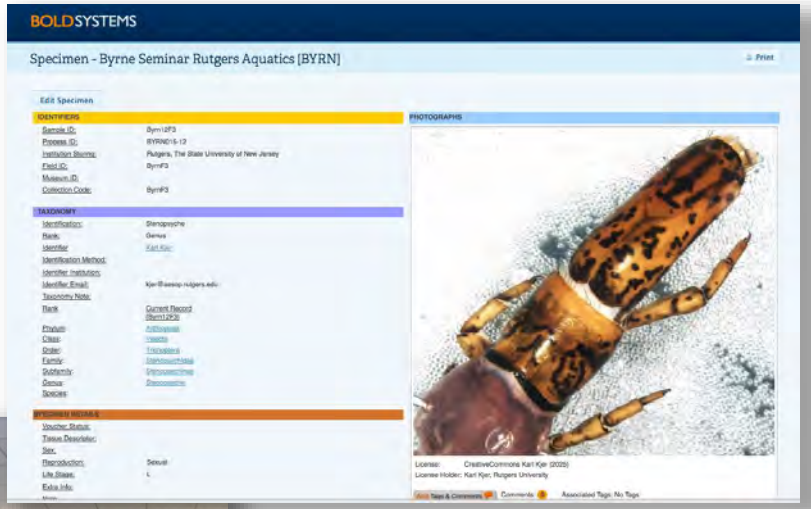
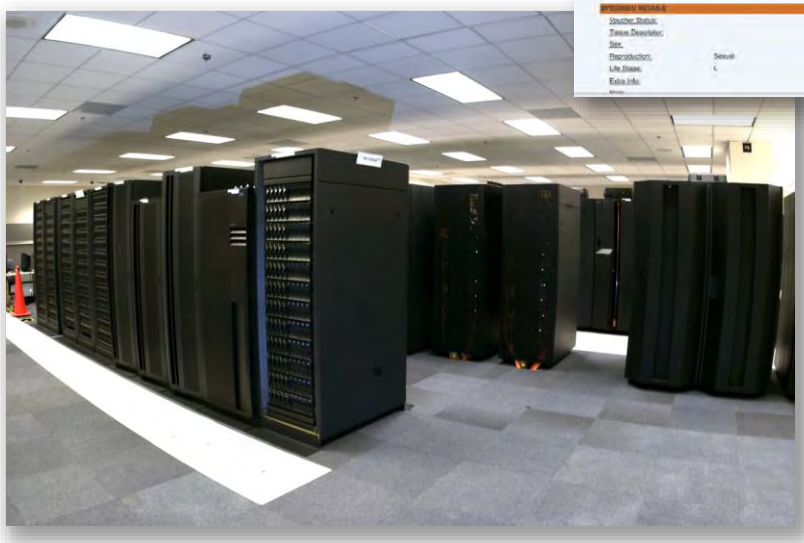
Integrating Trichoptera barcodes and big data: today's systematics

6th International Barcode of Life Conference
August 18-21, 2015





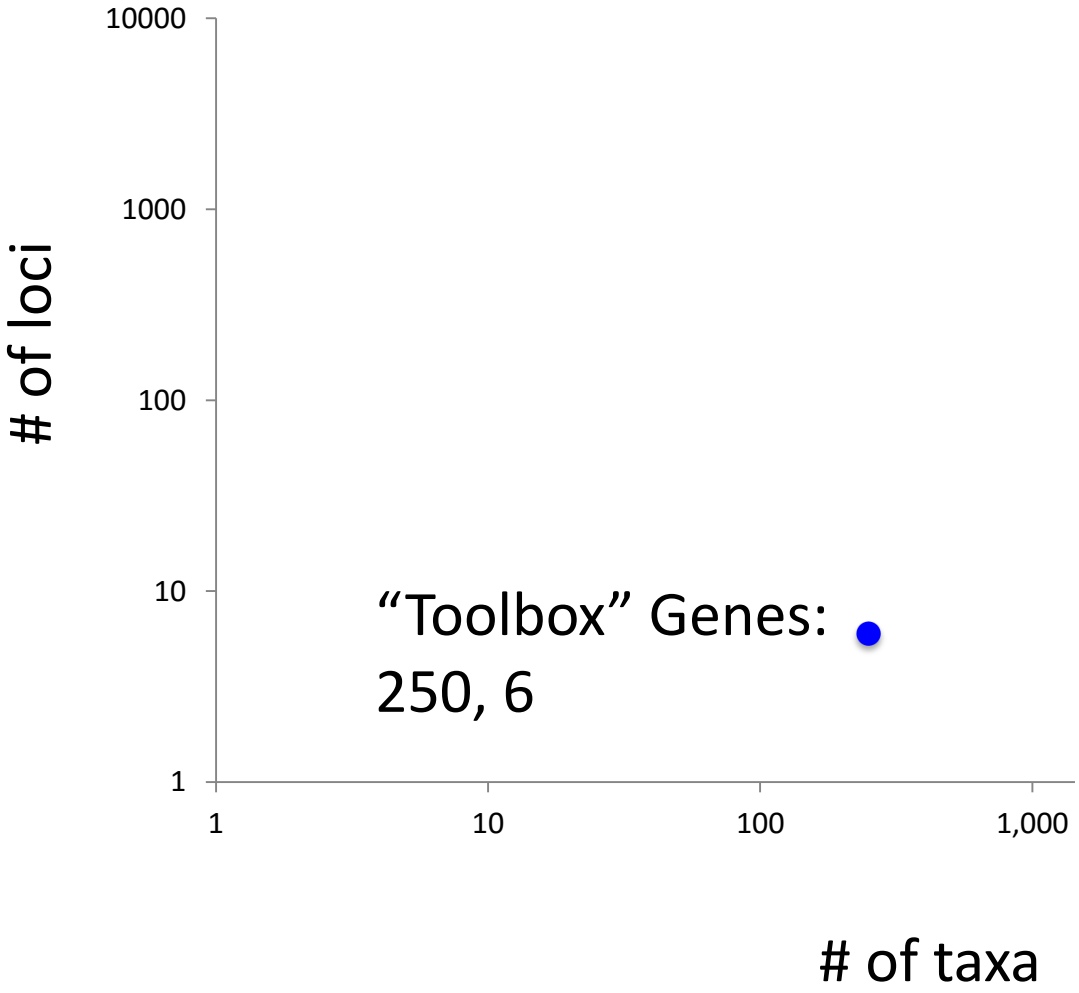
Karl Kjer, UC Davis,
Dept. of Entomology and Nematology

1KITE
1000 INSECT TRANSCRIPTOME EVOLUTION

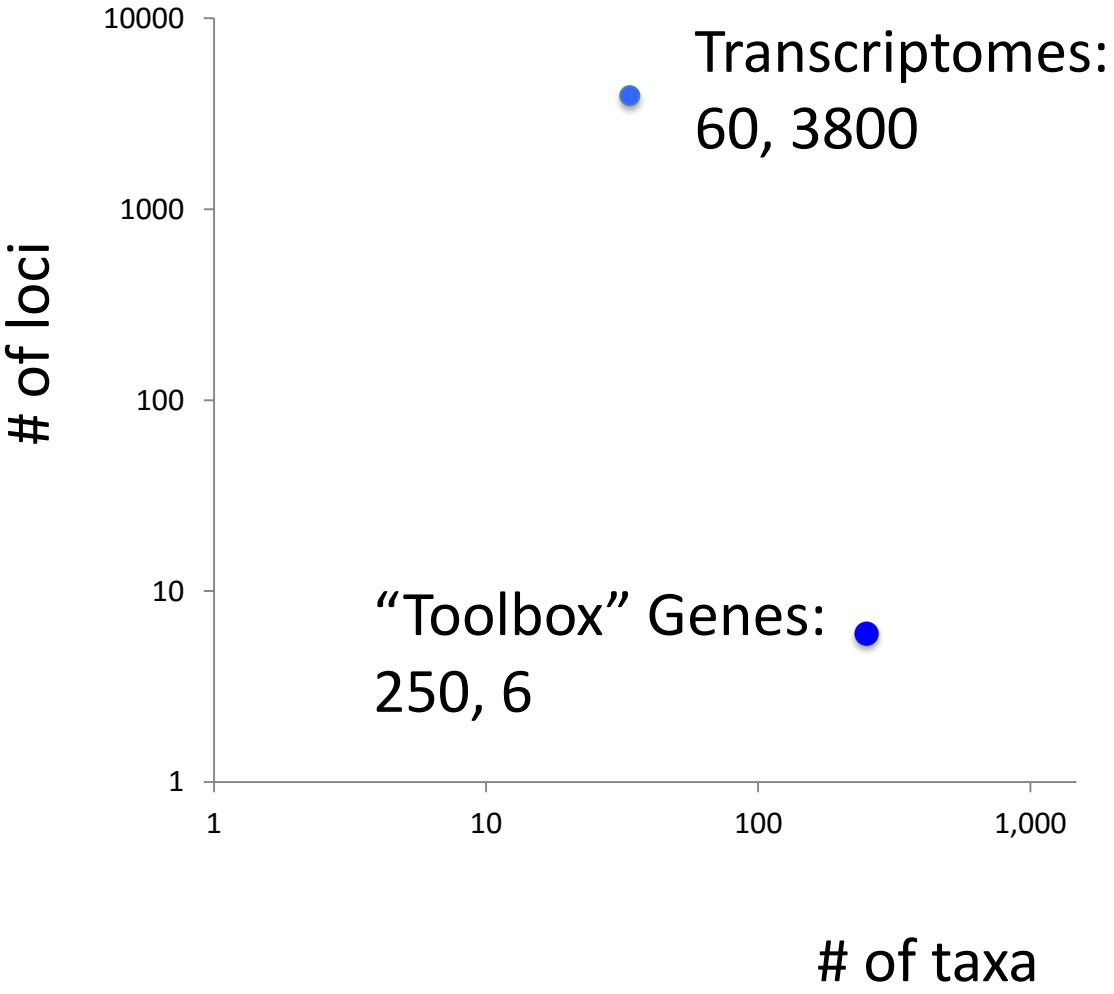
1. Barcodes
2. Transcriptomes
3. Targeted Enrichment

Theme: How do we stitch these very different data into a coherent picture of phylogeny?

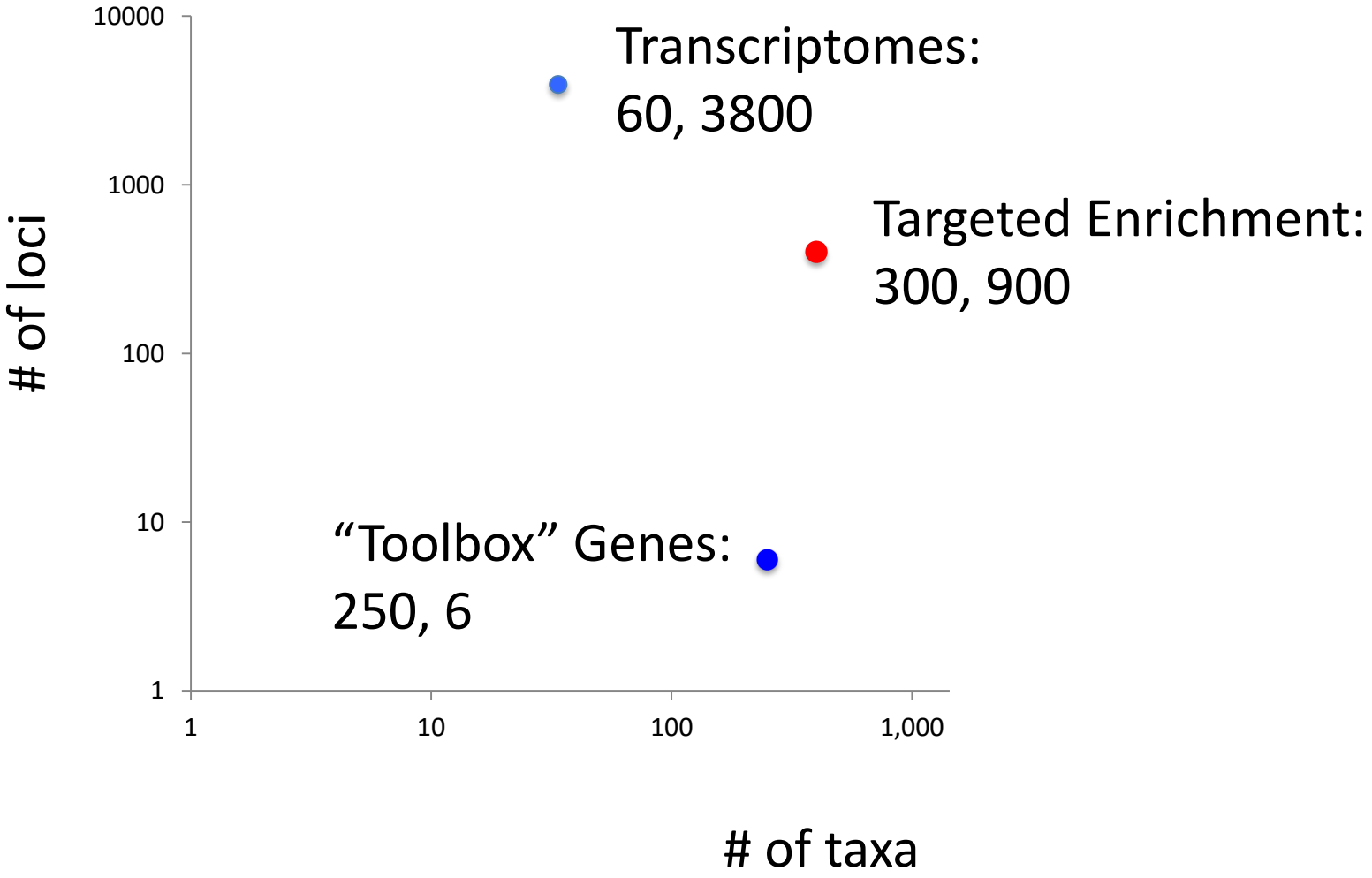
The Trichoptera Data



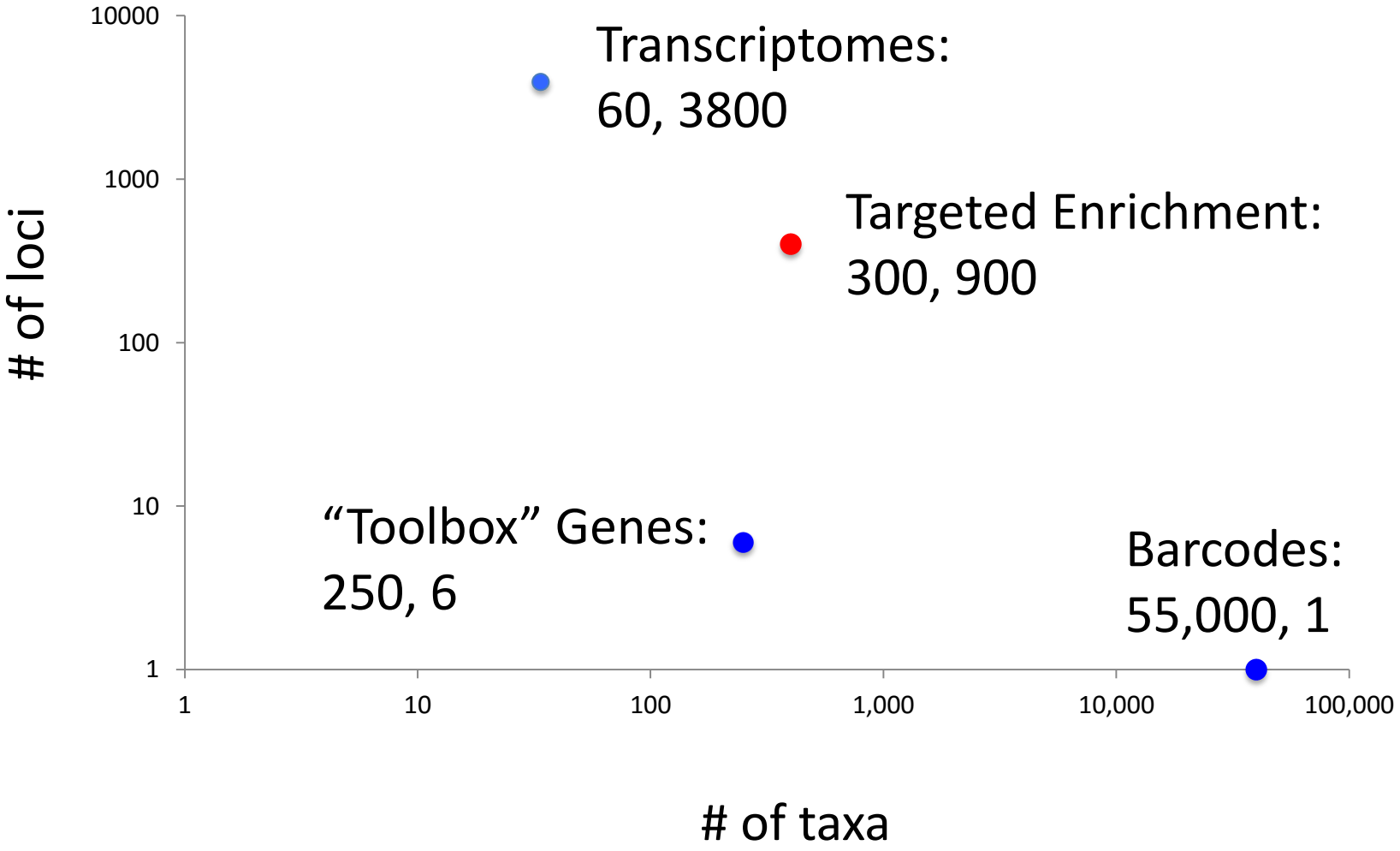
The Trichoptera Data



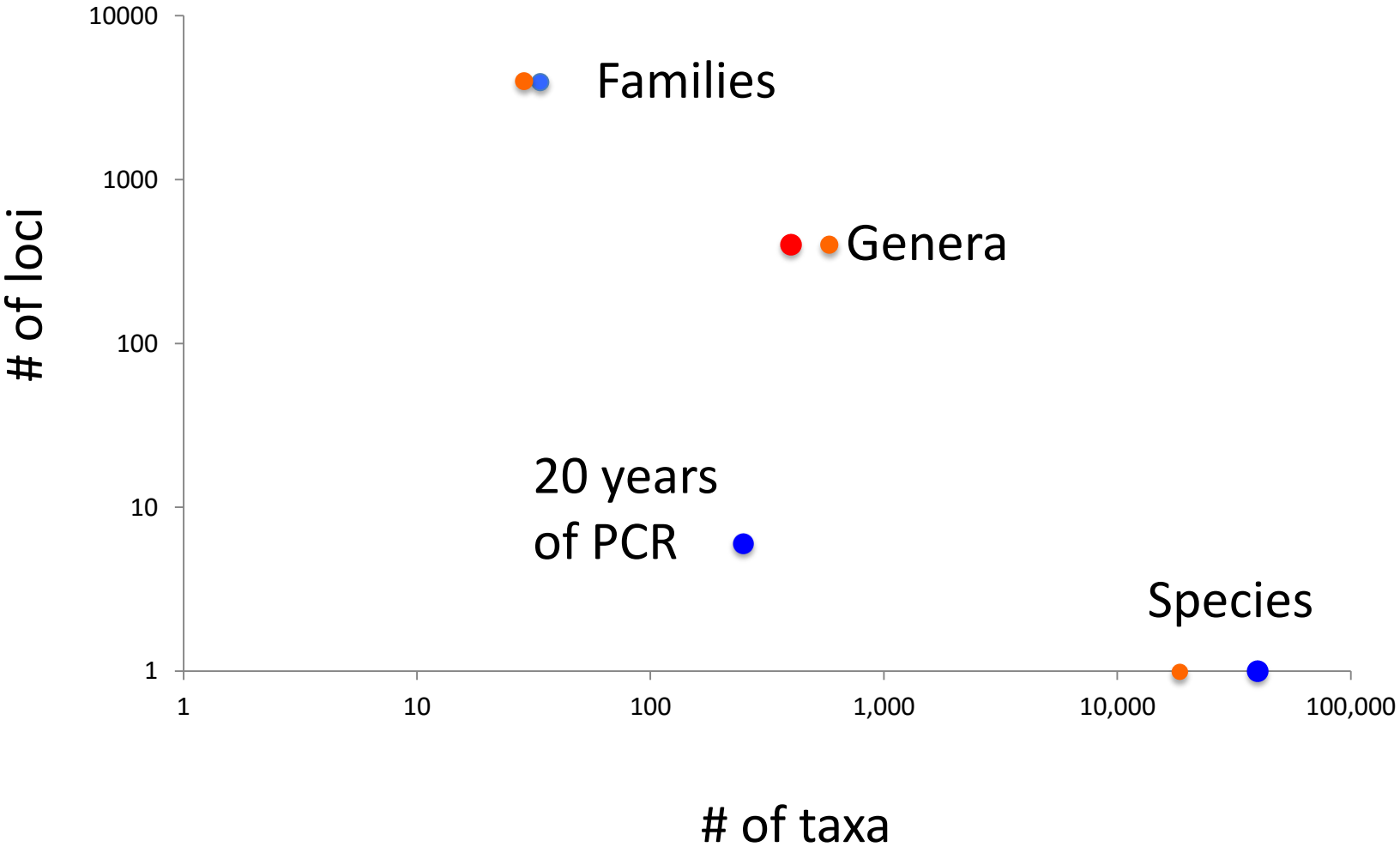
The Trichoptera Data



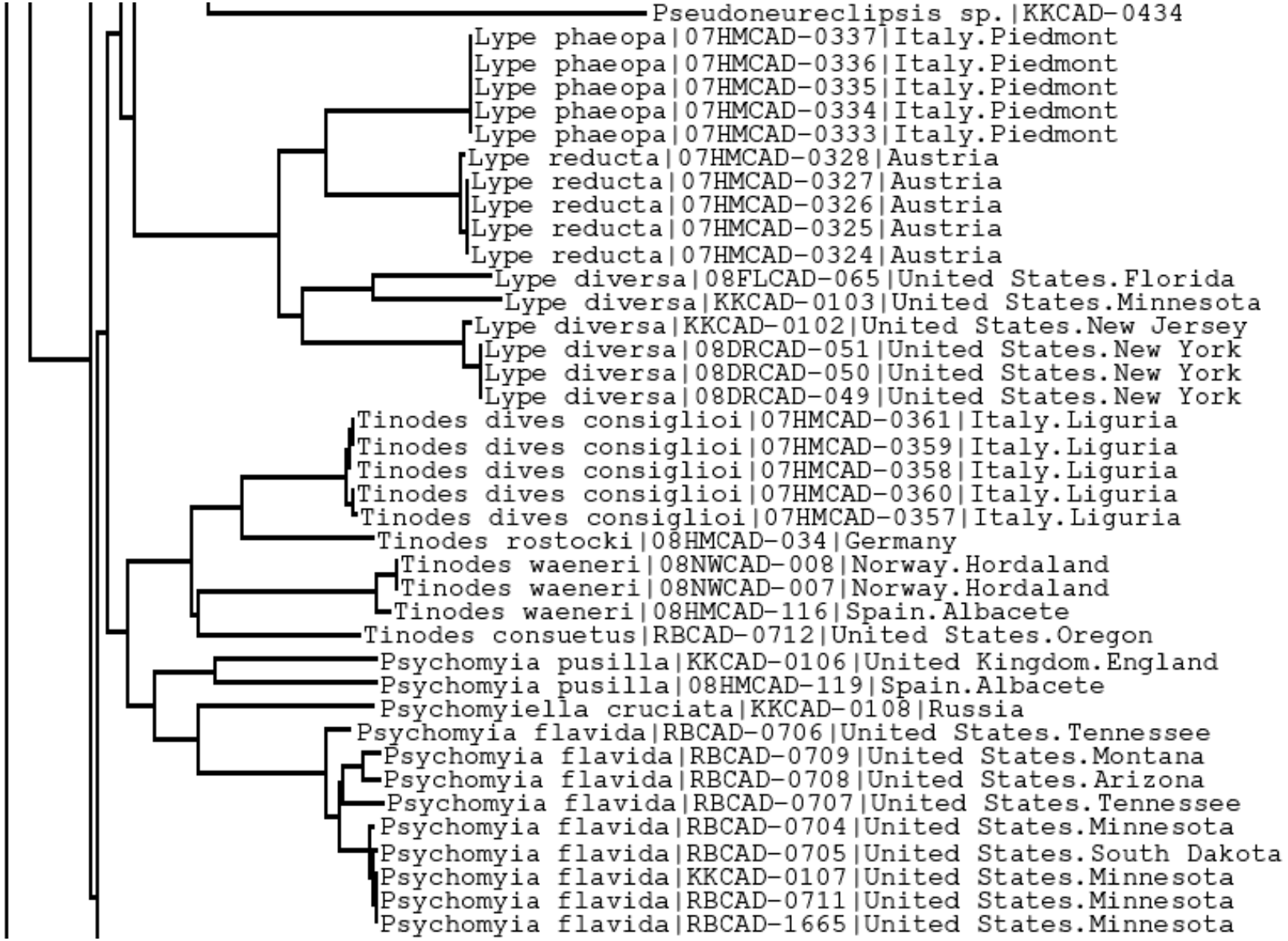
The Trichoptera Data



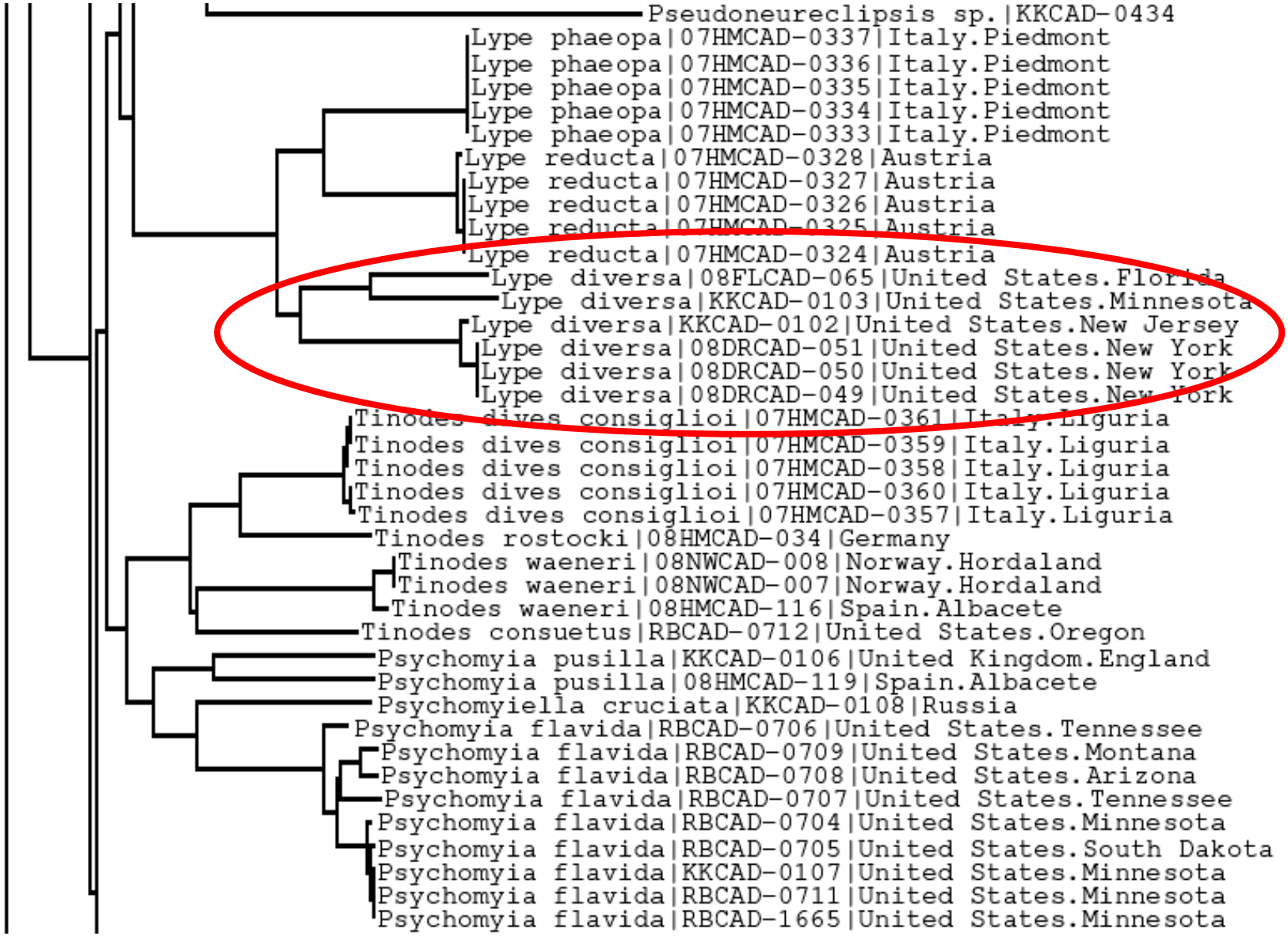
The Trichoptera Data

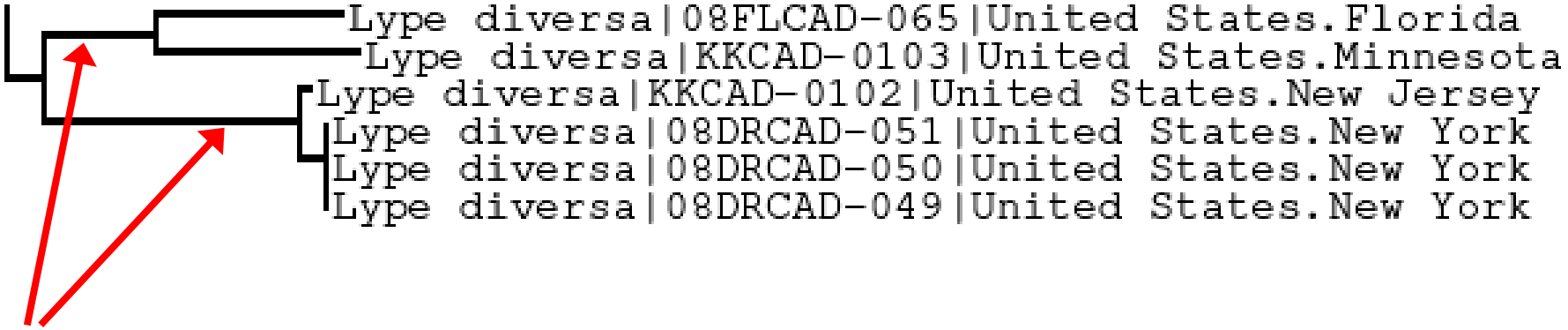


We have 2000 times this, from over 4,500 Trichoptera species.
And the interesting one is…



