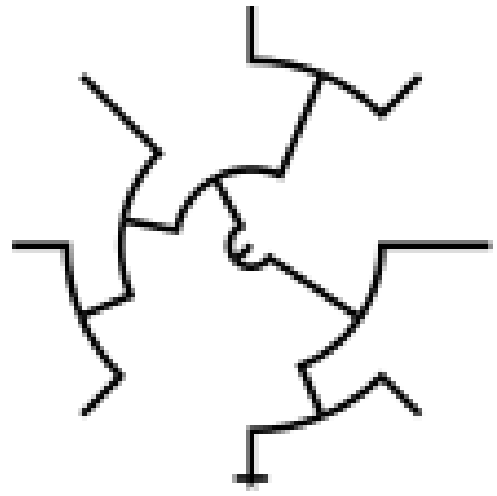
2.4 million species



Dark lineages have at least one representative in an input source tree

How will we fill in the gaps?





phylogenies

only 16% of phylogenies published 2000-2013 are digitally available (Drew et al. PLoS Biology 2013)

only 16% of phylogenies published 2000-2013 are digitally available (Drew et al. PLoS Biology 2013)

20% of phylogenies published 2013-2017 are digitally available

(McTavish et al. BioEssays 2018)

Synthesizing evolutionary knowledge across the tree of life requires community investment and involvement

- Getting digital access to phylogenetic estimates
- Data curation, potentially by multiple people
- Adding new studies as they are published

Our approach to cultivating community collaboration for the Open Tree of Life

- Recognizing and minimize barriers to sharing and accessing data
- Maximize benefits to curating data
- Targeted in person outreach

Data curation

- Trees can be uploaded from any source, does not have to be own data.
- Easy to use browser based interface
- Track curation attribution by name or pseudonym
- Data store is hosted publicly on GitHub

`github.com/OpenTreeOfLife/phylesystem-1`

McTavish et al. Bioinformatics 2015

Community Curation

235 individual curators of uploaded studies

Community Curation

235 individual curators of uploaded studies

Rapid curation progress at taxon focused in-person working groups, in collaboration with FuturePhy

Community Curation

235 individual curators of uploaded studies

Rapid curation progress at taxon focused in-person working groups, in collaboration with FuturePhy

Based on display name: 39 women, 61 men, 135 unknown

Community Curation

235 individual curators of uploaded studies

Rapid curation progress at taxon focused in-person working groups, in collaboration with FuturePhy

Based on display name: 39 women, 61 men, 135 unknown

Some success at approaching gender parity, but racial and ethnic diversity of participants is low.

Community Curation

235 individual curators of uploaded studies

Rapid curation progress at taxon focused in-person working groups, in collaboration with FuturePhy

Based on display name: 39 women, 61 men, 135 unknown

Some success at approaching gender parity, but racial and ethnic diversity of participants is low.

Open Tree Developers

12 main developers - 2 women, 10 men

Next steps: Broaden community participation

Better incentives for data uploading and curation

- Automated updating
- Branch lengths / Node ages
- Private data

Hands on workshops targeted to groups under-represented in systematics

Holder, McTavish  ABI 2018-2022

Cranston

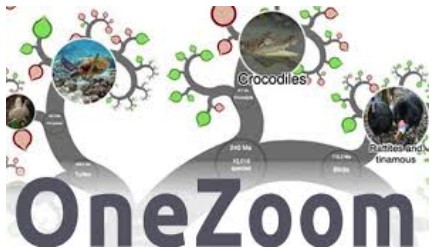
Open Tree resources are available via a range of implementations

- Browser interface, `tree.opentreeoflife.org`
- Open Tree of Life API
- R Open Tree of Life (rotl) 

Open Tree resources are available via a range of implementations

- Browser interface, `tree.opentreeoflife.org`
- Open Tree of Life API
- R Open Tree of Life (rotl) ROpenSci

Open Tree provides the tree backend for:

- DATElife
- 

- PHYLOtastic

Provides access at any technical skill level

Conclusions

- Phylogenetic information should be accessible and reusable
- Open Science is not automatically diverse or inclusive, and can be the opposite
- A variety of tools and approaches provides wide access to Open Tree resources
- Expansion of the contributor and developer community is necessary

Contribute your knowledge!

`tree.opentreeoflife.org/curator`

Thank You

OpenTree of Life project
NSF ABI 1759846
Mark Holder
Karen Cranston

NSF AVATOL 1208809
AVATOL PI'S: Burleigh, Crandall, Gude, Hibbett, Katz, Ree, Smith, Soltis, Williams

Lab group:
Martha Kandziora
Lesly Lopez Fang
Jasper Toscani-Field

Stephanie Meirmans
Maurine Neiman

August 19-22, 2018
Montpellier, France