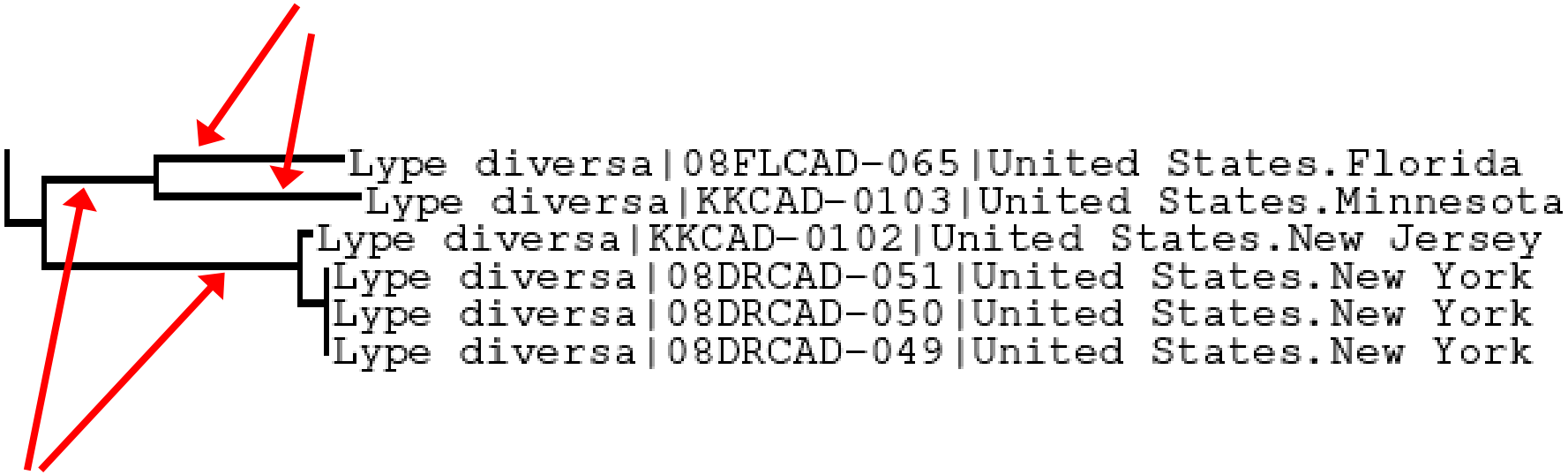
Here. Lype diversa





Are these 2 different species?

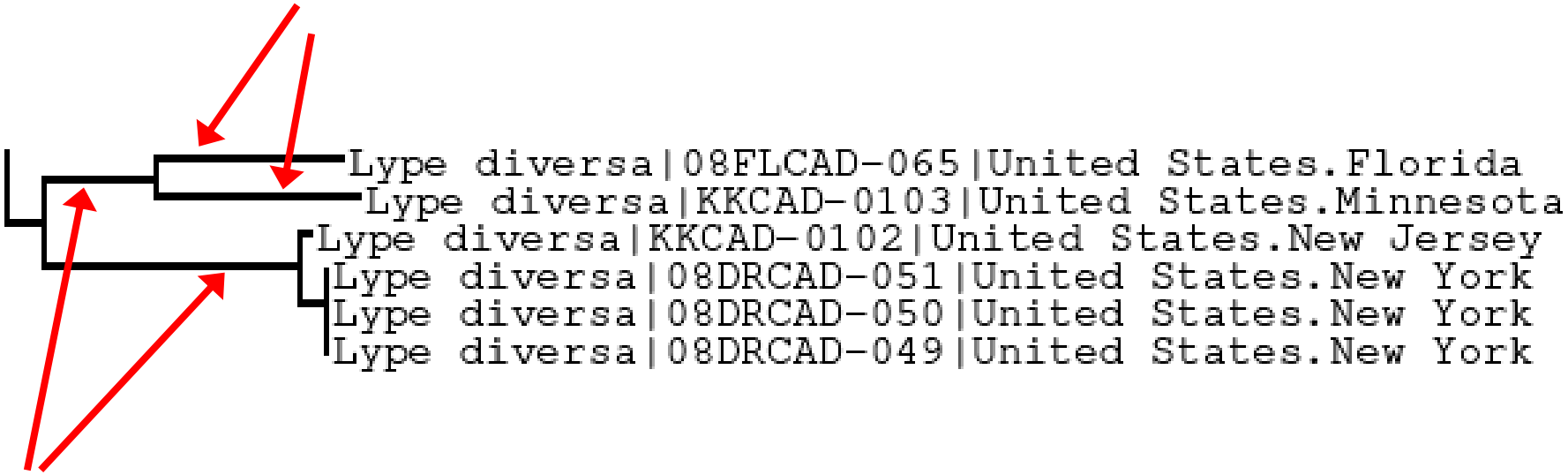
Maybe…what about these?



Are these 2 different species?

We have BINs and MOTUs (which are great, and useful). Are they species?

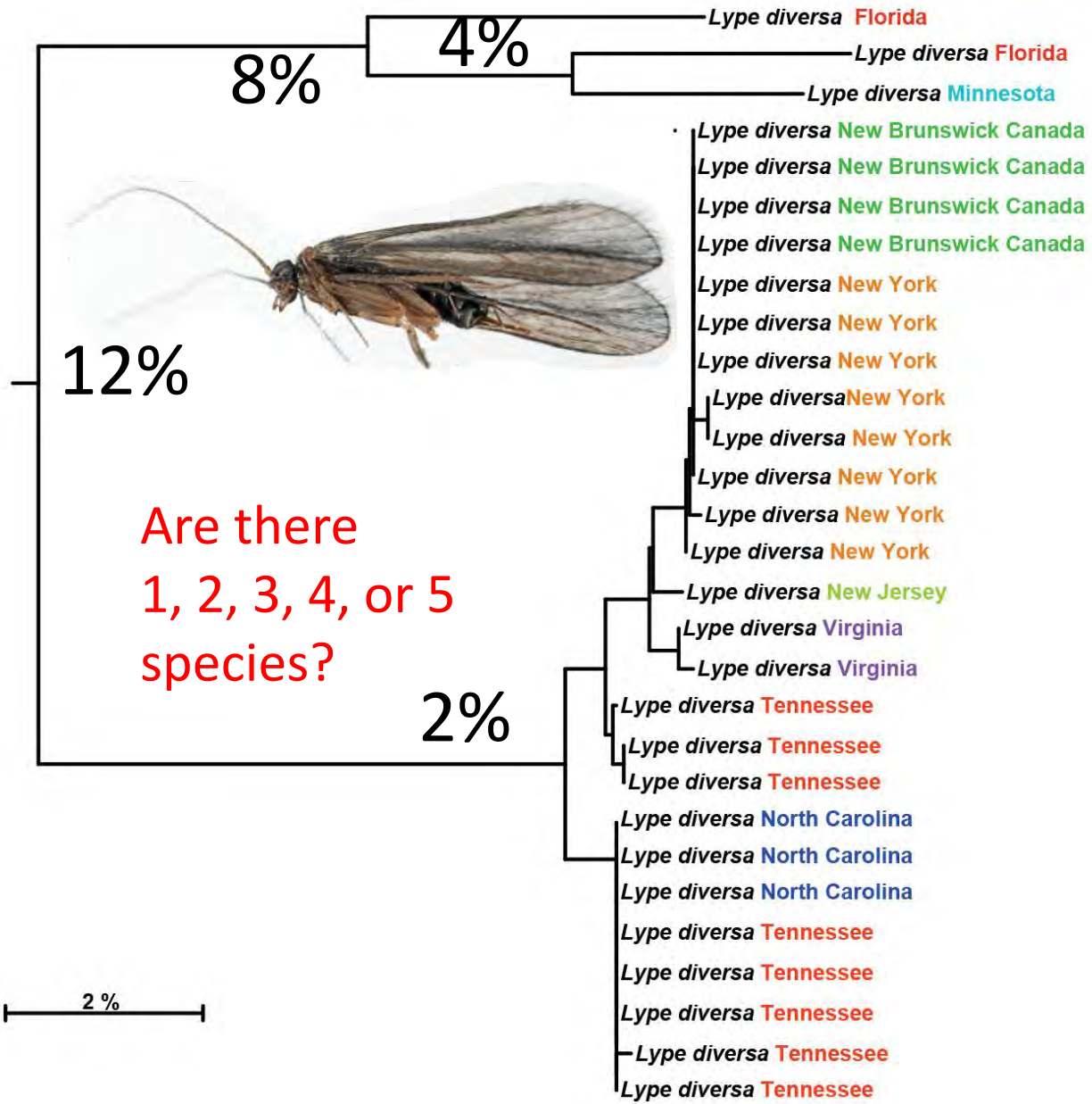
Maybe…what about these?



Are these 2 different species?

We need more information than barcodes can provide

Lype diversa



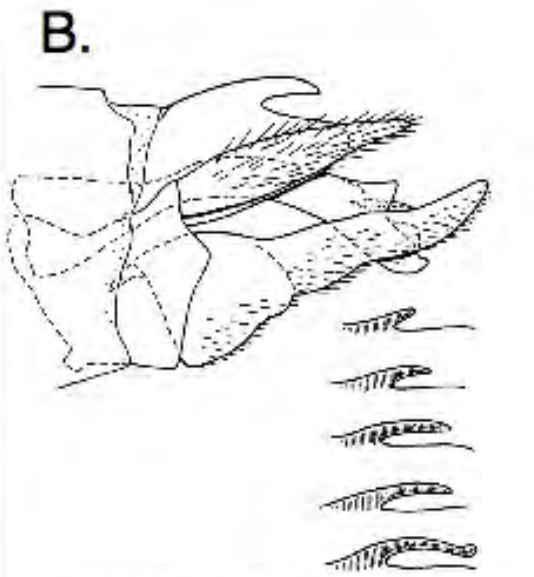
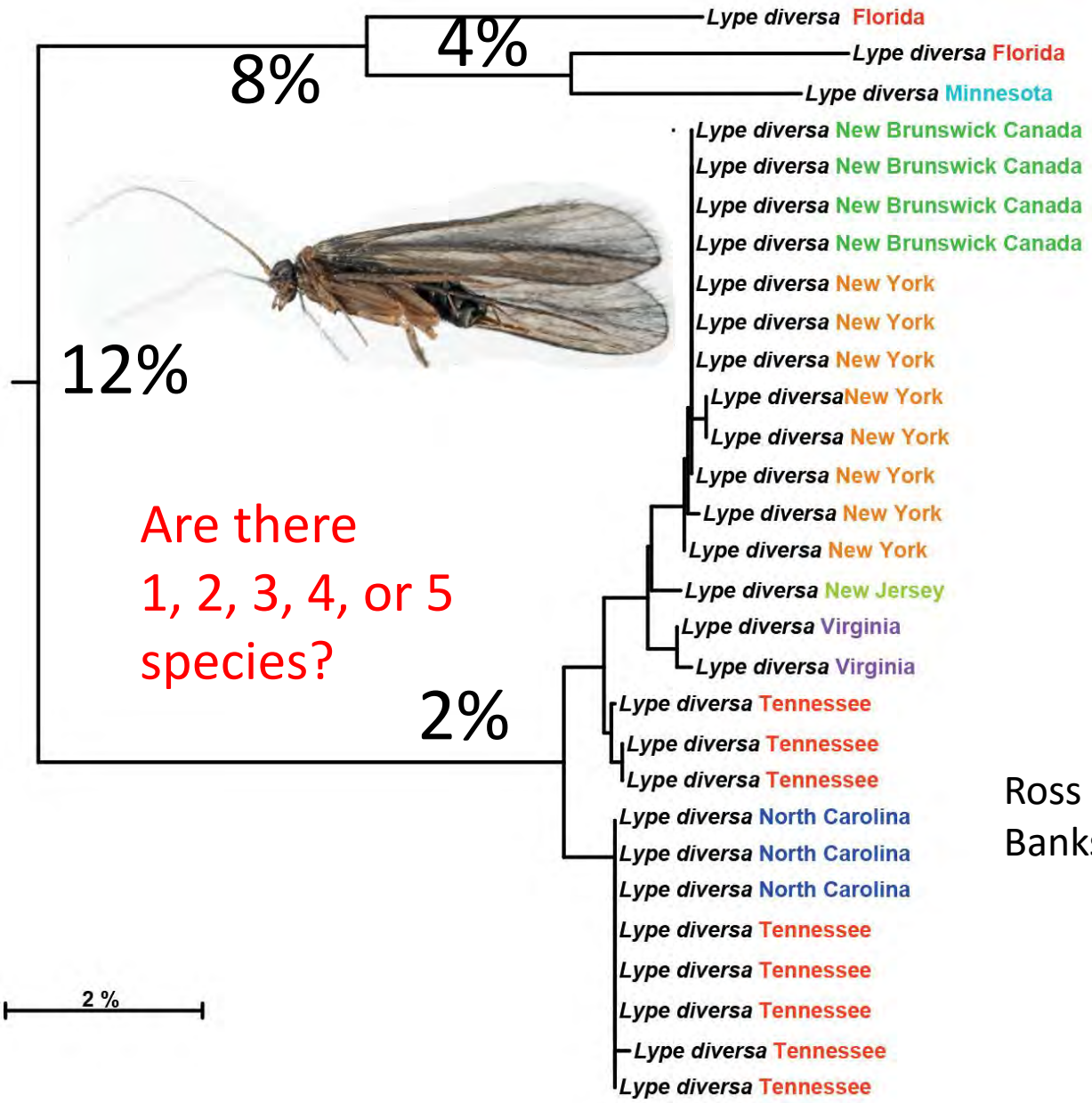


Fig. 275. - *Lype diversa*, male genitalia, showing variations of dorsal horn or the tenth tergite. (Ross 1944)



Are there
1, 2, 3, 4, or 5
species?

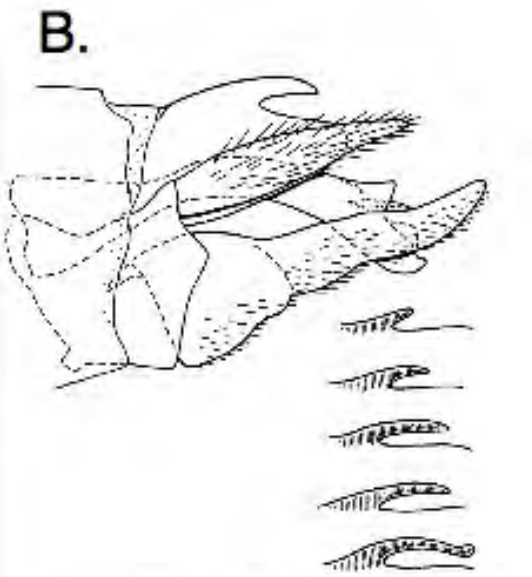
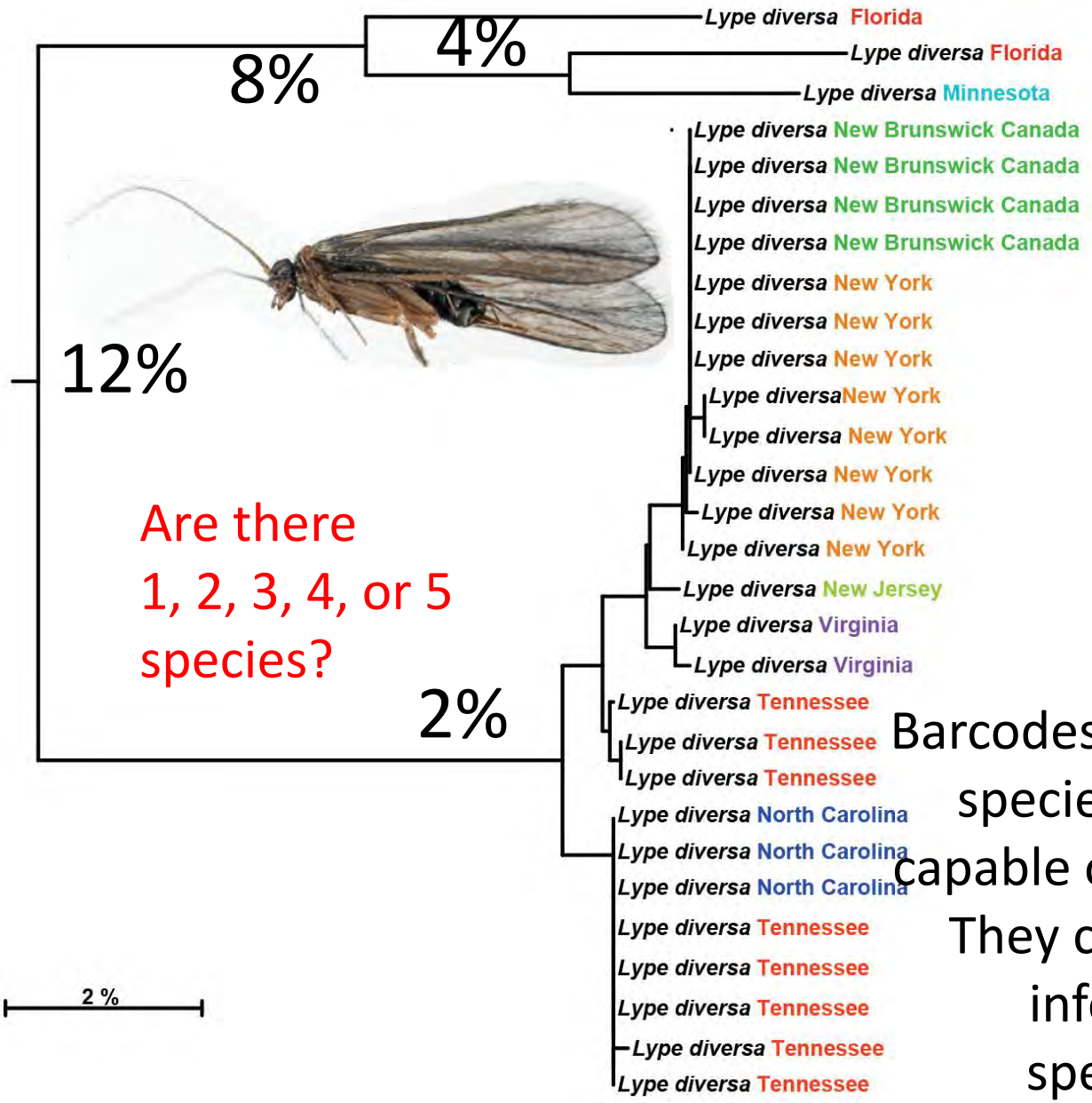


Fig. 275. - *Lype diversa*, male genitalia, showing variations of dorsal horn or the tenth tergite. (Ross 1944)

Ross noticed this back in ’44
Banks must have noticed it in ‘14



Are there 1, 2, 3, 4, or 5 species?

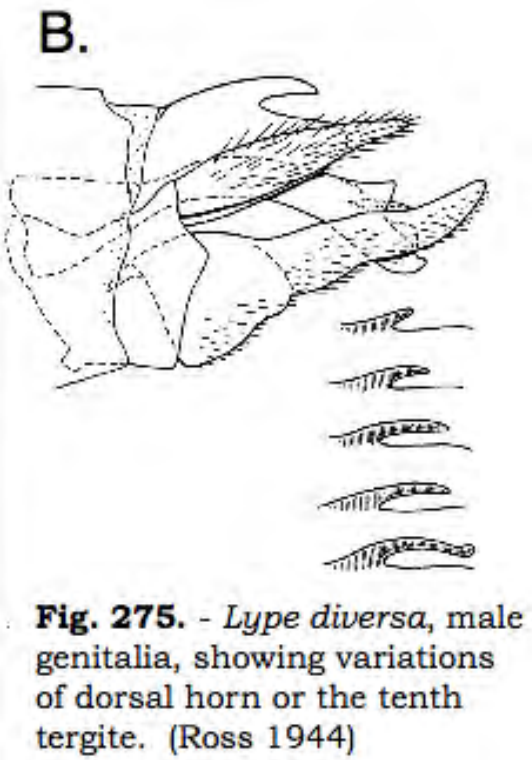
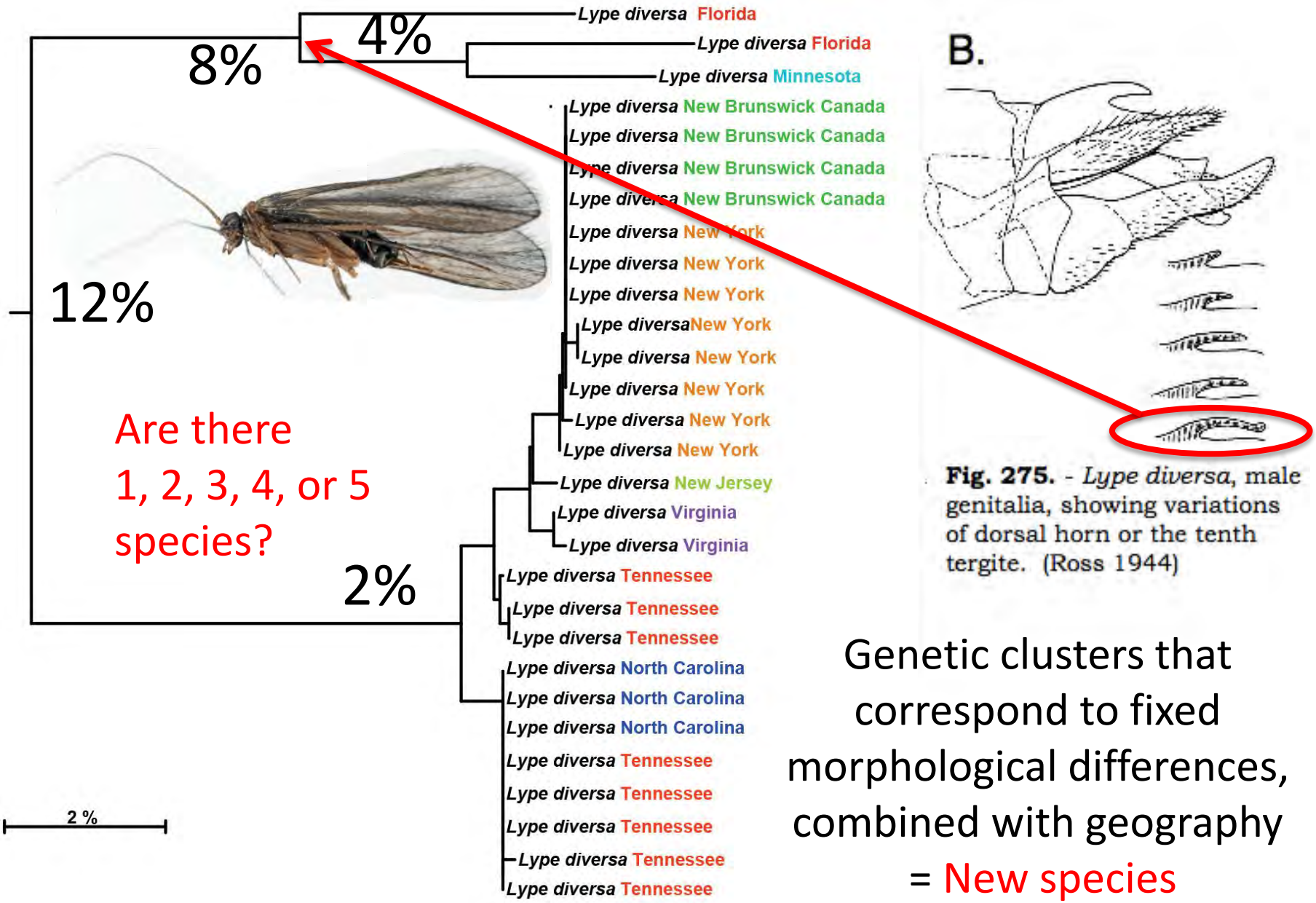


Fig. 275. - *Lype diversa*, male genitalia, showing variations of dorsal horn or the tenth tergite. (Ross 1944)

Barcodes alone may not define species, but are perfectly capable of describing diversity. They can be a reciprocally informative part of species description

This illustrates both the power, and the challenge of barcodes.

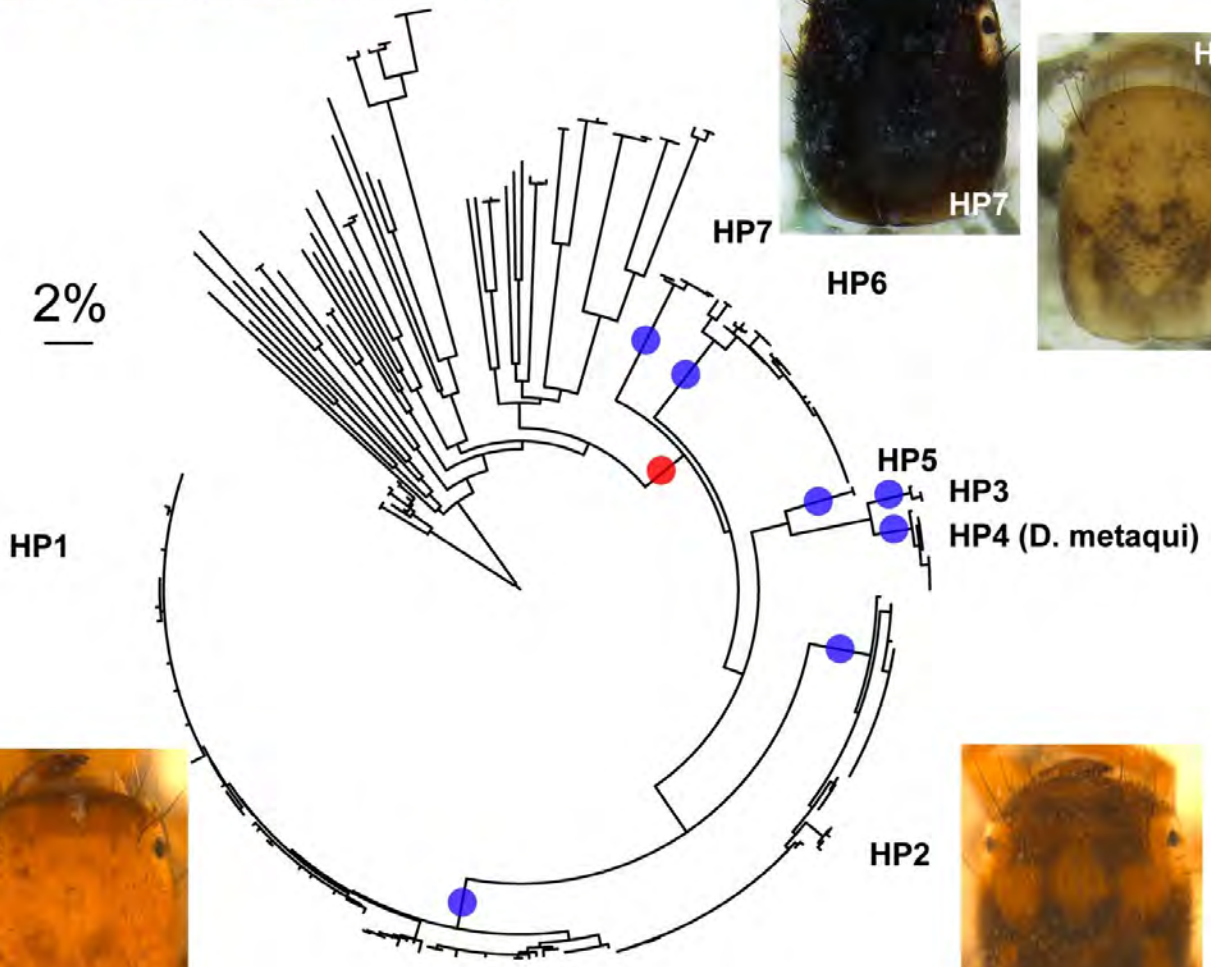


Genetic clusters that correspond to fixed morphological differences, combined with geography = New species

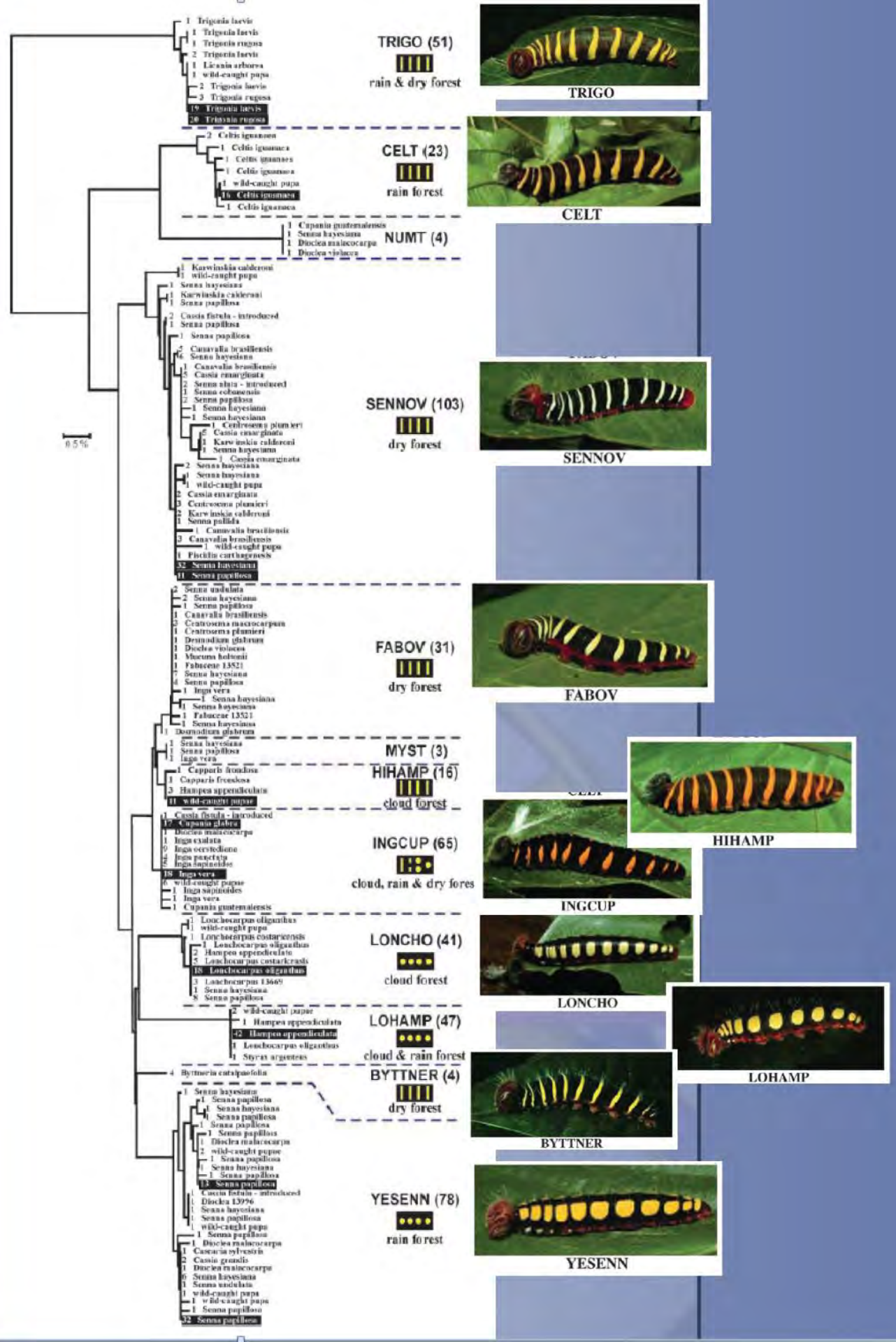
We see this all the time

World Diplectrona COI NJ tree

- Diplectrona modesta complex
- COI haplotype clusters



Harvey, Geraci, Robinson, Morse, Kjer, Zhou, 2012 Terr. Arth. Rev.

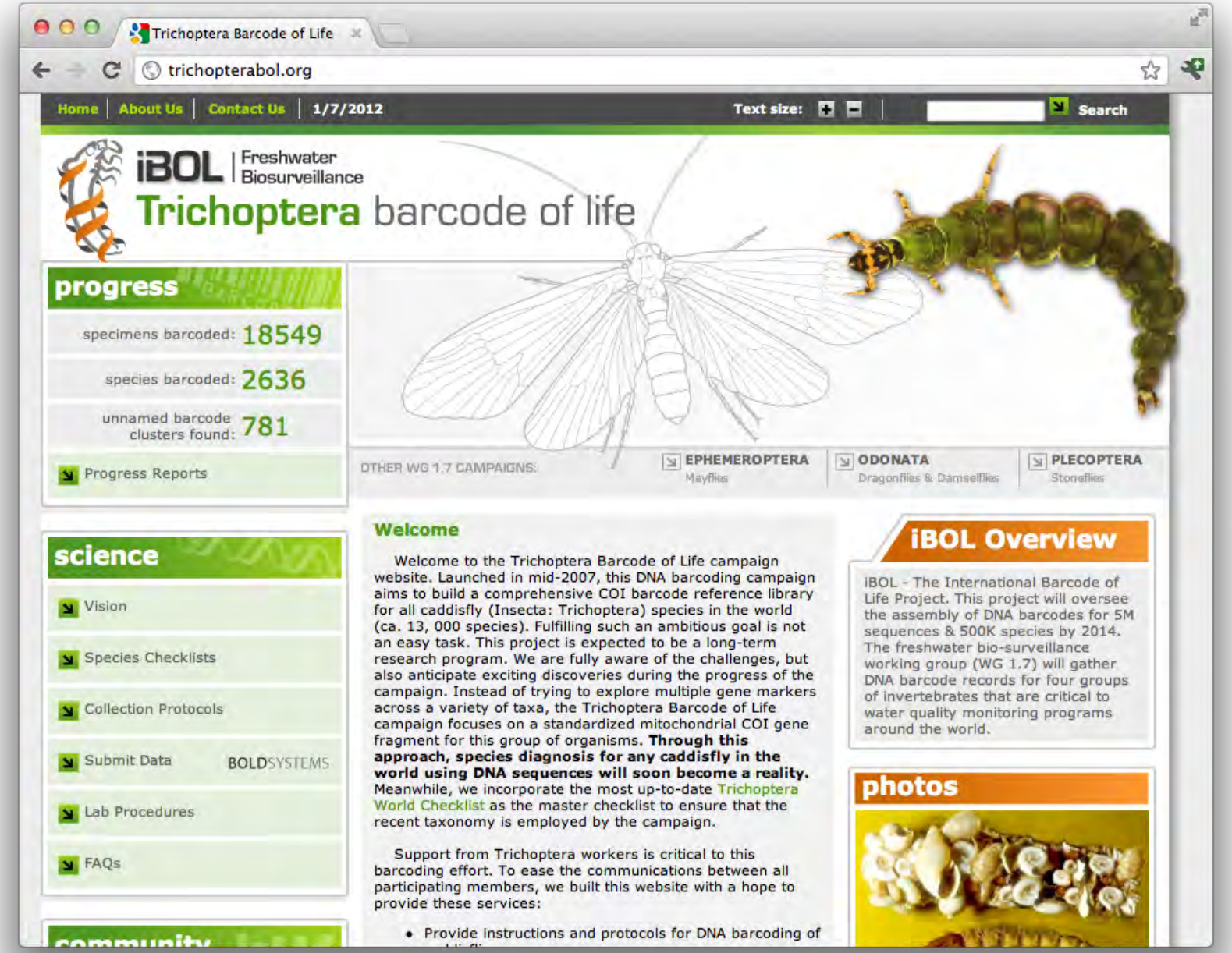

TRIGO (51)
rain & dry forest
TRIGO
CELT (23)
rain forest
CELT
NUMT (4)
SENNOV (103)
dry forest
SENNOV
0.5%
FABOV (31)
dry forest
FABOV
MYST (3)
HIHAMP (16)
cloud forest
HIHAMP
INGCUP (65)
cloud, rain & dry fores
INGCUP
LONCHO (41)
cloud forest
LONCHO
LOHAMP (47)
cloud & rain forest
LOHAMP
BYTTNER (4)
dry forest
BYTTNER
YESENN (78)
rain forest
YESENN

Challenge:

It is hard to distinguish between
inter-species and intra-species variation

Opportunity:
If we can associate distinct BINs with fixed morphological characters, we can increase confidence in species delineation and accelerate formal species description.

The database



Here is how it all started

Xin and I talk about China in 2002

We conclude that China is stepping up on the world stage, and will have resources
What do they want? Big science.
What impresses them? Big names, *Science* papers

Xin and I talk about China

We saw that the barcode initiative was big science, and Paul Hebert is a name China would know: Postdoc there.

Xin and I talk about China

Big science.
Famous guy.
Xin postdoc

iBOL | Freshwater Biosurveillance
Trichoptera barcode of life

progress
specimens barcoded: 18549
species barcoded: 2636
unnamed barcode clusters found: 781
Progress Reports

science
Vision
Species Checklists
Collection Protocols
Submit Data BOLDSYSTEMS

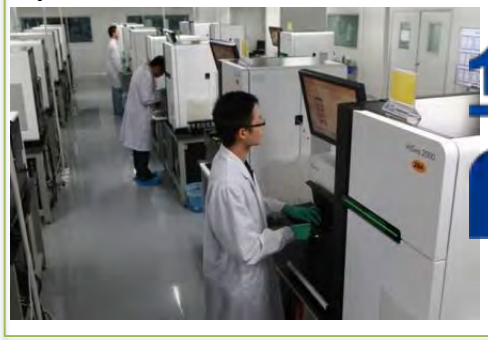
Leadership Team

Campaign Coordinator: Xin Zhou

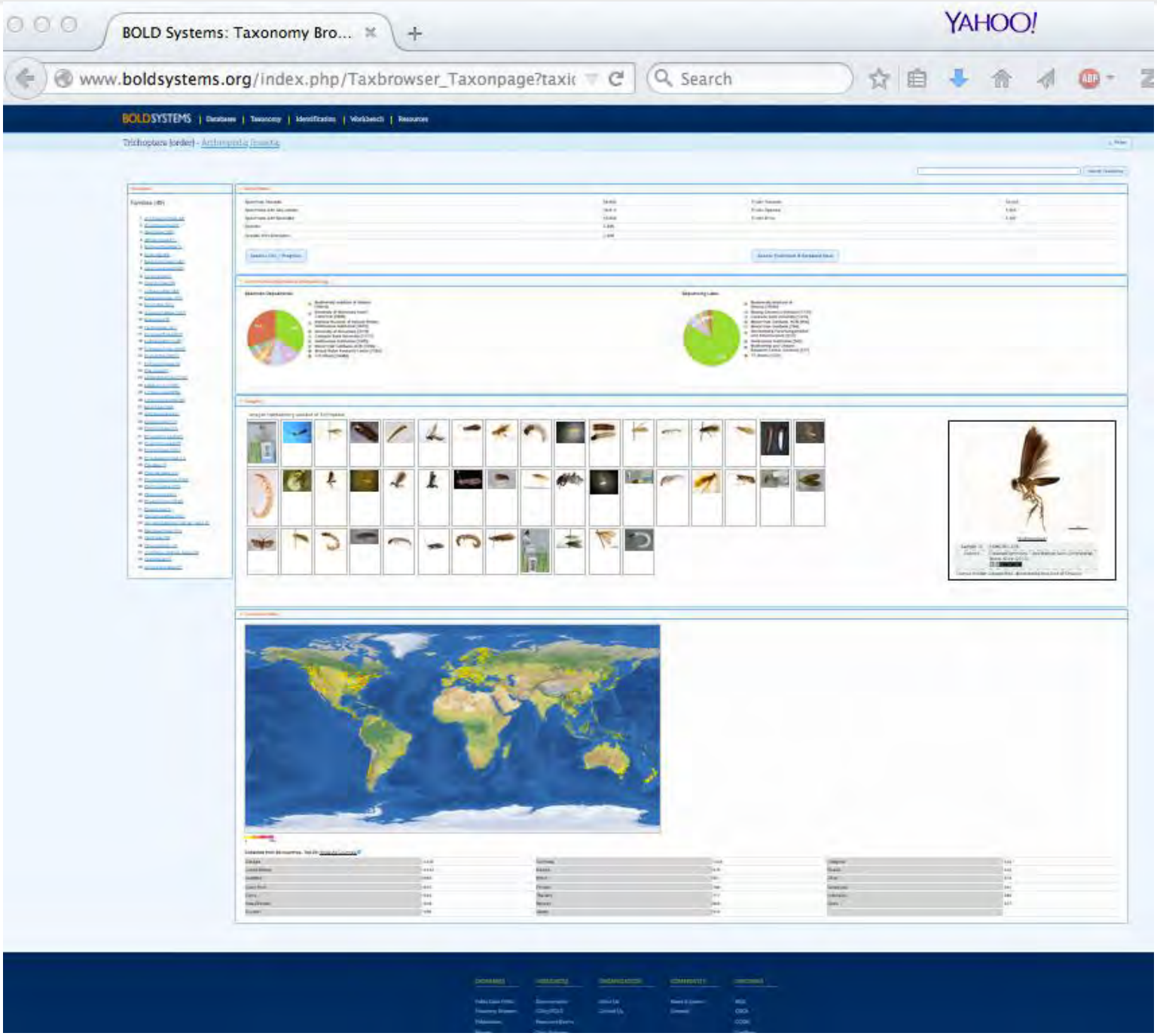
Canadian Centre for DNA Barcoding
Biodiversity Institute of Ontario
Department of Integrative Biology
University of Guelph
579 Gordon Street
Guelph, Ontario N1G 2W1
Canada

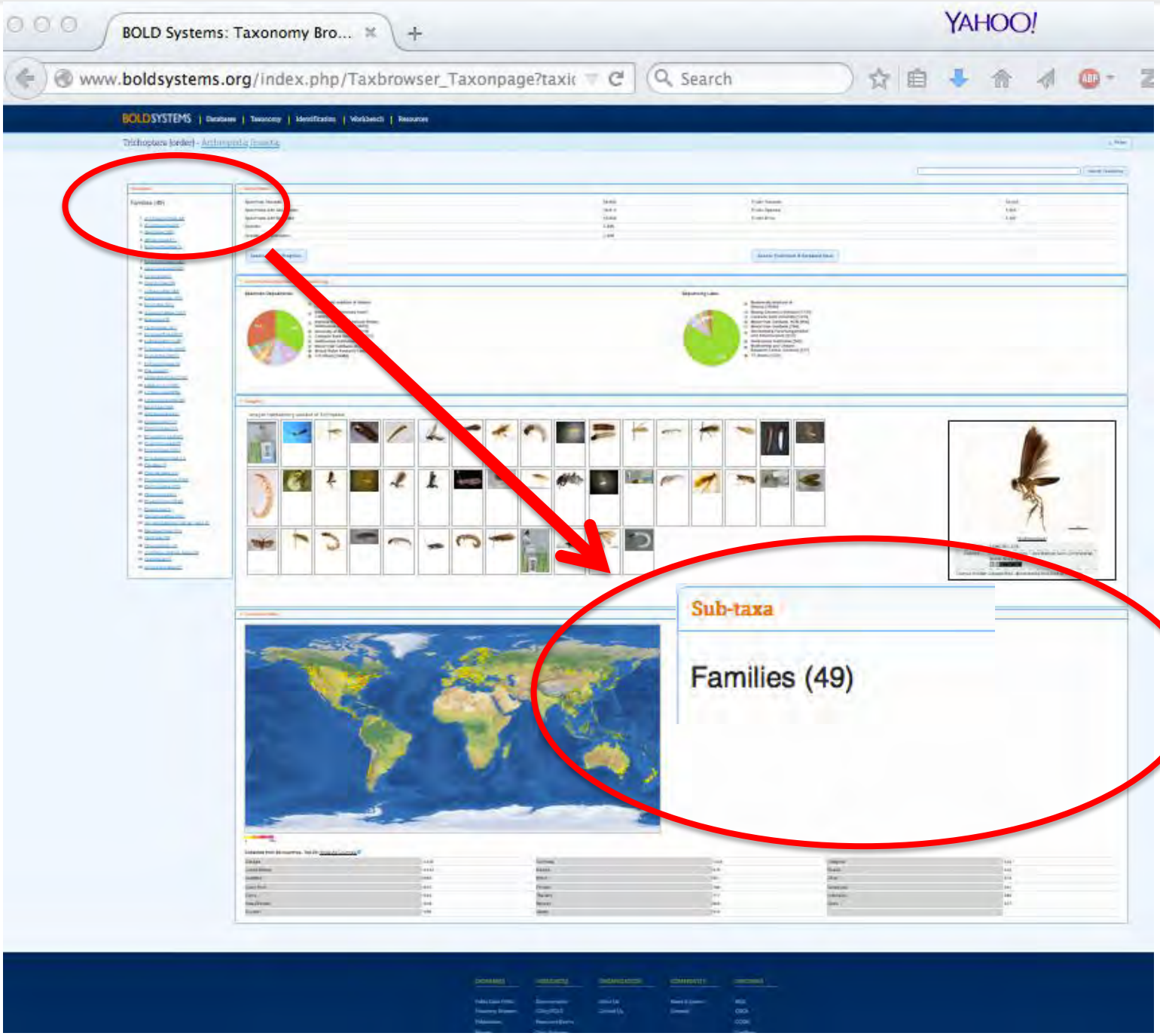
Tel: 1-(519)-824-4120 ext. 53800
Fax: 1-(519)-824-5703
Email: xinzhou@uoguelph.ca

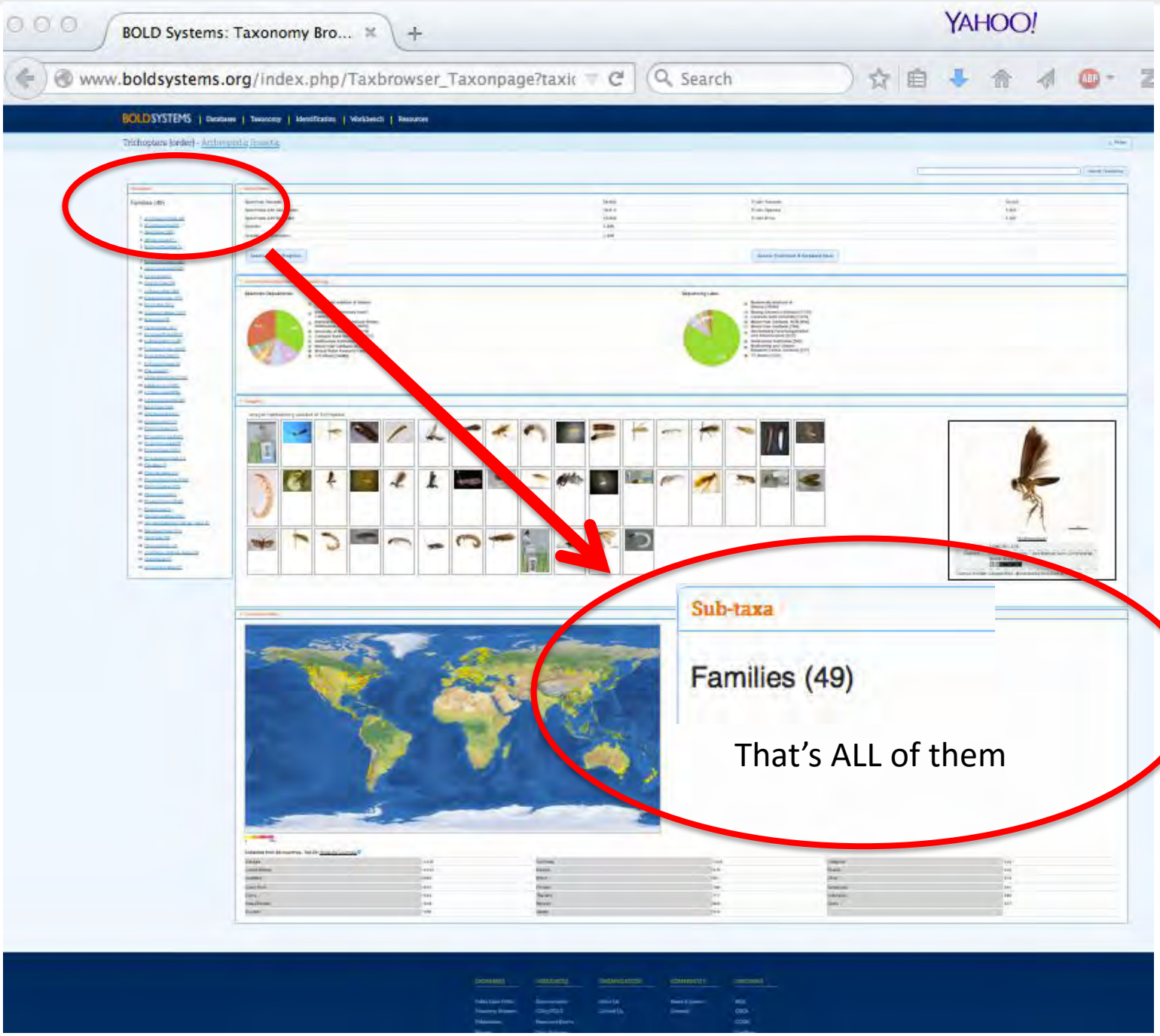
They built the institute we predicted in 2002. Xin was ready



华大基因 BGI


BOLD Systems: Taxonomy Bro...
YAHOO!
www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxi
Search
BOLDSYSTEMS | Databases | Taxonomy | Identification | Workbench | Resources


BOLD Systems: Taxonomy Bro...
YAHOO!
www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxi
Search
BOLDSYSTEMS
Sub-taxa
Families (49)


BOLD Systems: Taxonomy Bro...
YAHOO!
www.boldsystems.org/index.php/Taxbrowser_Taxonpage?taxi
Search
BOLDSYSTEMS
Sub-taxa
Families (49)
That's ALL of them

Trichoptera {order} - Arthropoda; Insecta;

Print

Search Taxonomy

Sub-taxa

Families (48)

1. Anomalopsychidae [19]
2. Antipodoeciidae [16]
3. Apataniidae [467]
4. Atriplectididae [5]
5. Barbarochthonidae [2]
6. Beraeidae [68]
7. Brachycentridae [826]
8. Calamoceratidae [254]
9. Calocidae [109]
10. Chathamiidae [25]
11. Conoesucidae [277]
12. Dipseudopsidae [149]
13. Ecnomidae [220]
14. Glossosomatidae [1090]
15. Goeridae [424]
16. Helicophidae [105]
17. Helicopsychidae [467]
18. Hydrobiosidae [1179]
19. Hydropsychidae [1034
20. Hydroptilidae [25
21. Hydrosalping [2]
22. Kokiriidae [
23. Lepidoston idae [1496]
24. Leptoceri [3983]
25. Limnephilida [348]
26. Limnocentropodi [6]
27. Molannidae [319]
28. Odontoceridae [490]
29. Oeconesidae [90]
30. Petrothrincidae [10]
31. Philopotamidae [2647]
32. Philorheithridae [110]
33. Phryganeidae [1201]
34. Phryganopsychidae [10]
35. Pisuliidae [3]
36. Plectrotarsidae [11]
37. Polycentropodidae [1684]
38. Psychomyiidae [503]
39. Ptilocolepidae [27]

Taxon Description (Wikipedia)

The **caddisflies** are an order, **Trichoptera**, of insects with approximately 12,000 described species. Also called **sedge-flies** or **rail-flies**, they are small moth-like insects having two pairs of hairy membranous wings. They are closely related to Lepidoptera (moths and butterflies) which have scales on their wings, and the two orders together form the superorder Amphiesmenoptera. Caddisflies have aquatic larvae and are found in a wide variety of habitats such as streams, rivers, lakes, ponds, spring seeps, and temporary waters (vernal pools). The larvae of many species make protective cases of silk decorated with gravel, sand, twigs or other debris. The name "Trichoptera" comes from Greek: (, "hair") + (, "wing"). full article at Wikipedia

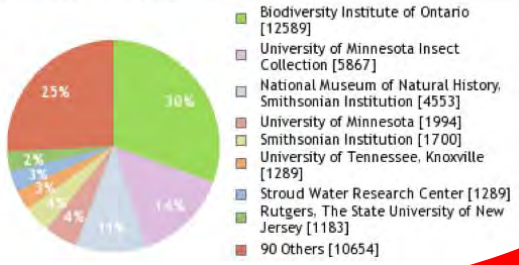
BOLD Stats

Specimen Records:	43,469	Public Records:	27,660
Specimens with Sequences:	35,437	Public Species:	1,254
Specimens with Barcodes:	30,873	Public BINs:	3,964
Species:	4,813		
Species With Barcodes:	3,702		

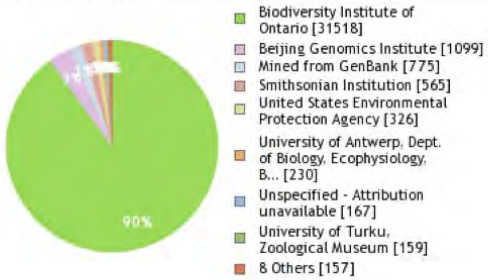
Species List – Progress

Access Published & Released Data

Specimen Depositories:



Sequencing Labs:



Imagery

images representing subtaxa of Trichoptera

Trichoptera {order} - Arthropoda; Insecta;

Print

Sub-taxa

Families (48)

1. Anomalopsychidae [19]
2. Antipodoeciidae [16]
3. Apataniidae [467]
4. Atriplectididae [5]
5. Barbarochthonidae [2]
6. Beraeidae [68]
7. Brachycentridae [826]
8. Calamoceratidae [254]
9. Calocidae [109]
10. Chathamiidae [25]
11. Conoesucidae [277]
12. Dipseudopsidae [149]
13. Ecnomidae [220]
14. Glossosomatidae [1090]
15. Goeridae [424]
16. Helicophidae [105]
17. Helicopsychidae [467]
18. Hydrobiosidae [1179]
19. Hydropsychidae [10340]
20. Hydroptilidae [255
21. Hydrosalpingi [2]
22. Kokiriidae [
23. Lepidoston idae [1496]
24. Leptoceridae [3983]
25. Limnephilida [7348]
26. Limnocentropodi [6]
27. Molannidae [319]
28. Odontoceridae [490]
29. Oeconesidae [90]
30. Petrothrincidae [10]
31. Philopotamidae [2647]
32. Philorheithridae [110]
33. Phryganeidae [1201]
34. Phryganopsychidae [10]
35. Pisuliidae [3]
36. Plectrotarsidae [11]
37. Polycentropodidae [1684]
38. Psychomyiidae [503]
39. Ptilocolepidae [27]

Taxon Description (Wik

The **caddisflies** are a
having two pairs of ha
form the superorder A
and temporary waters
comes from Greek: (

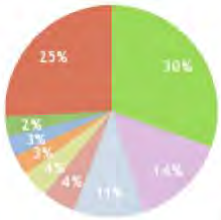
BOLD Stats

Specimen Records:
Specimens with Sequer
Specimens with Barcod
Species:
Species With Barcodes:

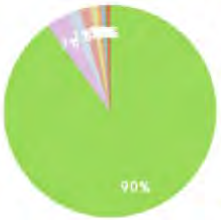
Species List – Prog

Contrib

Specimen Depositorie



Biodiversity Institute of Ontario [12589]
University of Minnesota Insect Collection [5867]
National Museum of Natural History, Smithsonian Institution [4553]
University of Minnesota [1994]
Smithsonian Institution [1700]
University of Tennessee, Knoxville [1289]
Stroud Water Research Center [1289]
Rutgers, The State University of New Jersey [1183]
90 Others [10654]



Biodiversity Institute of Ontario [31518]
Beijing Genomics Institute [1099]
Mined from GenBank [775]
Smithsonian Institution [565]
United States Environmental Protection Agency [326]
University of Antwerp, Dept. of Biology, Ecophysiology, B... [230]
Unspecified - Attribution unavailable [167]
University of Turku, Zoological Museum [159]
8 Others [157]

Imagery

images representing subtaxa of Trichoptera

Trichoptera {order} - Arthropoda; Insecta;

Print

Sub-taxa

Families (48)

1. Anomalopsychidae [19]
2. Antipodoeciidae [16]
3. Apataniidae [467]
4. Atriplectididae [5]
5. Barbarochthonidae [2]
6. Beraeidae [68]
7. Brachycentridae [826]
8. Calamoceratidae [254]
9. Calocidae [109]
10. Chathamiidae [25]
11. Conoesucidae [277]
12. Dipseudopsidae [149]
13. Ecnomidae [220]
14. Glossosomatidae [1090]
15. Goeridae [424]
16. Helicophidae [105]
17. Helicopsychidae [467]
18. Hydrobiosidae [1179]
19. Hydropsychidae [10340]
20. Hydroptilidae [255
21. Hydrosalpingi [2]
22. Kokiriidae [
23. Lepidoston idae [1496]
24. Leptoceridi [3983]
25. Limnephilida [7348]
26. Limnocentropodi [6]
27. Molannidae [319]
28. Odontoceridae [490]
29. Oeconesidae [90]
30. Petrothrincidae [10]
31. Philopotamidae [2647]
32. Philorheithridae [110]
33. Phryganeidae [1201]
34. Phryganopsychidae [10]
35. Pisuliidae [3]
36. Plectrotarsidae [11]
37. Polycentropodidae [1684]
38. Psychomyiidae [503]
39. Ptilocolepidae [27]

Taxon Description (Wik

The **caddisflies** are a
having two pairs of ha
form the superorder A
and temporary waters
comes from Greek: (

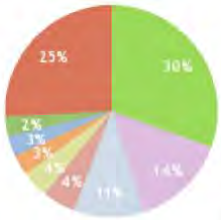
BOLD Stats

Specimen Records:
Specimens with Sequer
Specimens with Barcod
Species:
Species With Barcodes:

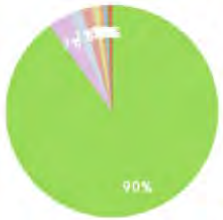
Species List – Prog

Contrib

Specimen Depositorie



[12589]
University of Minnesota Insect Collection [5867]
National Museum of Natural History, Smithsonian Institution [4553]
University of Minnesota [1994]
Smithsonian Institution [1700]
University of Tennessee, Knoxville [1289]
Stroud Water Research Center [1289]
Rutgers, The State University of New Jersey [1183]
90 Others [10654]

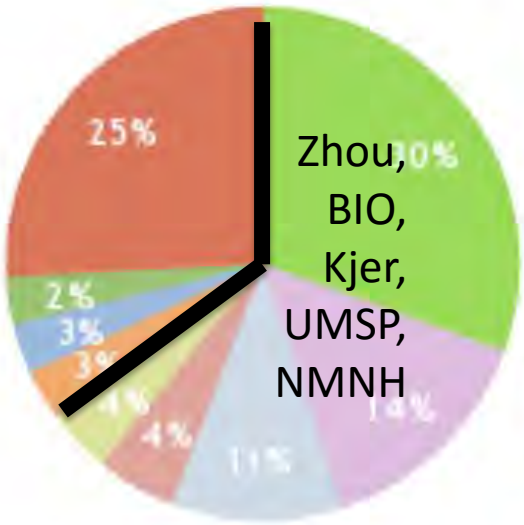


Biodiversity Institute of Ontario [31518]
Beijing Genomics Institute [1099]
Mined from GenBank [775]
Smithsonian Institution [565]
United States Environmental Protection Agency [326]
University of Antwerp, Dept. of Biology, Ecophysiology, B... [230]
Unspecified - Attribution unavailable [167]
University of Turku, Zoological Museum [159]
8 Others [157]

Imagery

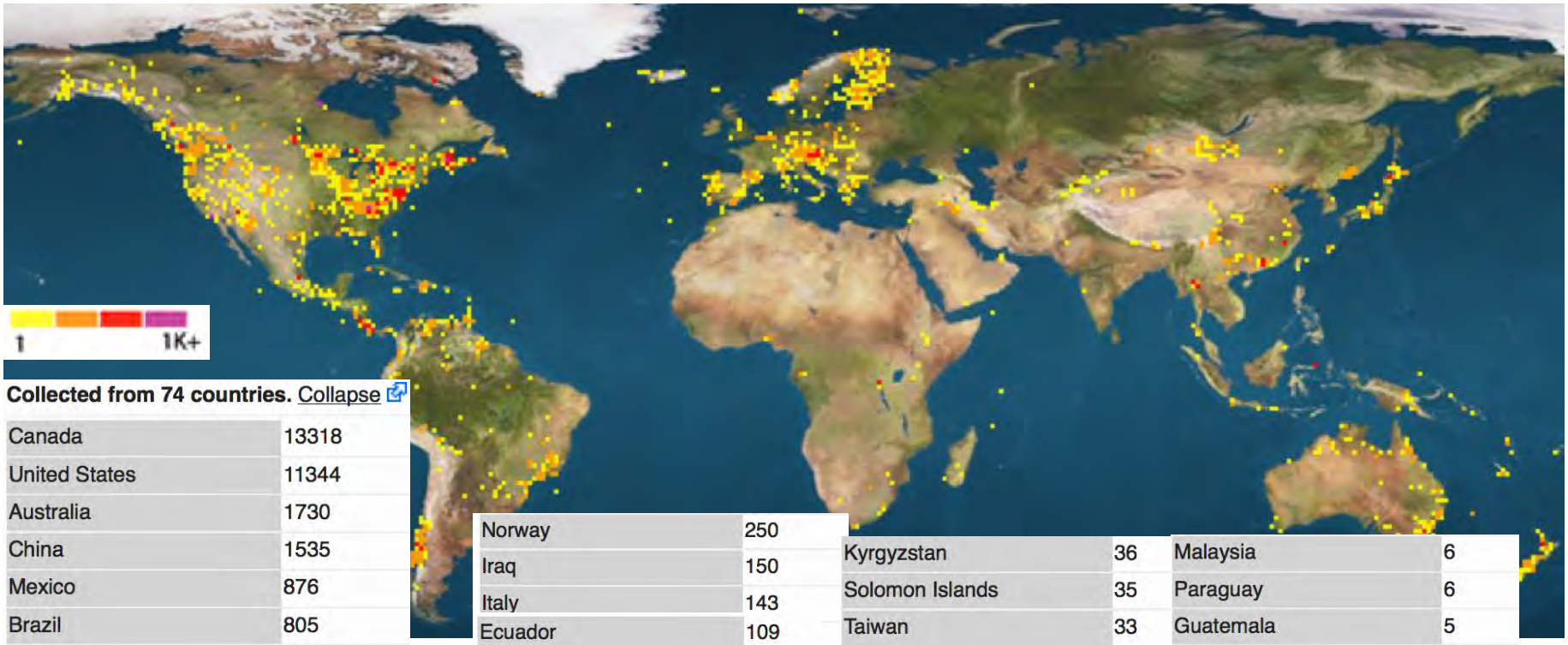
images representing subtaxa of Trichoptera

Specimen Depositories:



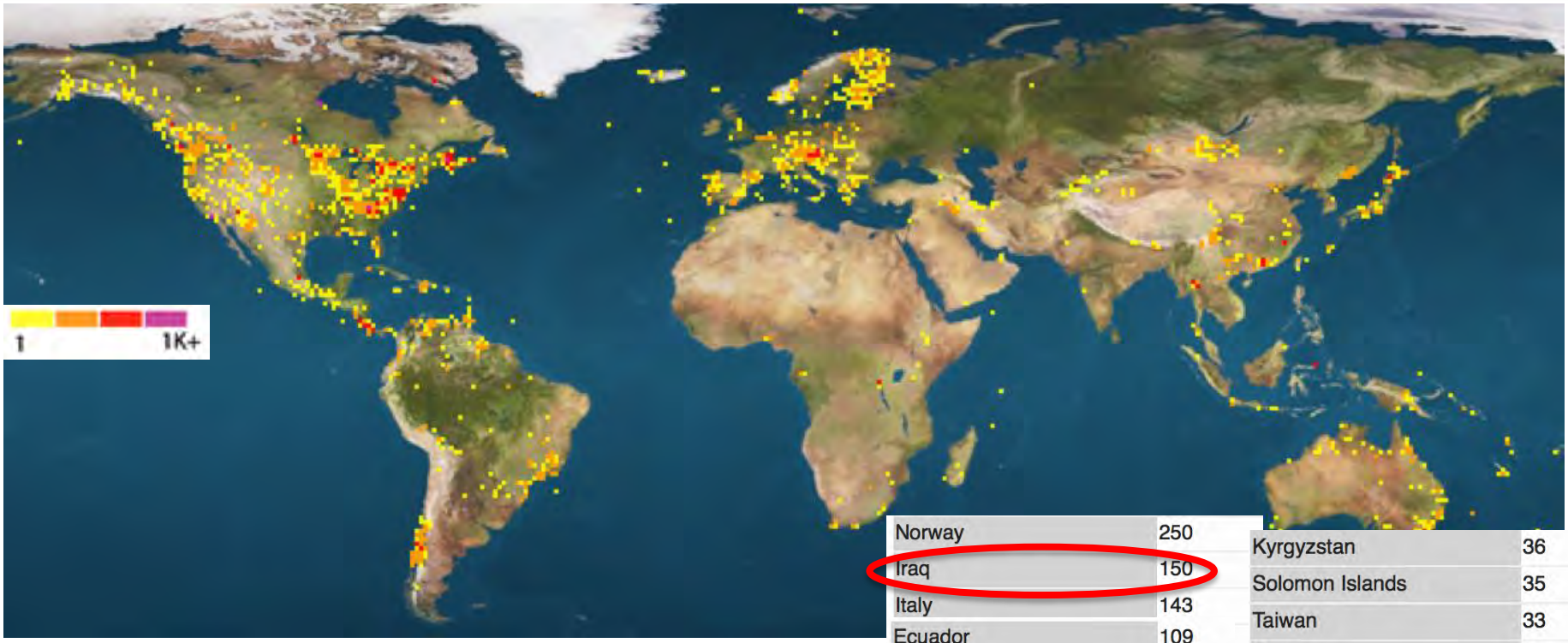
Zhou, BIO, Kjer, UMSP, NMNH

- ✓ Biodiversity Institute of Ontario [12589]
- ✓ University of Minnesota Insect Collection [5867]
- ✓ National Museum of Natural History, Smithsonian Institution [4553]
- ✓ University of Minnesota [1994]
- ✓ Smithsonian Institution [1700]
- University of Tennessee, Knoxville [1289]
- Stroud Water Research Center [1289]
- ✓ Rutgers, The State University of New Jersey [1183]
- 90 Others [10654]



Collected from 74 countries. Collapse

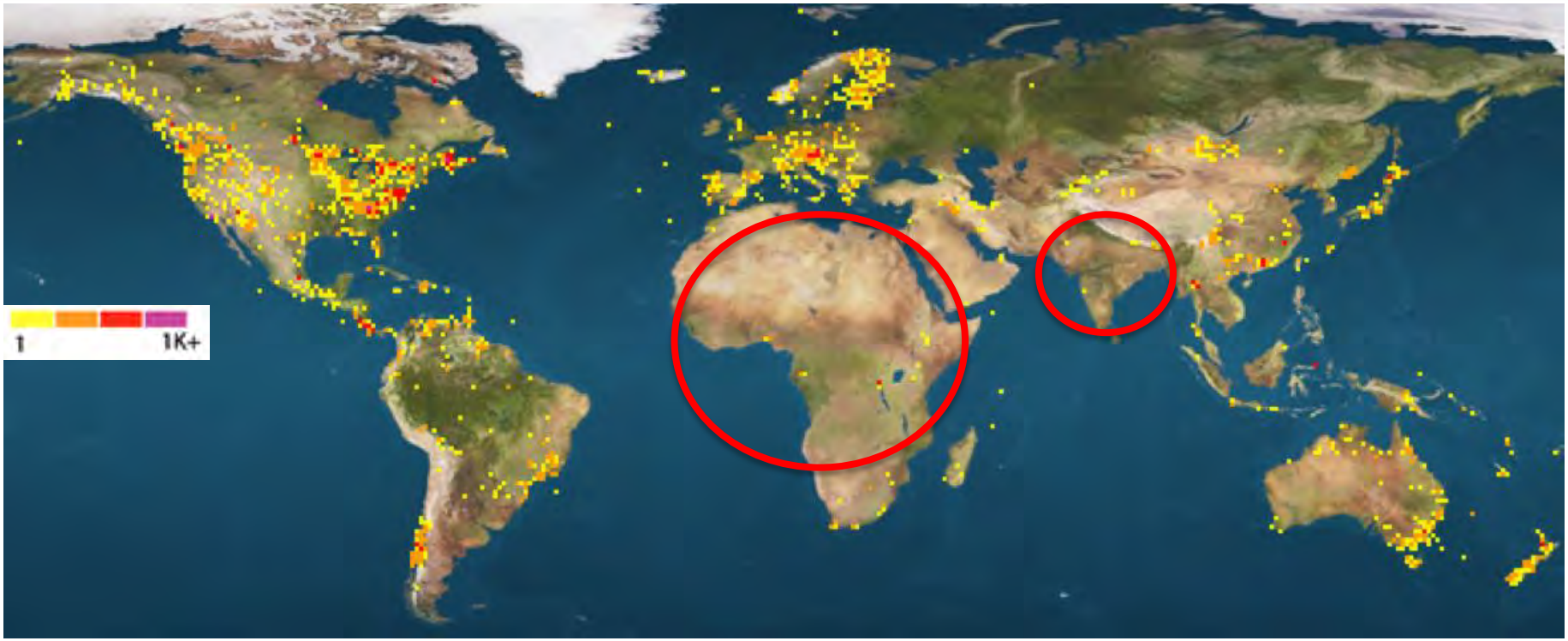
Canada	13318
United States	11344
Australia	1730
China	1535
Mexico	876
Brazil	805
Finland	757
Costa Rica	703
New Zealand	654
Japan	546
Mongolia	530
Thailand	452
Russia	401
Chile	378
Venezuela	368
Germany	351
Indonesia	346
Spain	299
Belgium	280
Austria	258
Norway	250
Iraq	150
Italy	143
Ecuador	109
Peru	98
Papua New Guinea	98
South Africa	95
Bolivia	90
Dominican Republic	74
Nigeria	70
Argentina	58
Madagascar	46
Greece	46
Sweden	44
France	40
Panama	38
Nicaragua	37
Cuba	37
Kyrgyzstan	36
Solomon Islands	35
Taiwan	33
Ghana	30
India	28
Kenya	25
Iran	24
Bhutan	20
Portugal	14
Gabon	14
Guyana	13
United Kingdom	12
Bulgaria	12
Faeroe Islands	11
Switzerland	10
New Caledonia	9
Laos	9
Malaysia	6
Paraguay	6
Guatemala	5
Netherlands	5
Pakistan	5
Seychelles	3
Kazakhstan	3
Micronesia	2
South Korea	2
Romania	2
Denmark	2
Poland	2
Trinidad and Tobago	1
Puerto Rico	1
Sri Lanka	1
Ethiopia	1
Croatia	1



Iraq, Bhutan, Cuba, the Faeroe Islands… This is not a quick and dirty sample of backyard insects!

Norway	250
Iraq	150
Italy	143
Ecuador	109
Peru	98
Papua New Guinea	98
South Africa	95
Bolivia	90
Dominican Republic	74
Nigeria	70
Argentina	58
Madagascar	46
Greece	46
Sweden	44
France	40
Panama	38
Nicaragua	37
Cuba	37

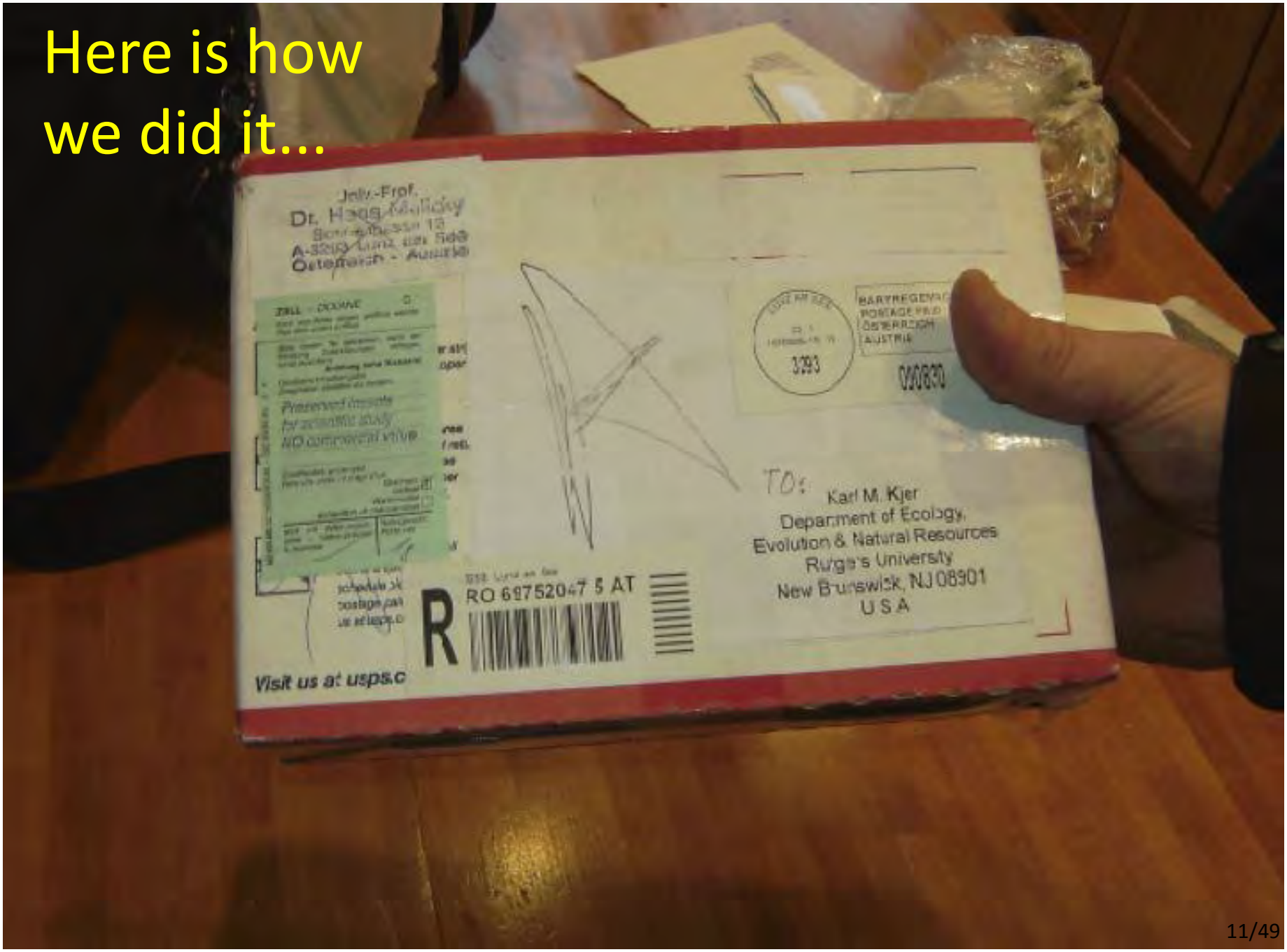
Kyrgyzstan	36
Solomon Islands	35
Taiwan	33
Ghana	30
India	28
Kenya	25
Iran	24
Bhutan	20
Portugal	14
Gabon	14
Guyana	13
United Kingdom	12
Bulgaria	12
Faeroe Islands	11
Switzerland	10
New Caledonia	9
Laos	9



Two big holes are a challenge

(and an opportunity)

Here is how we did it...



TO:
Karl M. Kjer
Department of Ecology
Evolution & Natural Resources
Rutgers University
New Brunswick, NJ 08901
U S A

Make it easy for collaborators, and partner with them on projects
5M Guanidine thiocyanate (GIT) Sigma G6639

The database is huge, international, and past a critical tipping point. (it's useful)

Useful for Identification: We used it in 1KITE to identify species and evaluate contamination

…more later

Barbarochthon sp.	Top Hit: Arthropoda - Trichoptera-Barbarochthon sp.(99.07%)
Anisocentropus **pallidus**	Top Hit: Arthropoda - Trichoptera - Anisocentropus pallidus (99.84%
Tamasia acuta	Top Hit: Arthropoda-Trichoptera - Tamisia acuta (99.07%)
Chathamia integripennis	Arthropoda-Insecta-Trichoptera-Calamoceratidae Anisocentropus
Conoesucus sp.	Top Hit: Arthropoda-Trichoptera (100%)
Costora delora	Top Hit: Arthropoda-Trichoptera Costora delora (98.75%)
Phylocentropus carolinus	Top Hit: Arthropoda-Trichoptera-Hydropsyche decalda (100%)
Dipseudopsis capensis	Arthropoda-Insecta-Trichoptera-Calamoceratidae-Anisocentropus
Parecnomina resima	Top Hit: Arthropoda-Trichoptera-Goera fuscula (99.2%)
Agapetus tomus	Top Hit: Arthropoda - Trichoptera - Agapetus tomus (99.85%)
Glossosoma nigrior	Top Hit: Arthropoda-Trichoptera - Glossosoma nigrior (100%)
Goera fuscula	Top Hit: Arthropoda-Trichoptera-Goera fuscula (99.2%)
Goera horni	Top Hit: Arthropoda - Trichoptera - Goera horni (99.23%)
Alloecella grisea	Top Hit: Arthropoda-Trichoptera-Fattigia pele (100%)
Helicopsyche borealis	Top Hit: Arthropoda - Trichoptera - Helicopsyche borealis (100%)
Hydrobiosidae sp.	No match
Apsilochorema sutshanum	Top Hit: Arthropoda-Trichoptera (98%)
Arctopsyche sp.	Top Hit: Arthropoda-Trichoptera-Triaenodes unanimis (100%)
Diplectrona sp.	Top Hit: Arthropoda - Trichoptera (97.67%)

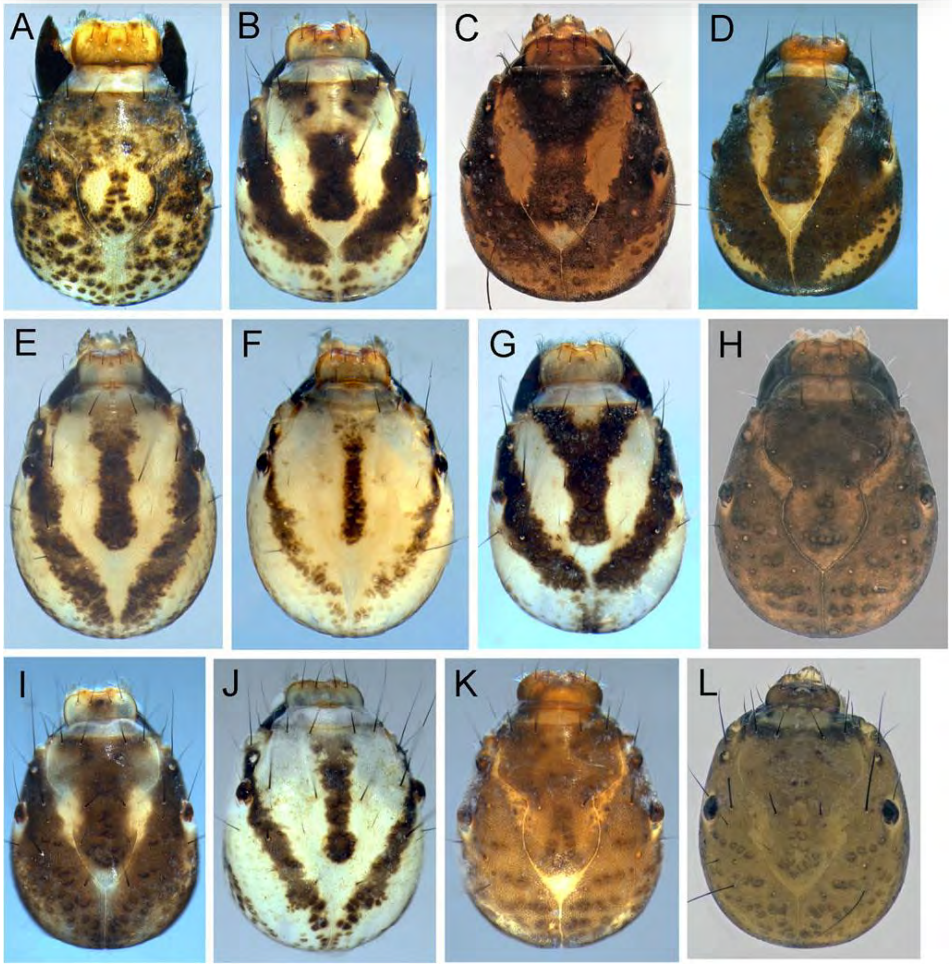
Ruiter *et al. BMC Ecology* 2013, **13**:5
http://www.biomedcentral.com/1472-6785/13/5

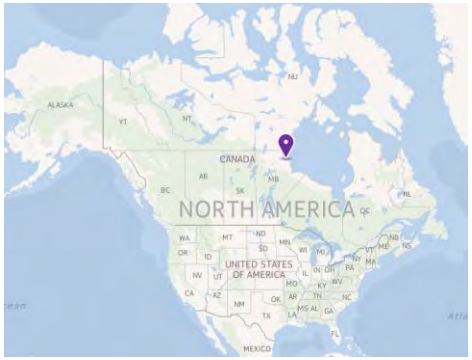
BMC Ecology

RESEARCH ARTICLE Open Access

DNA barcoding facilitates associations and diagnoses for Trichoptera larvae of the Churchill (Manitoba, Canada) area

David E Ruiter[1], Elizabeth E Boyle[2] and Xin Zhou[3*]



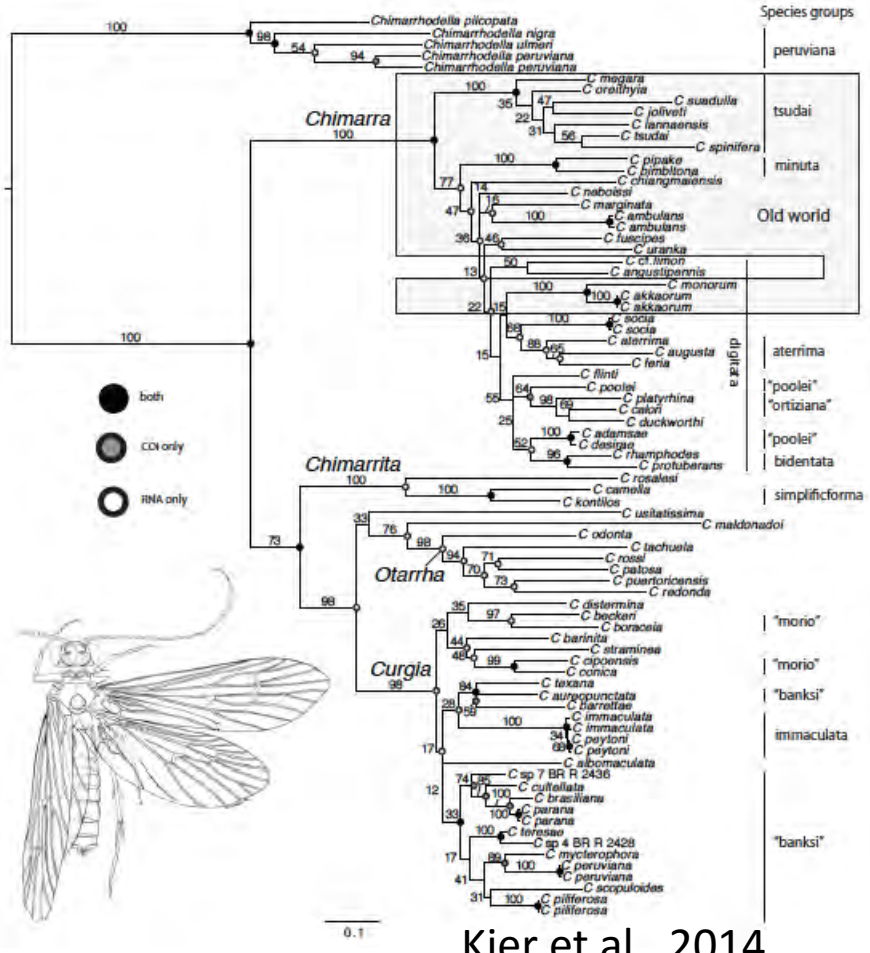


- **3,310 COI barcodes (adults + larvae)**
- **2/3 larval caddisflies from Churchill now identifiable**

Useful for Identification:
Many applications for larval associations

Ruiter, Boyle, Zhou*. 2013. *BMC Ecology*

The database is useful for phylogeny: Add another gene for a species-level phylogeny



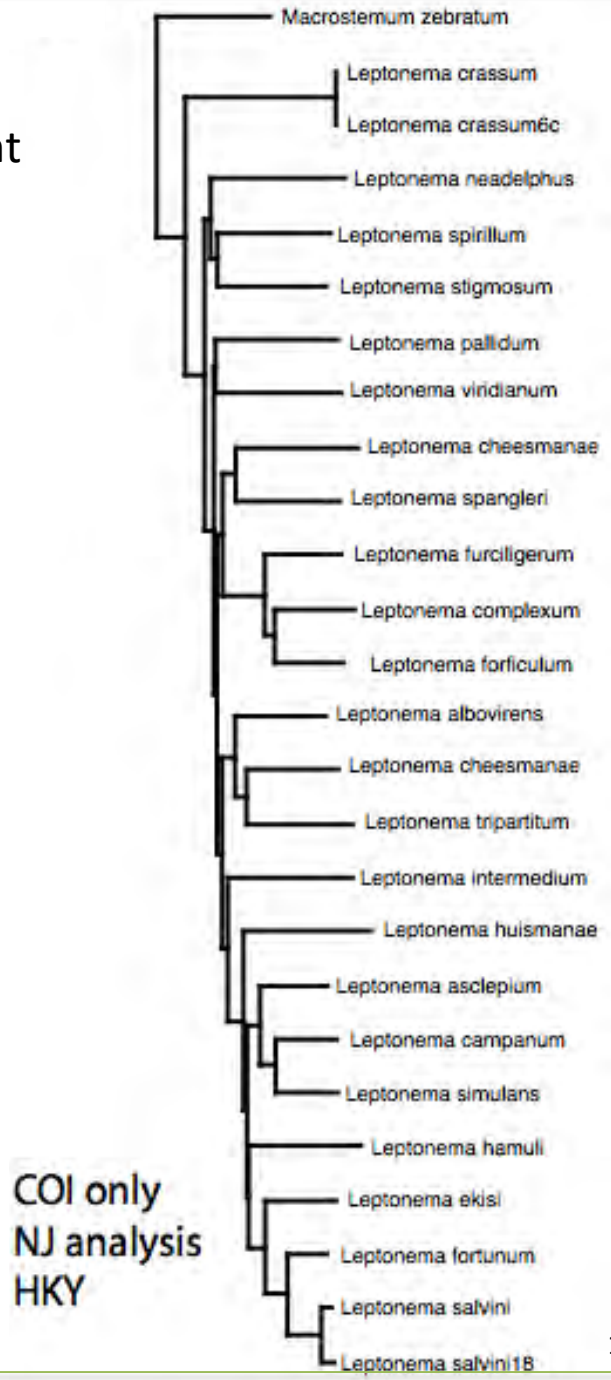
Kjer et al., 2014

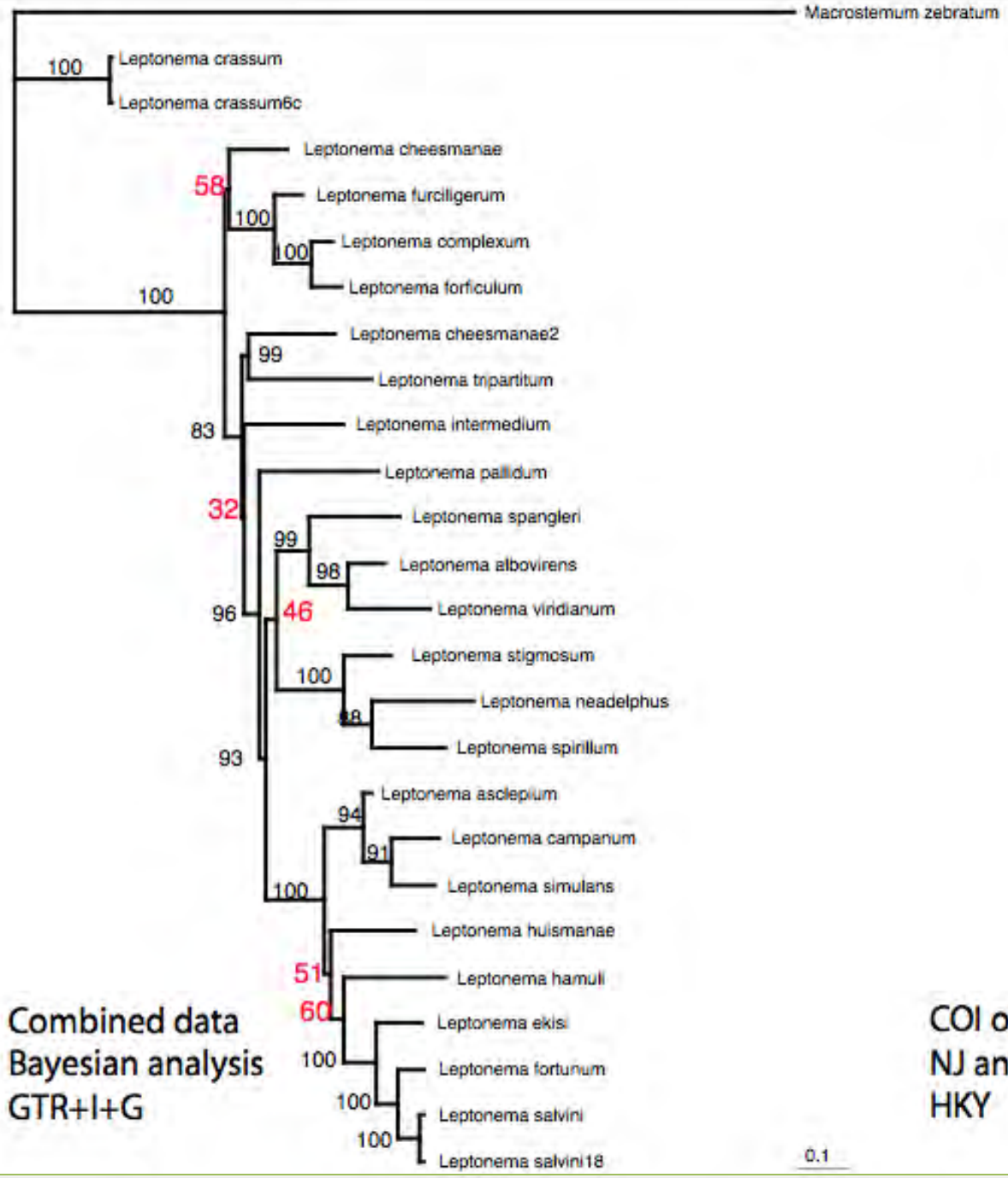
COI is ideal for barcoding because a large enough fragment is not expected to be identical across species

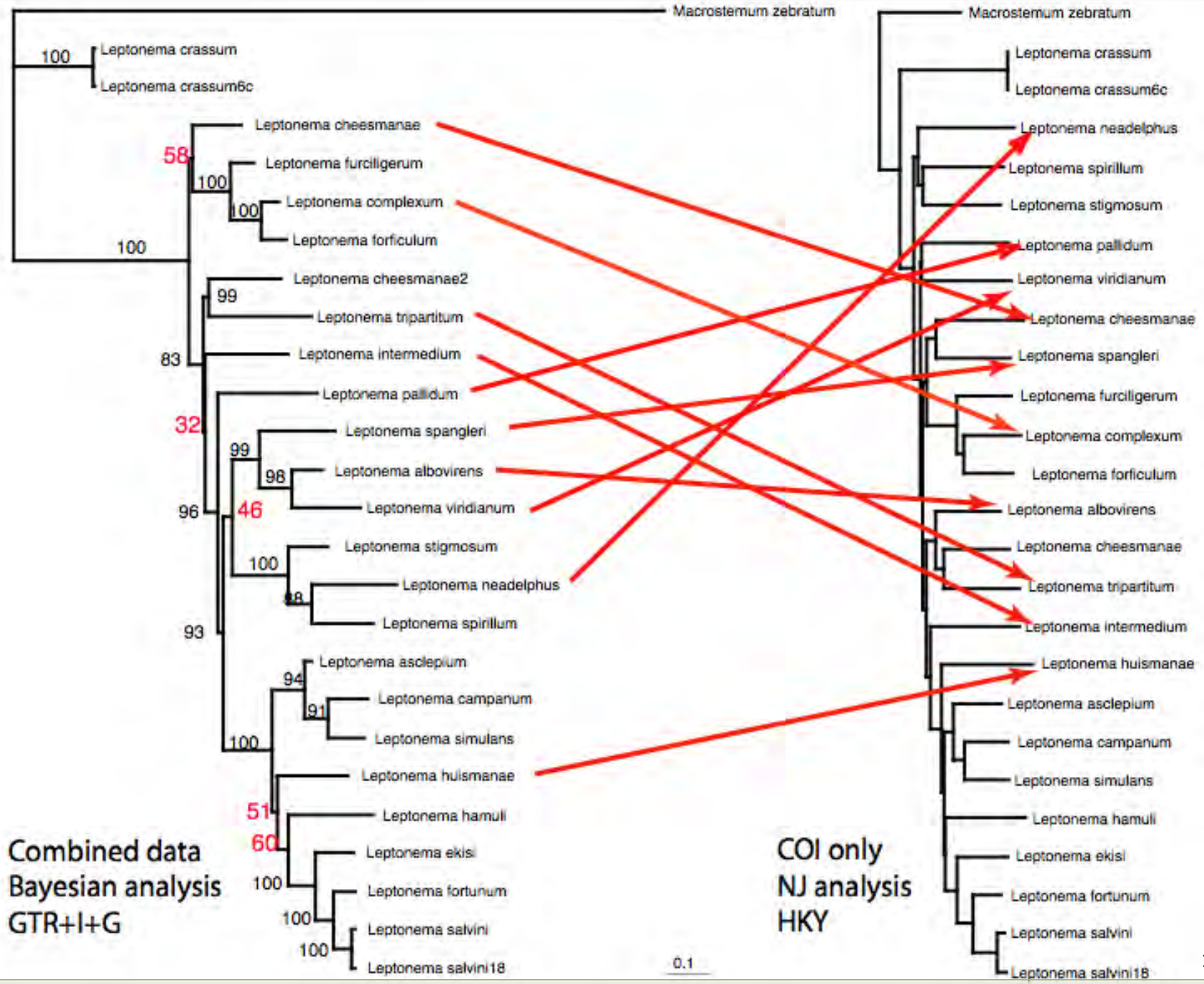
But fast evolving genes are problematic for phylogenetics

A phylogram with long terminal branches, and short internodes is the classic “hard” problem in phylogenetics.

COI, by itself, especially when analyzed with NJ is at best, an interesting hypothesis.

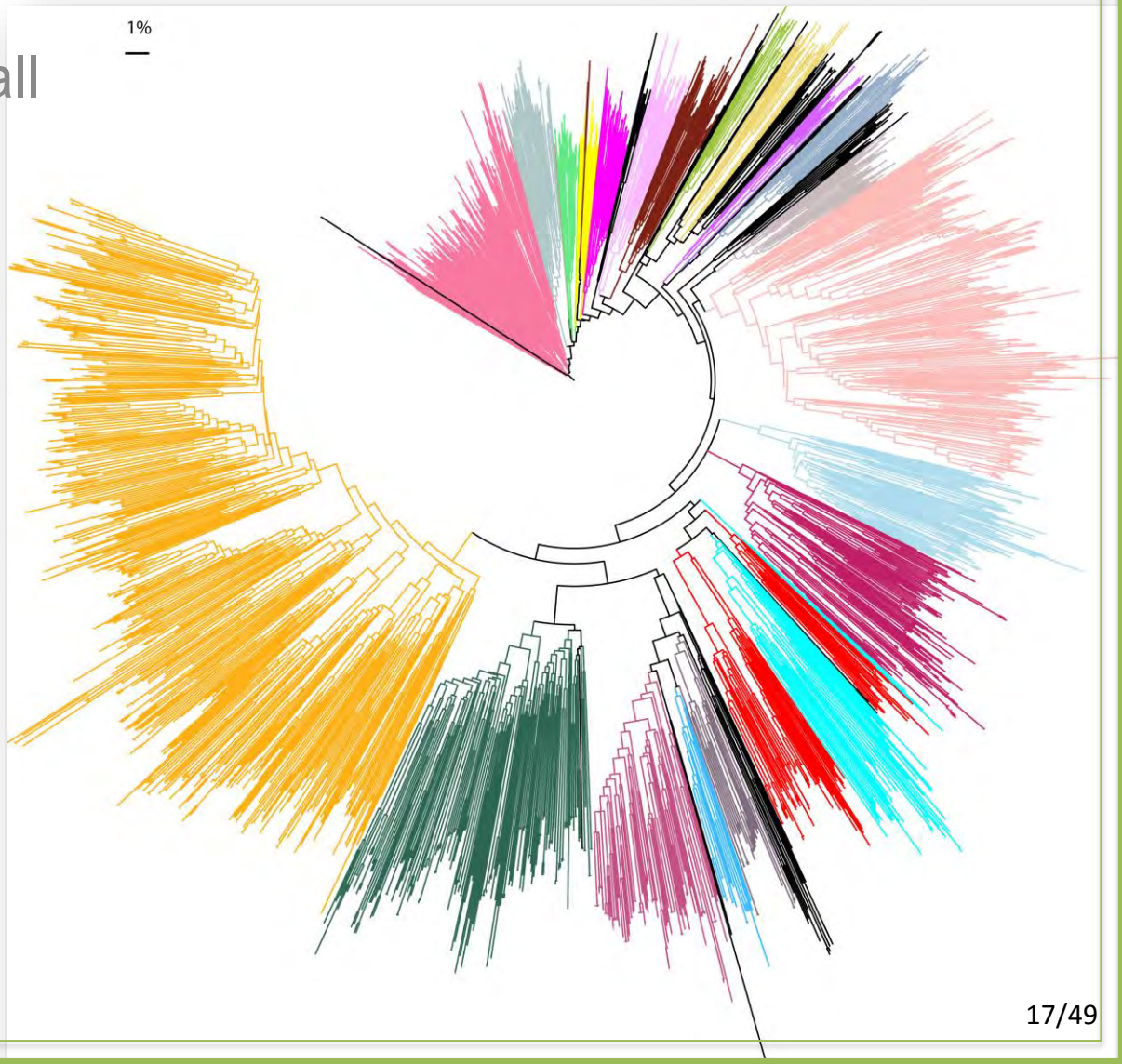


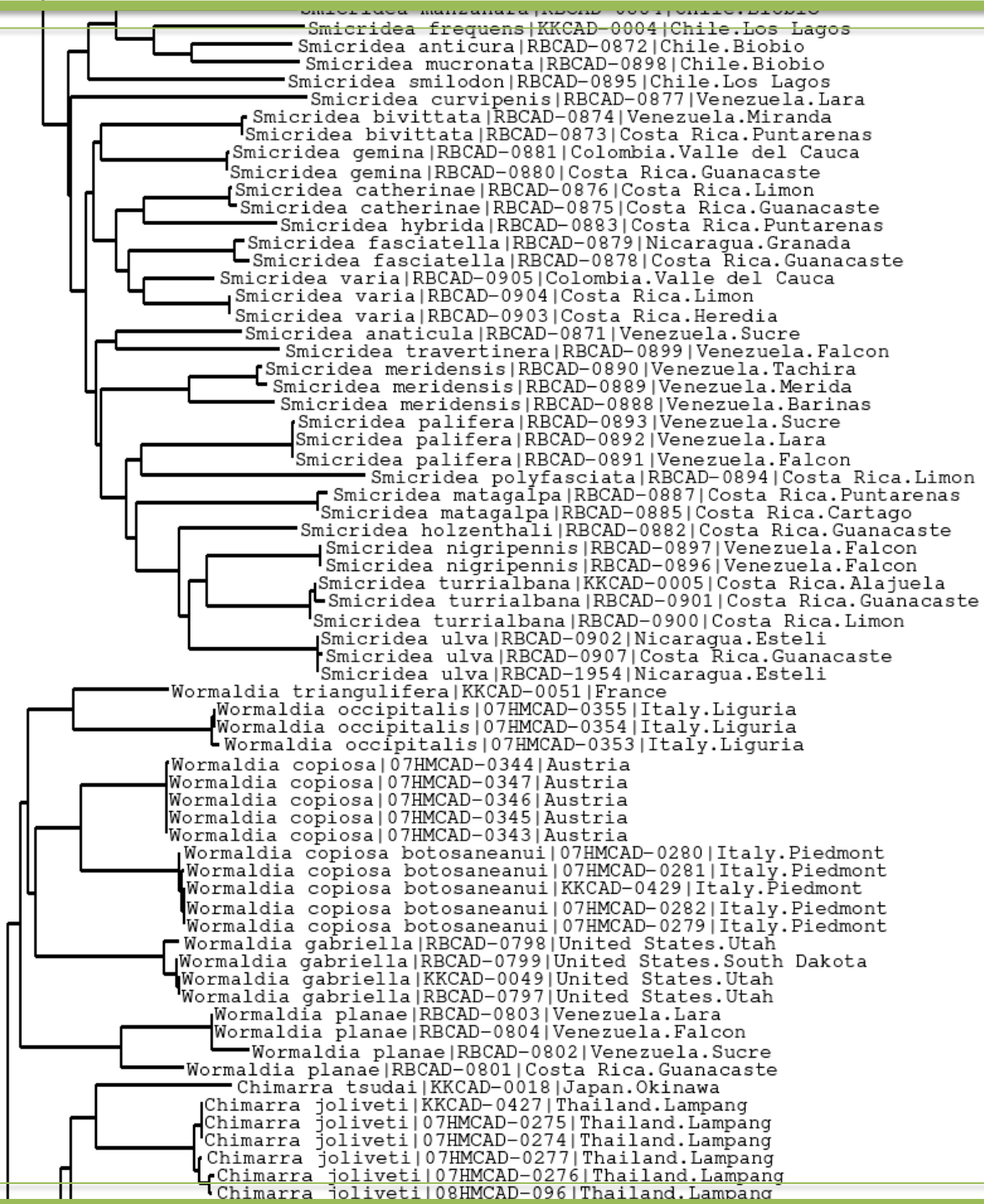

Macrostemum zebratum
100
Leptonema crassum
Leptonema crassum6c
Leptonema cheesmanae
58
Leptonema furciligerum
100
Leptonema complexum
100
Leptonema forficulum
100
Leptonema cheesmanae2
99
Leptonema tripartitum
Leptonema intermedium
83
Leptonema pallidum
32
Leptonema spangleri
99
98
Leptonema albovirens
96
46
Leptonema viridianum
Leptonema stigmosum
100
Leptonema neadelphus
Leptonema spirillum
93
Leptonema asclepium
94
Leptonema campanum
91
Leptonema simulans
100
Leptonema huismanae
51
Leptonema hamuli
60
Leptonema ekisi
100
Leptonema fortunum
100
Leptonema salvini
100
Leptonema salvini18
0.1
Combined data
Bayesian analysis
GTR+I+G
Macrostemum zebratum
Leptonema crassum
Leptonema crassum6c
Leptonema neadelphus
Leptonema spirillum
Leptonema stigmosum
Leptonema pallidum
Leptonema viridianum
Leptonema cheesmanae
Leptonema spangleri
Leptonema furciligerum
Leptonema complexum
Leptonema forficulum
Leptonema albovirens
Leptonema cheesmanae
Leptonema tripartitum
Leptonema intermedium
Leptonema huismanae
Leptonema asclepium
Leptonema campanum
Leptonema simulans
Leptonema hamuli
Leptonema ekisi
Leptonema fortunum
Leptonema salvini
Leptonema salvini18
COI only
NJ analysis
HKY


Macrostemum zebratum
Leptonema crassum
Leptonema crassum6c
Leptonema cheesmanae
Leptonema furciligerum
Leptonema complexum
Leptonema forficulum
Leptonema cheesmanae2
Leptonema tripartitum
Leptonema intermedium
Leptonema pallidum
Leptonema spangleri
Leptonema albovirens
Leptonema viridianum
Leptonema stigmosum
Leptonema neadelphus
Leptonema spirillum
Leptonema asclepium
Leptonema campanum
Leptonema simulans
Leptonema huismanae
Leptonema hamuli
Leptonema ekisi
Leptonema fortunum
Leptonema salvini
Leptonema salvini18
100
58
100
100
100
83
99
32
99
98
96
46
100
93
94
91
100
51
60
100
100
100
Combined data
Bayesian analysis
GTR+I+G
0.1
Macrostemum zebratum
Leptonema crassum
Leptonema crassum6c
Leptonema neadelphus
Leptonema spirillum
Leptonema stigmosum
Leptonema pallidum
Leptonema viridianum
Leptonema cheesmanae
Leptonema spangleri
Leptonema furciligerum
Leptonema complexum
Leptonema forficulum
Leptonema albovirens
Leptonema cheesmanae
Leptonema tripartitum
Leptonema intermedium
Leptonema huismanae
Leptonema asclepium
Leptonema campanum
Leptonema simulans
Leptonema hamuli
Leptonema ekisi
Leptonema fortunum
Leptonema salvini
Leptonema salvini18
COI only
NJ analysis
HKY

The Trichoptera tree of life

45,000 barcodes from all families



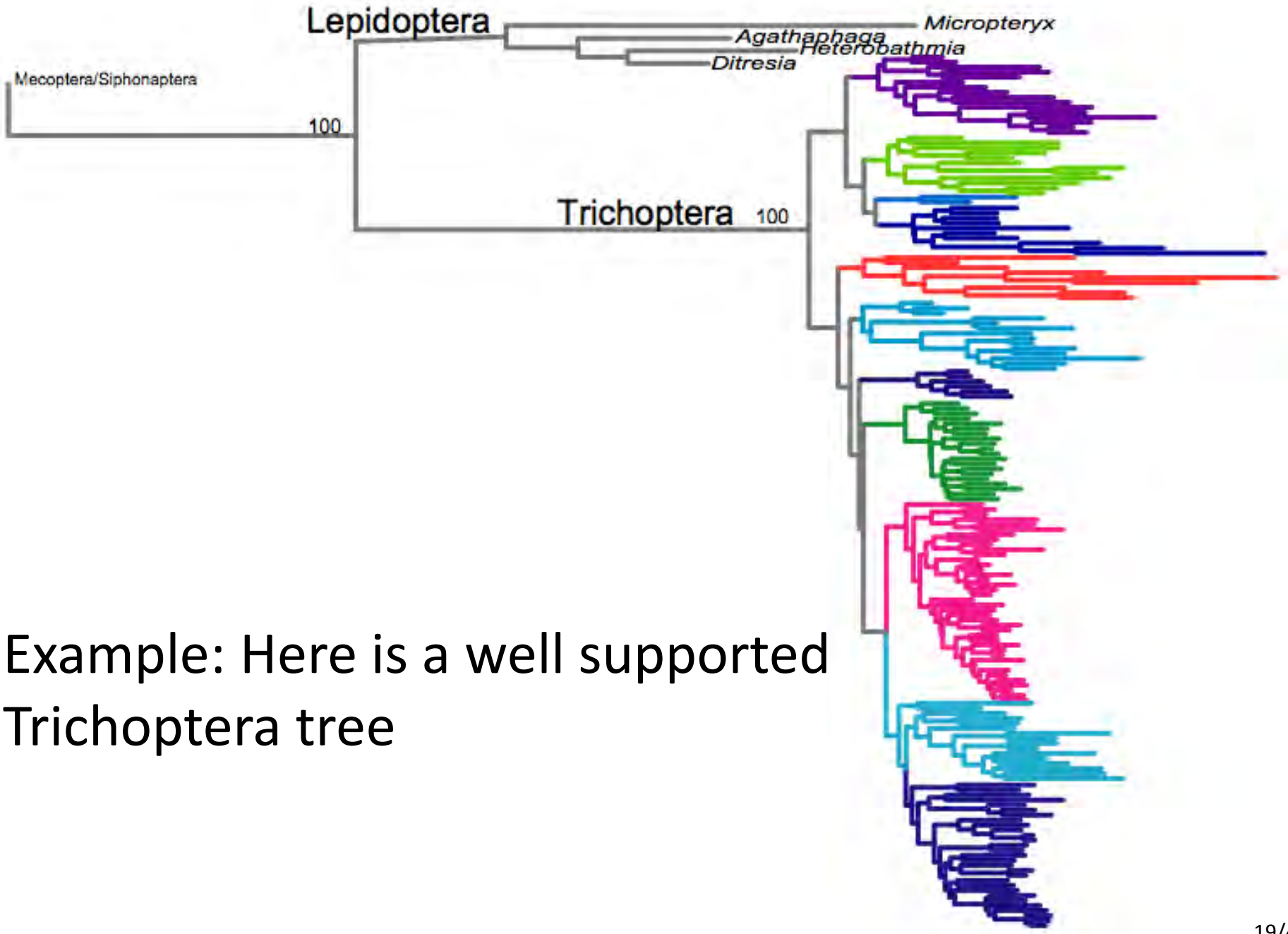


We want to complete the Trichoptera "tree of life"

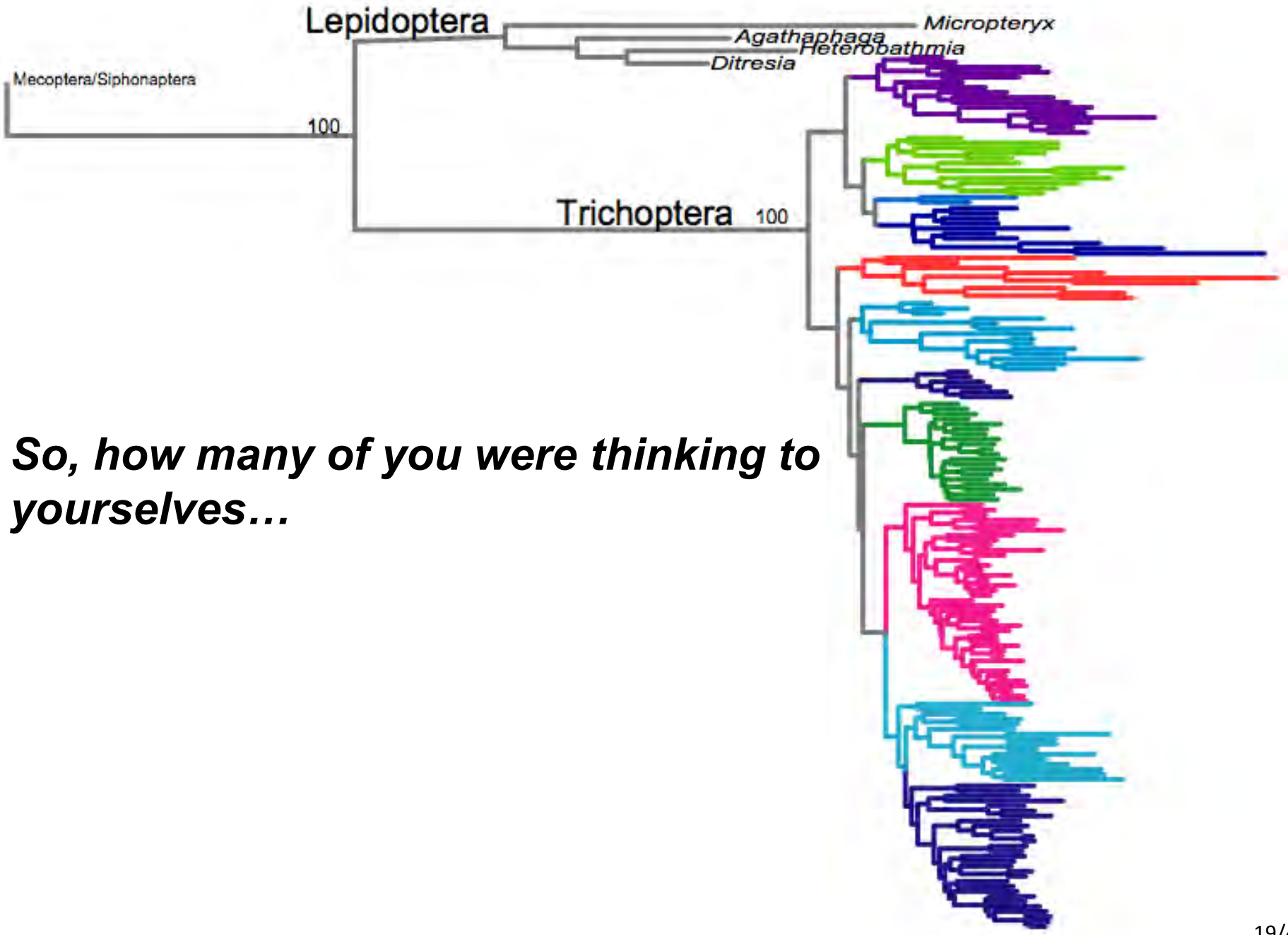
Strategy:

Fix nodes with other data, and then fill out the terminal taxa with COI

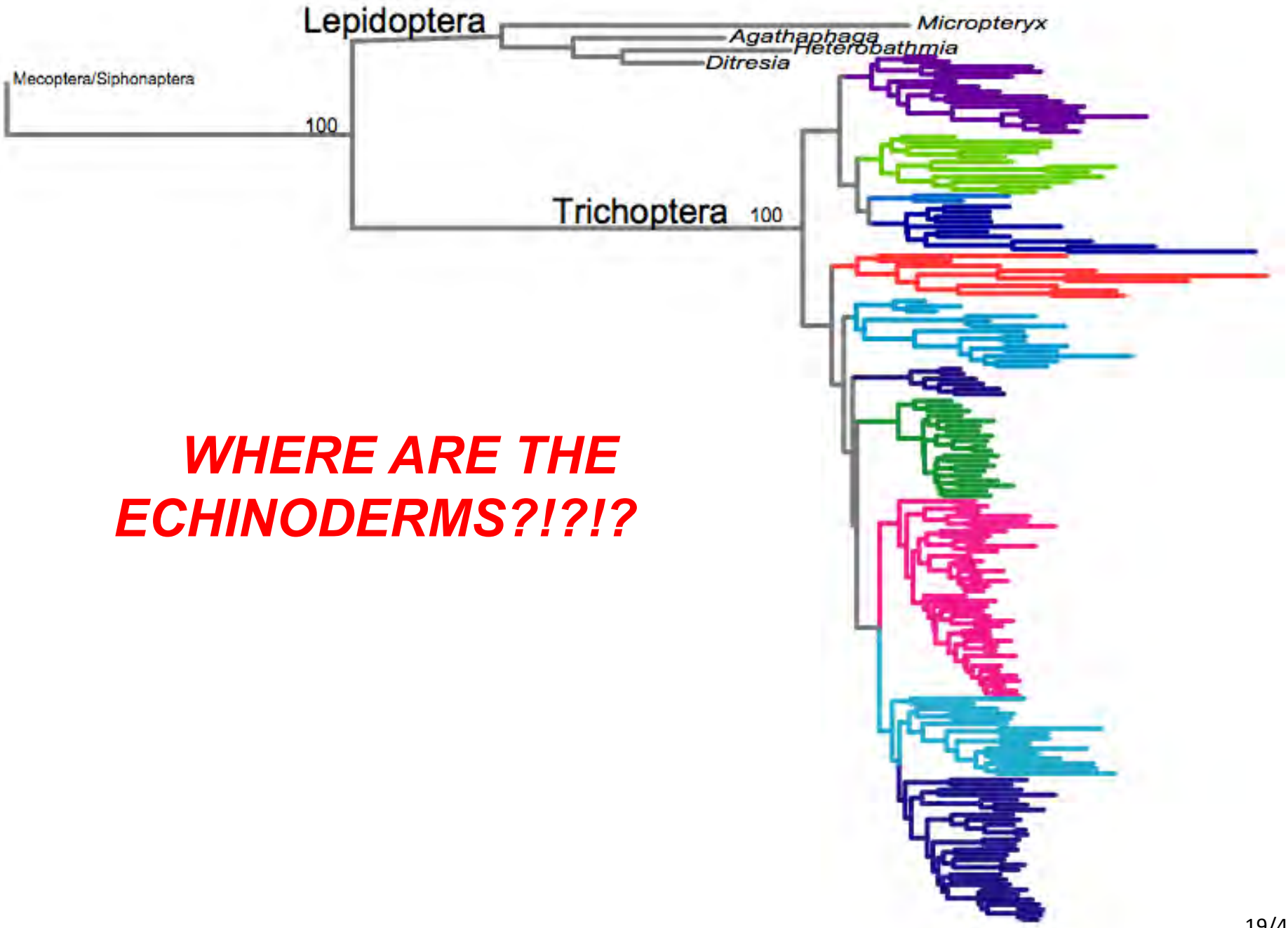
Can we do this?



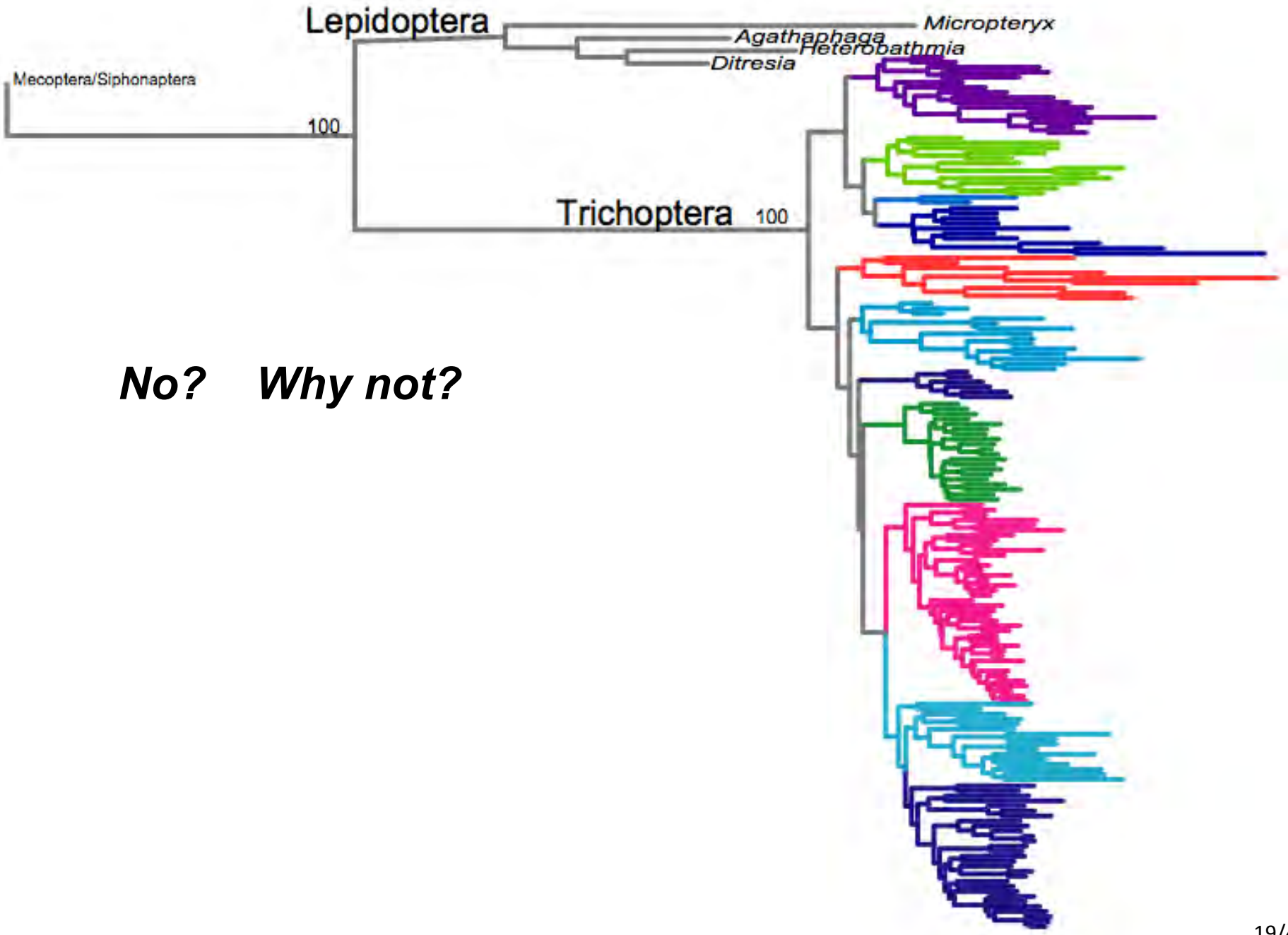
Example: Here is a well supported Trichoptera tree



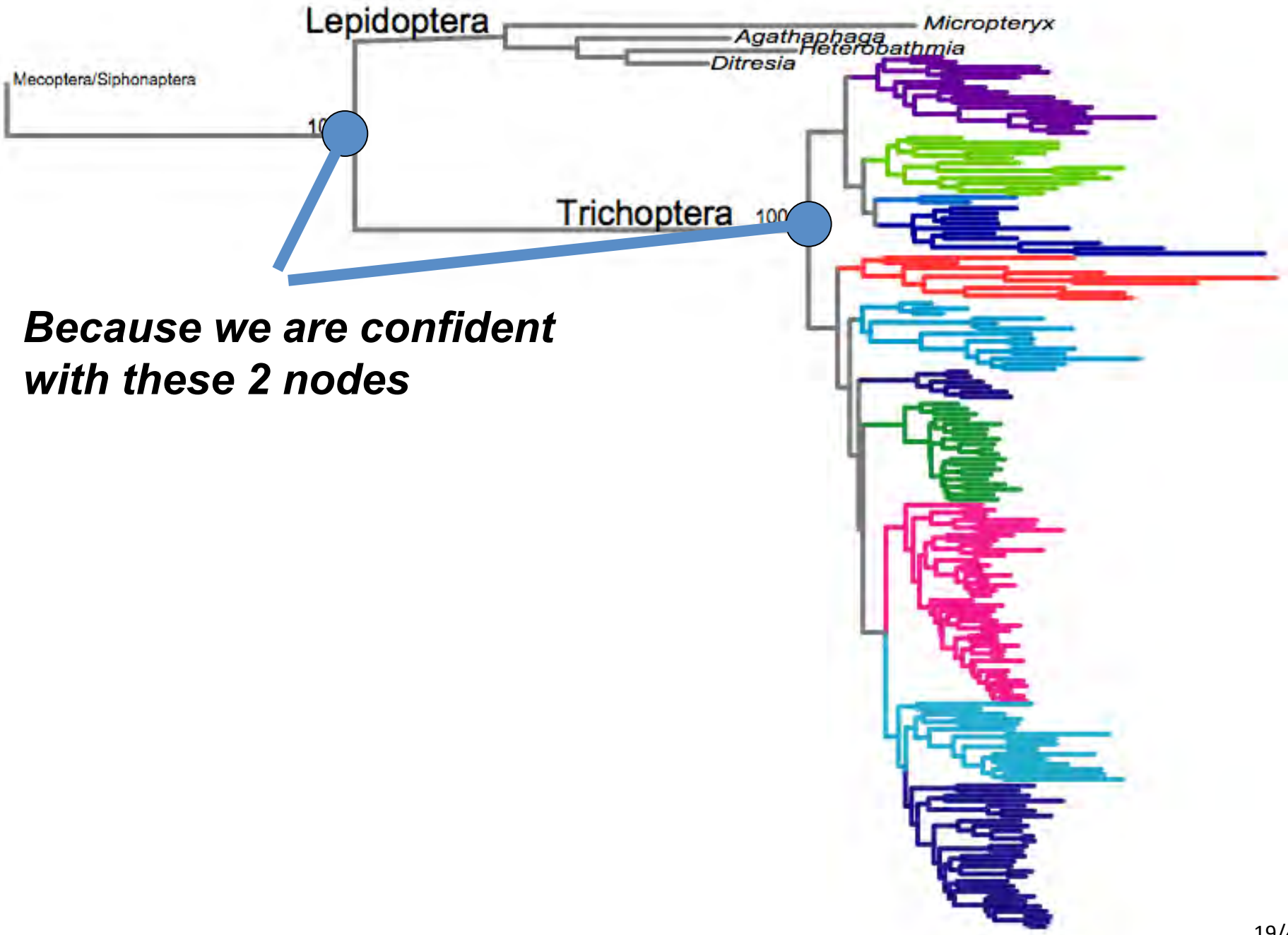
So, how many of you were thinking to yourselves…



WHERE ARE THE ECHINODERMS?!?!?

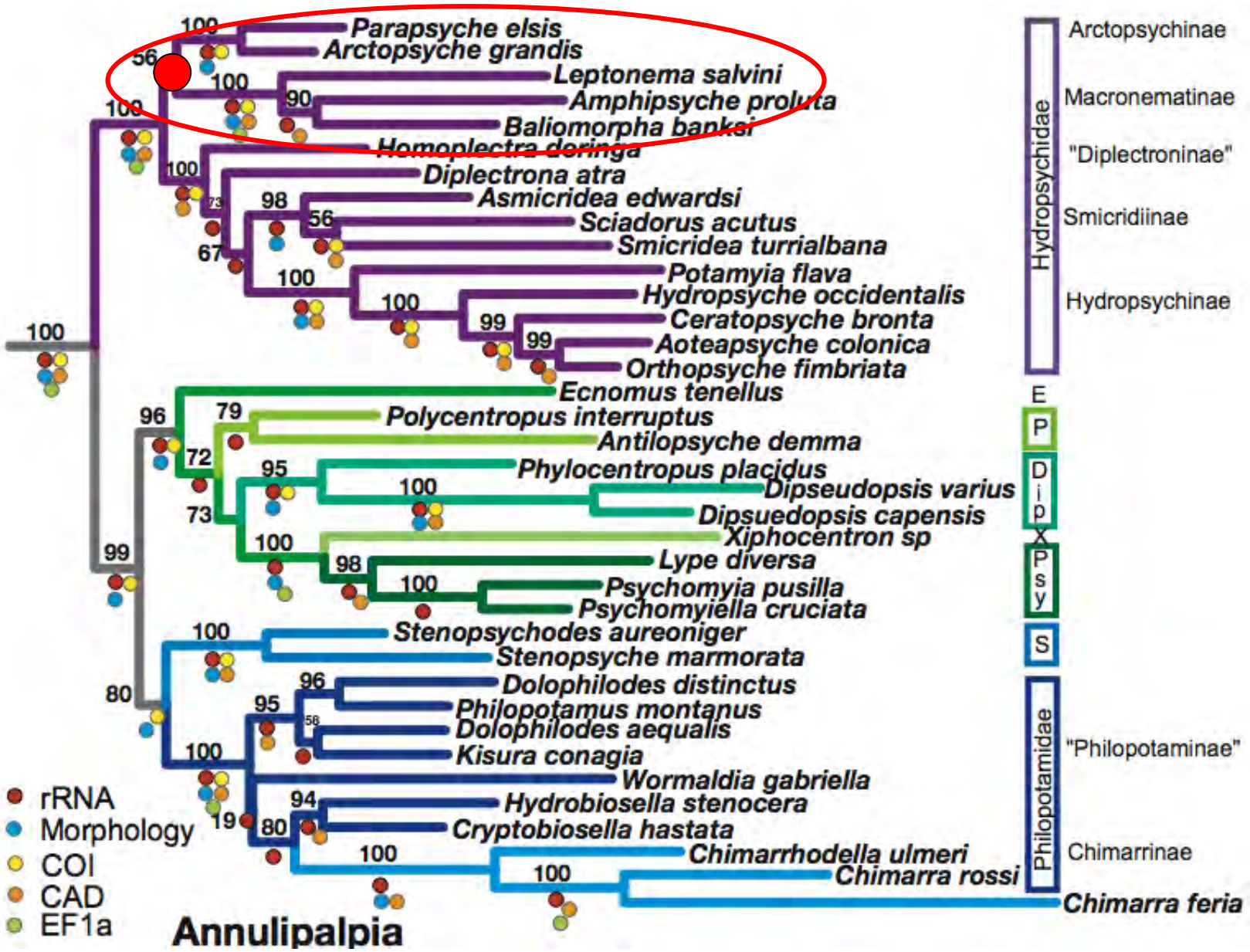

Lepidoptera
Micropteryx
Agathaphaga
Heterobathmia
Ditresia
Mecoptera/Siphonaptera
100
Trichoptera
100

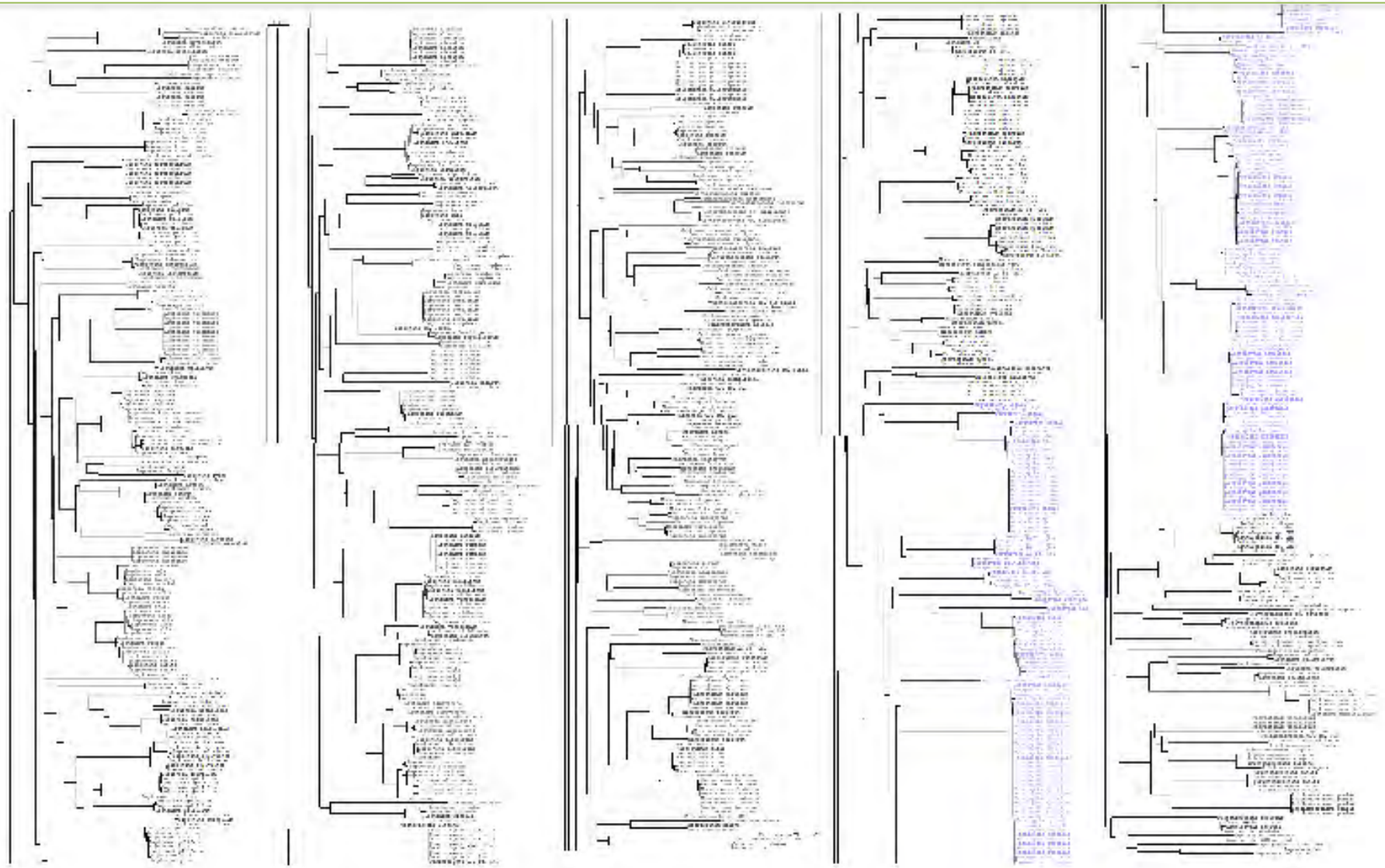
No? Why not?



Because we are confident with these 2 nodes

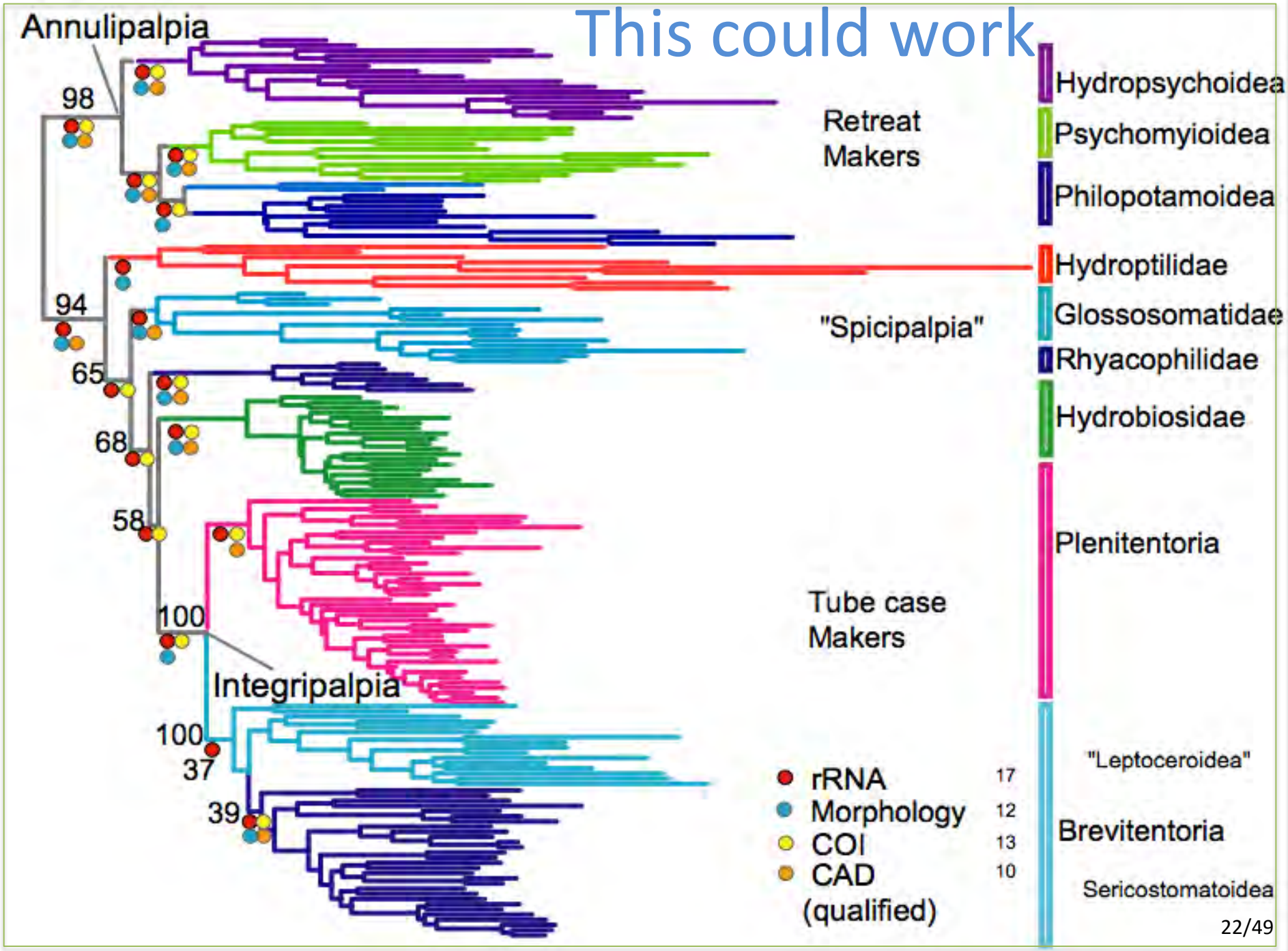
Using the same logic, we can estimate a phylogeny of Macronematinae, rooted with Arctopsychinae…all in Hydropsychidae



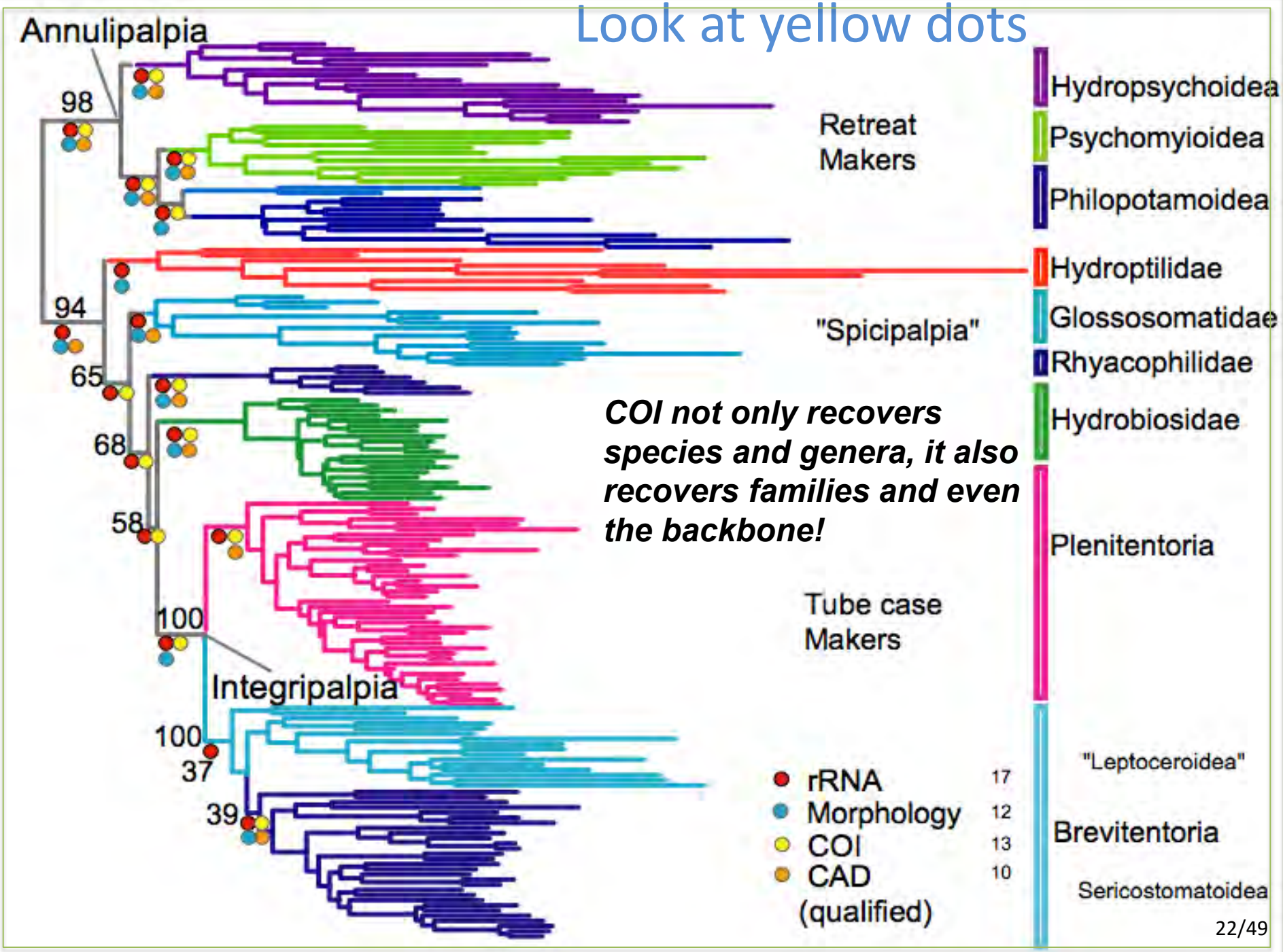


Here is a tree with 1,000 macronematines, and arctopsychines, one rooting the other

This could work

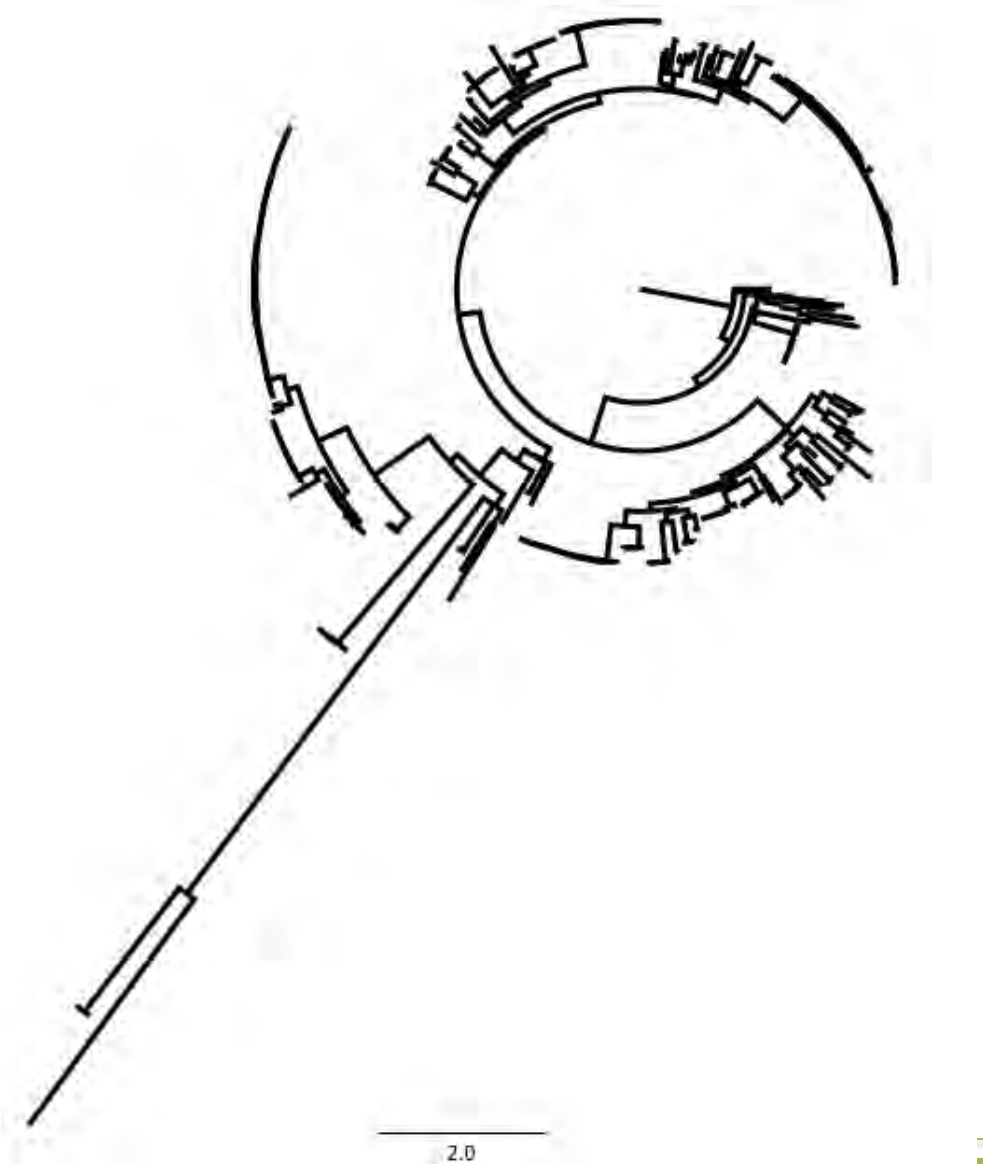


Look at yellow dots

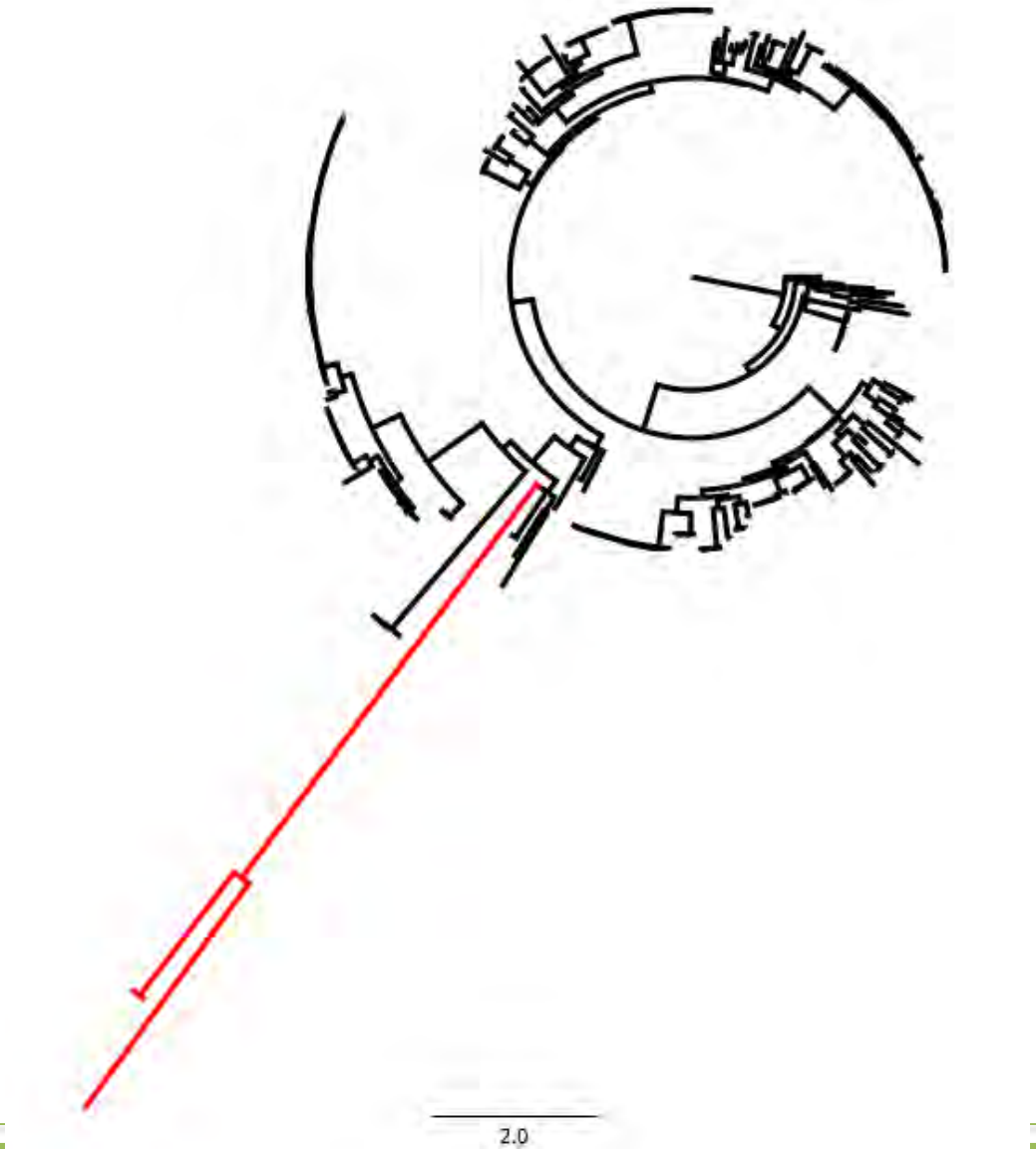


COI not only recovers species and genera, it also recovers families and even the backbone!

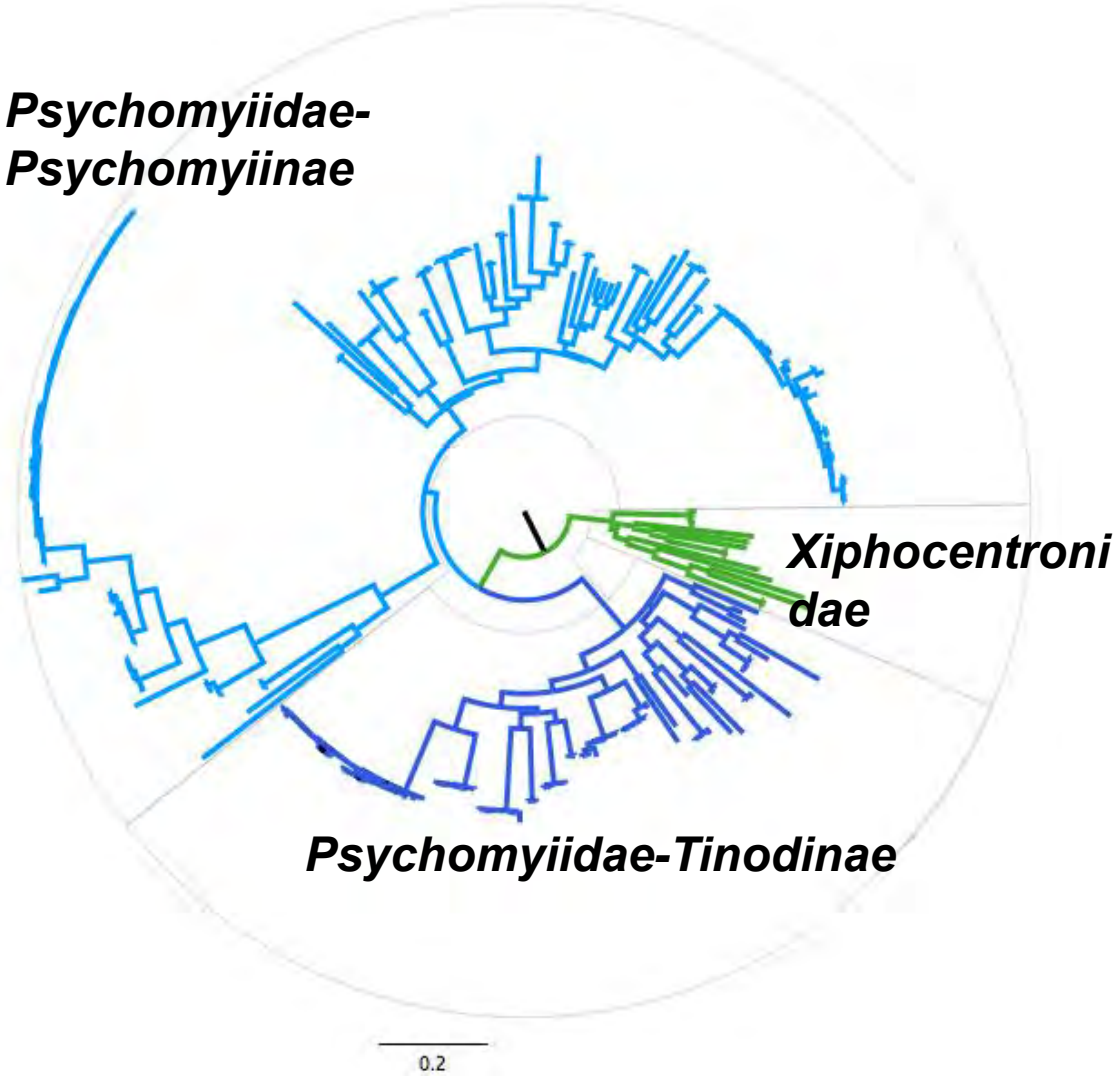
Trees with barcodes



Trees with barcodes



Trees with barcodes



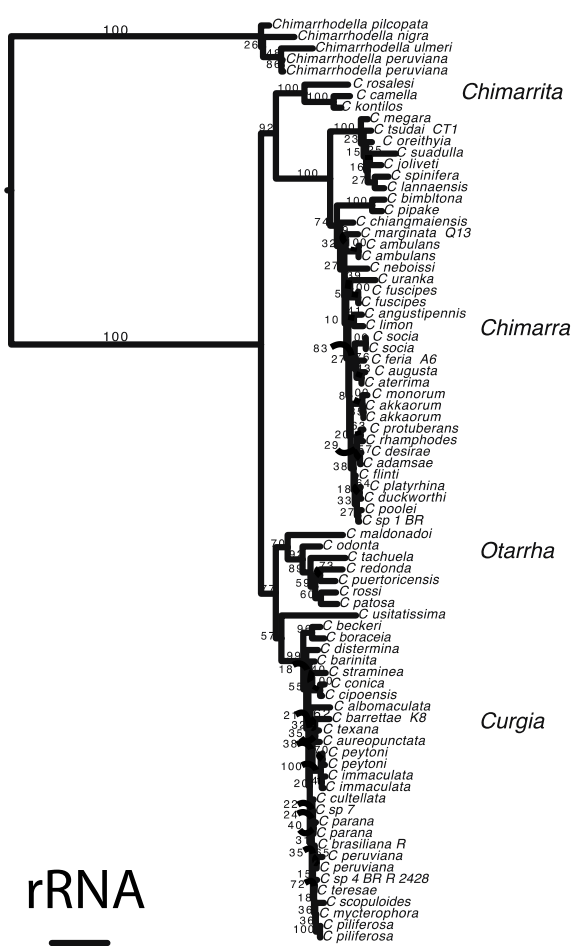
Genus ***Chimarra***:

The “little black ones”

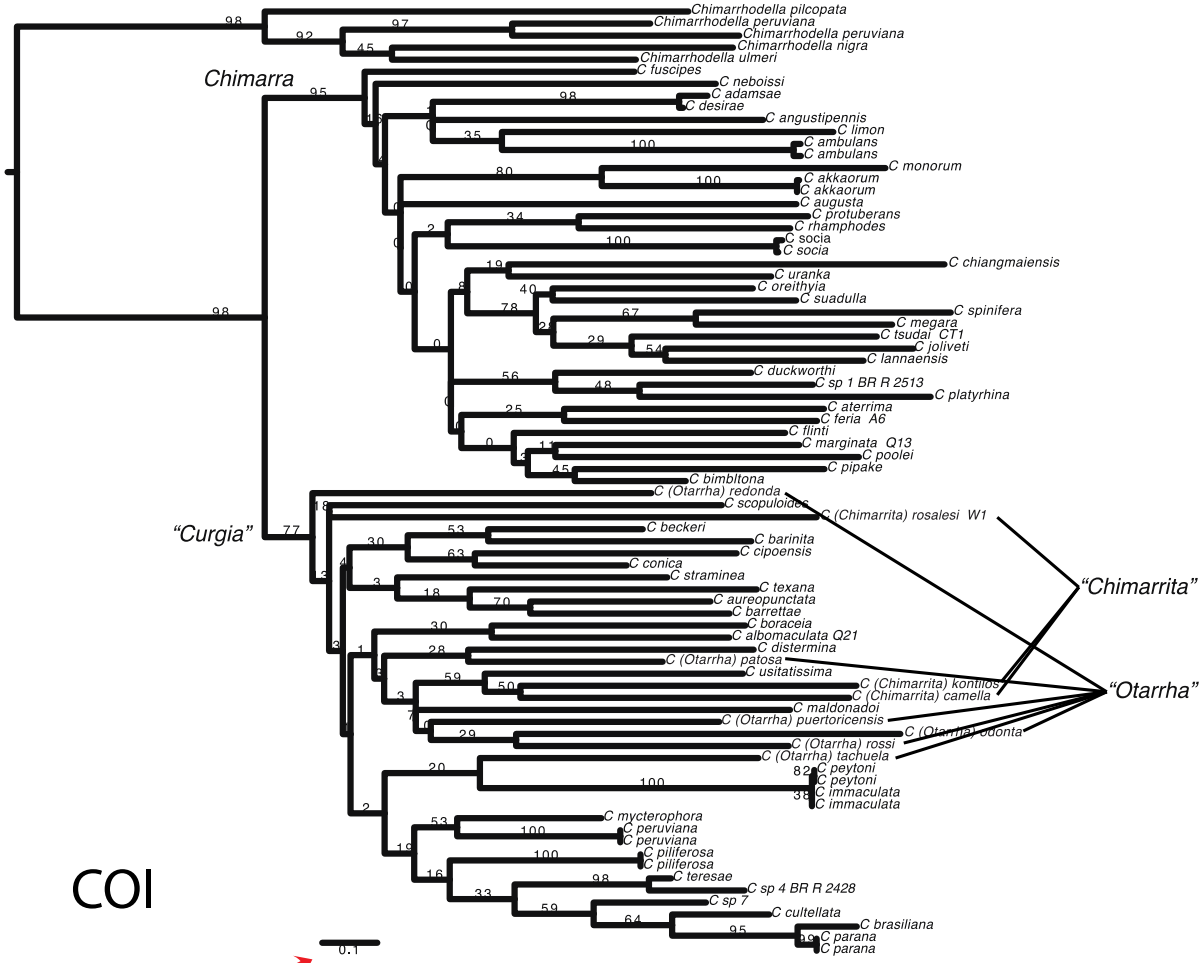
Largest in Trichoptera with ~750 species

BugGuide.net Photo by Ken

Which one will dominate a combined analysis?



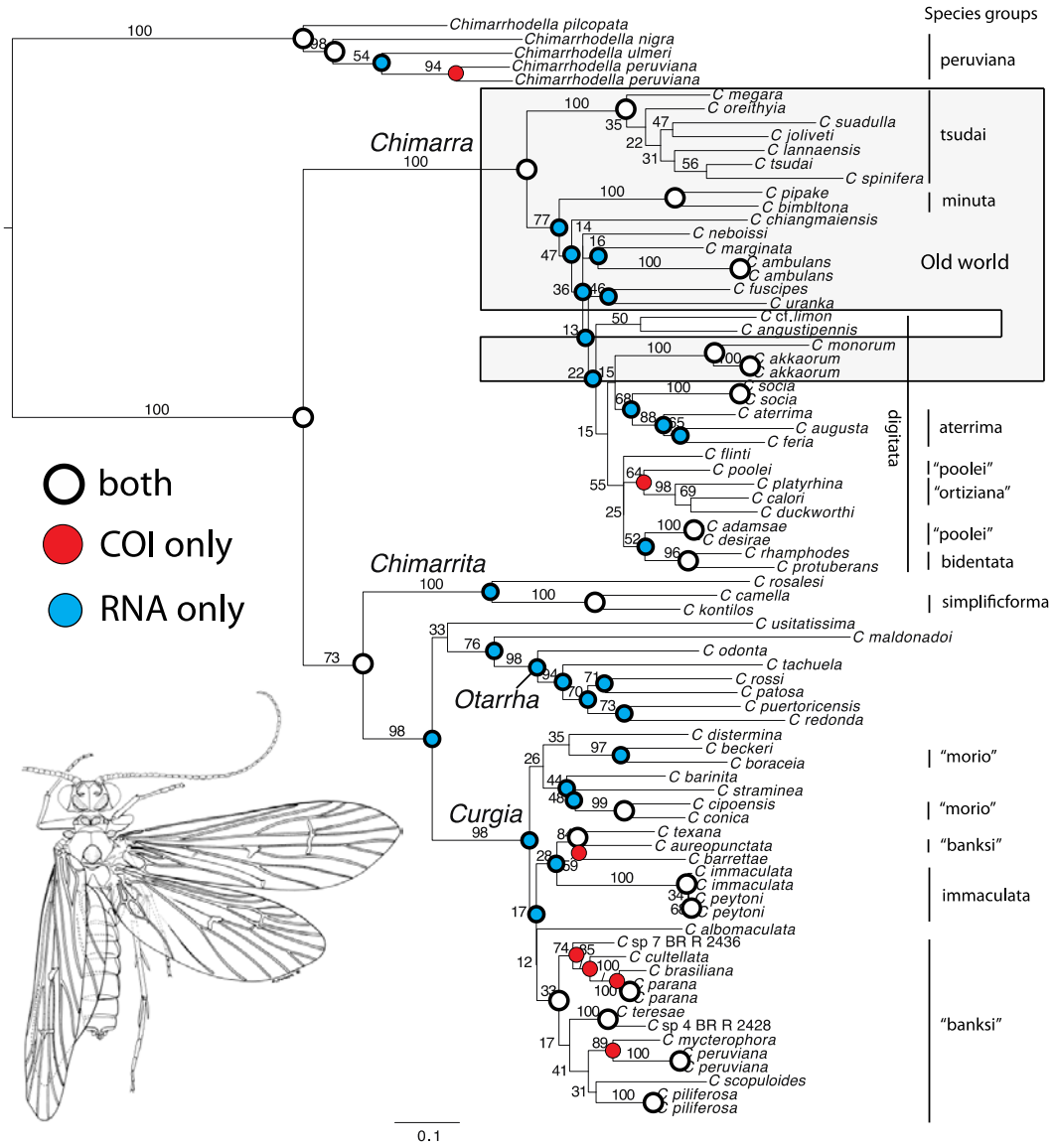
rRNA



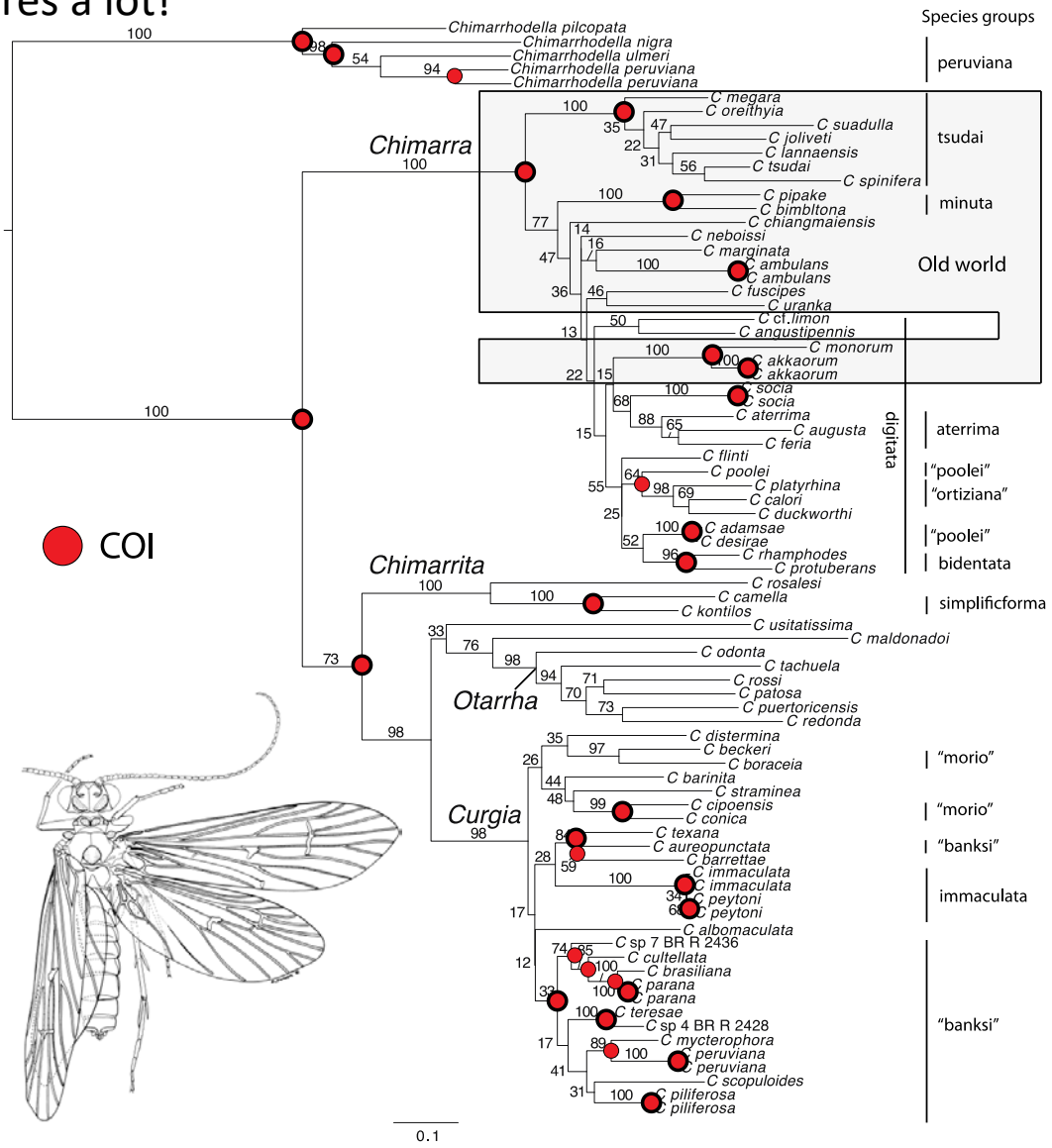
COI

Same Scale

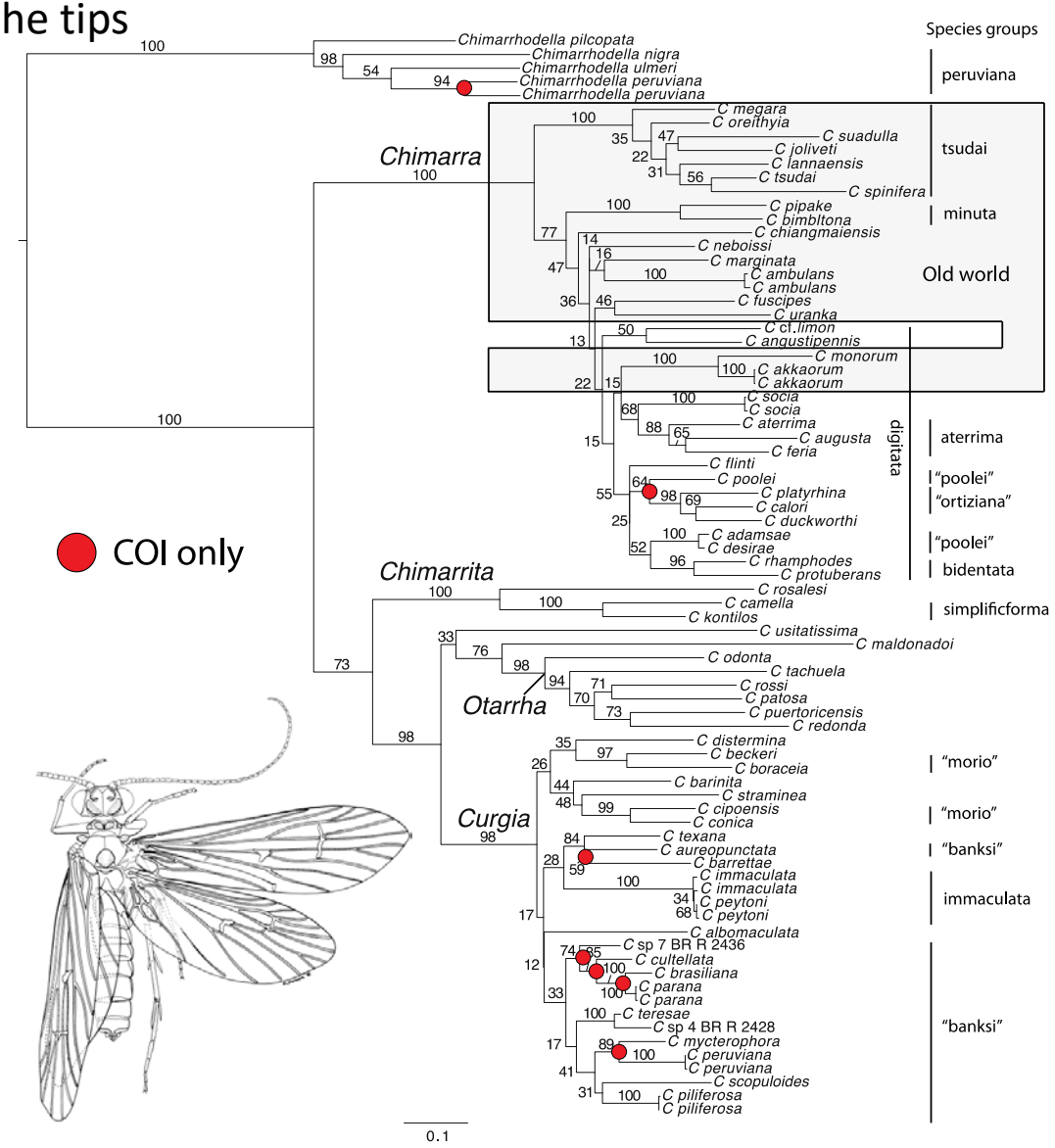
Combined: Judged by morphology, this is a great tree!



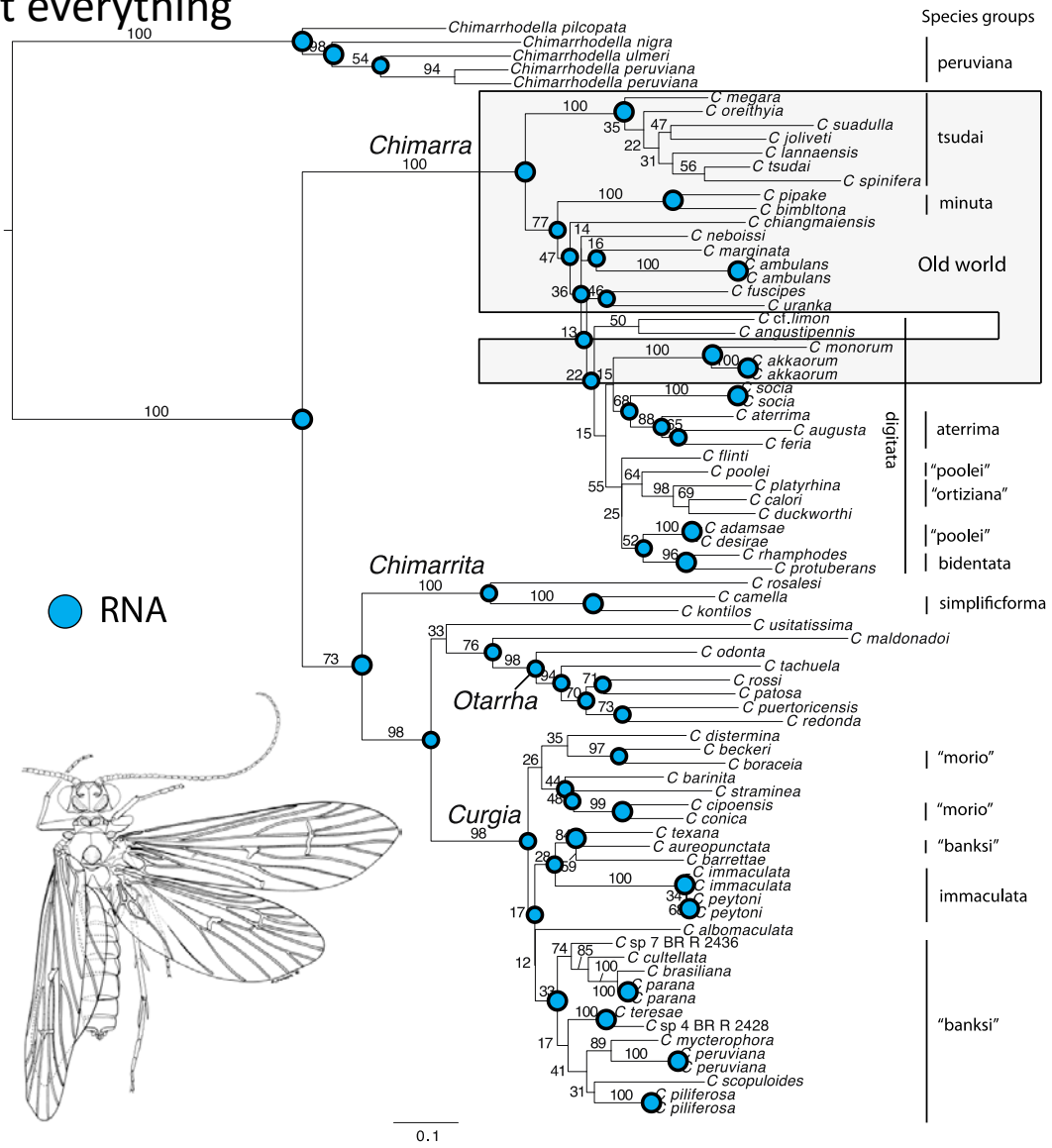
Barcode captures a lot!



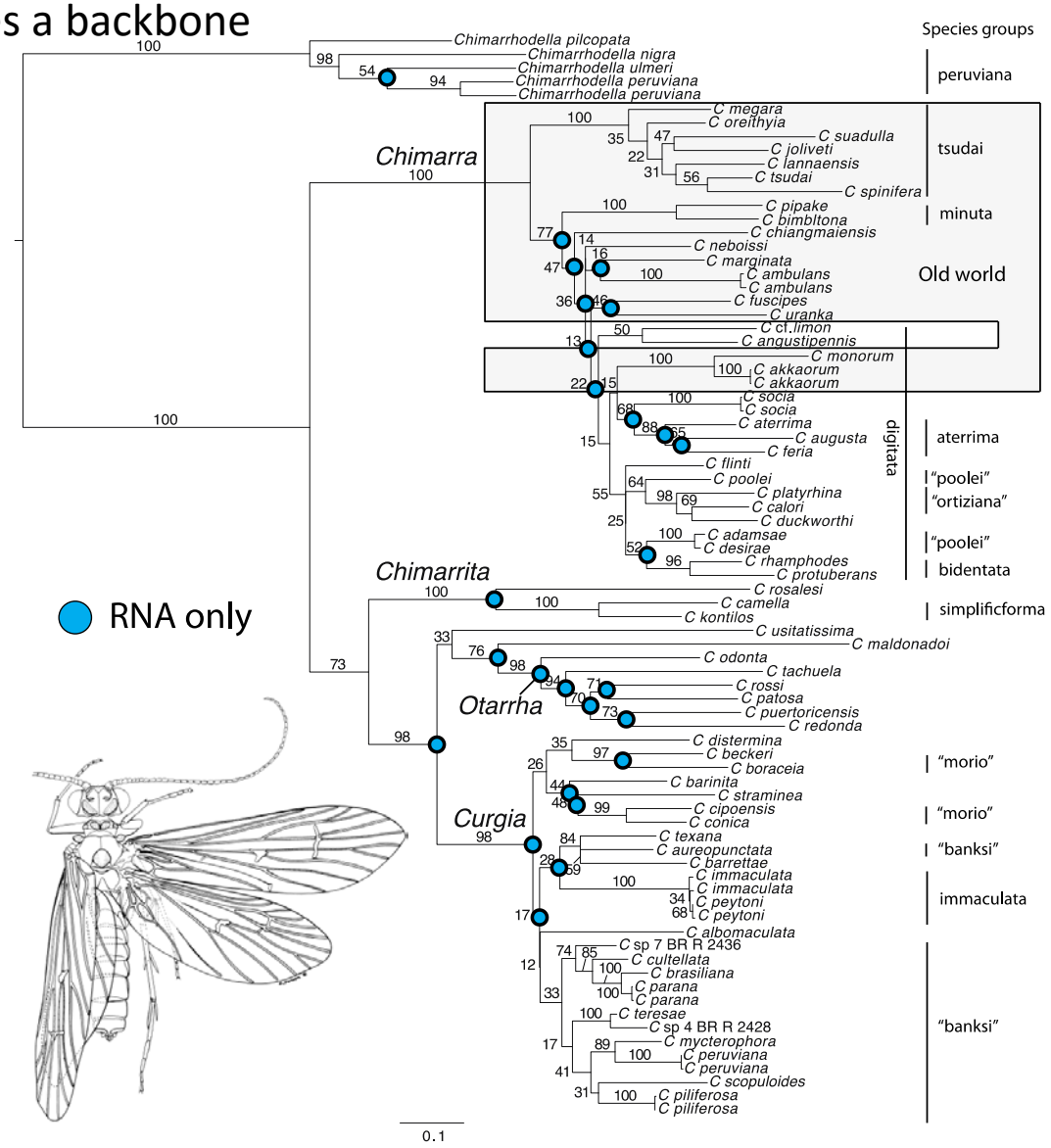
But favors the tips



RNA gets almost everything



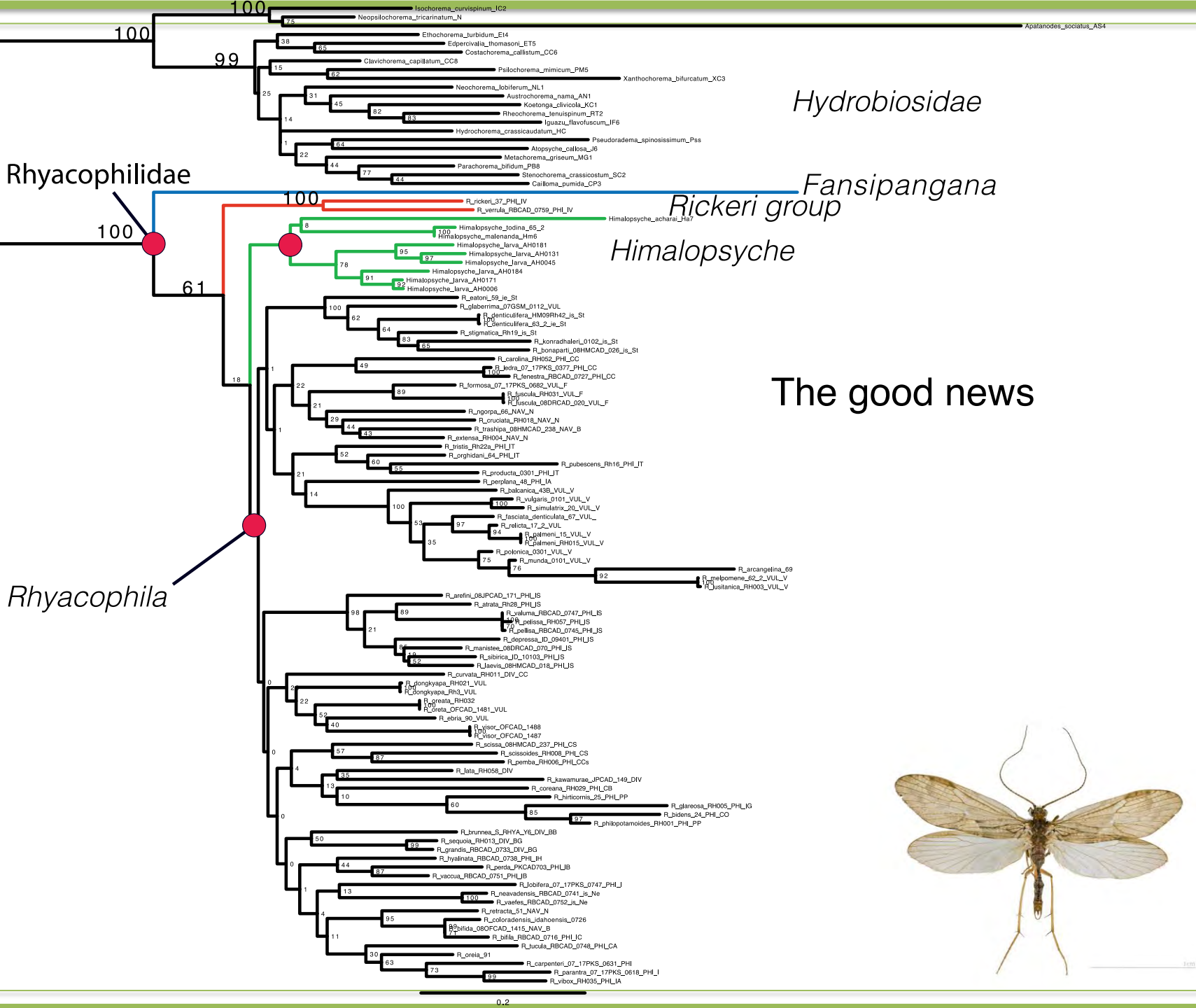
...and provides a backbone

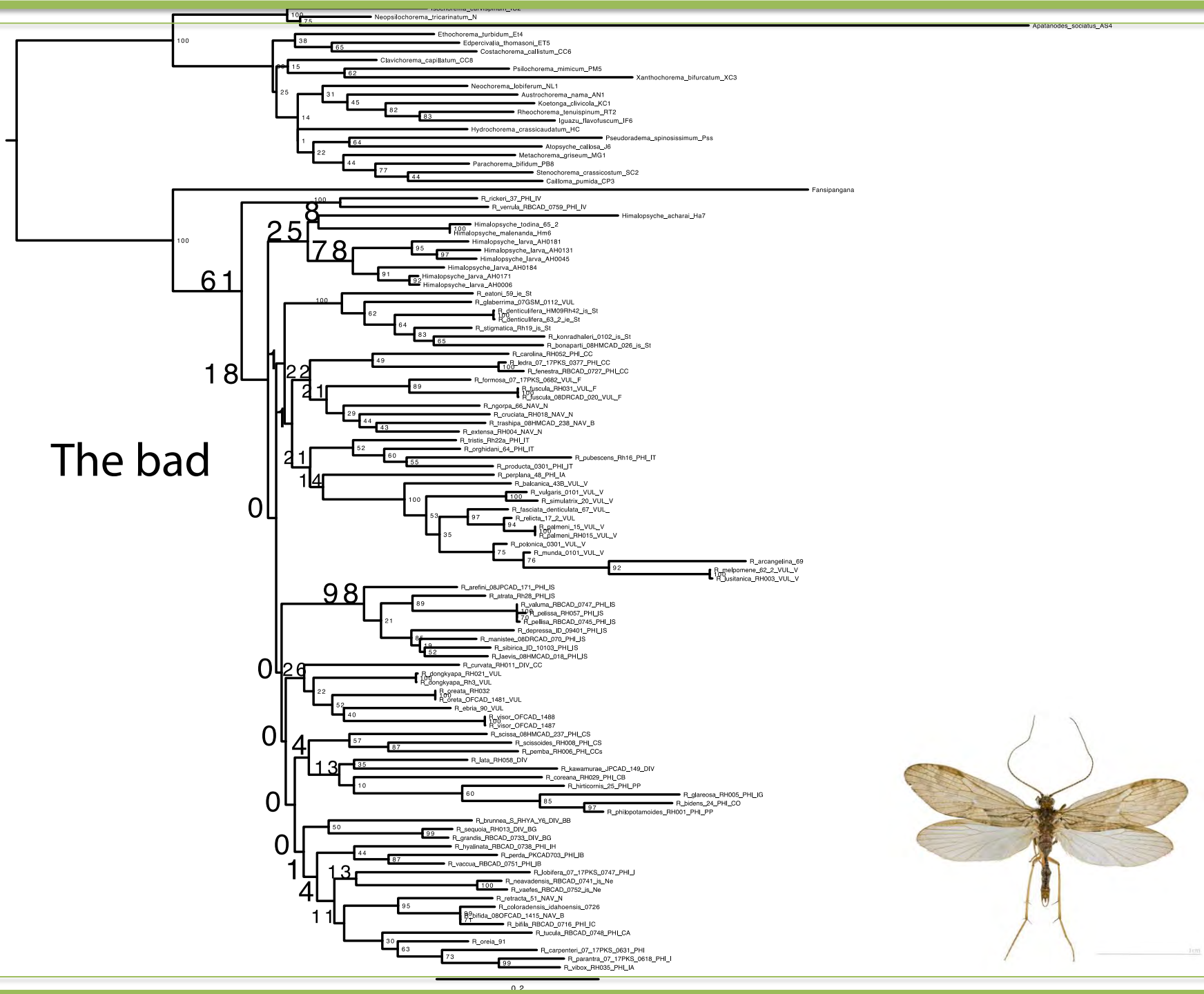


How about ***Rhyacophila***?

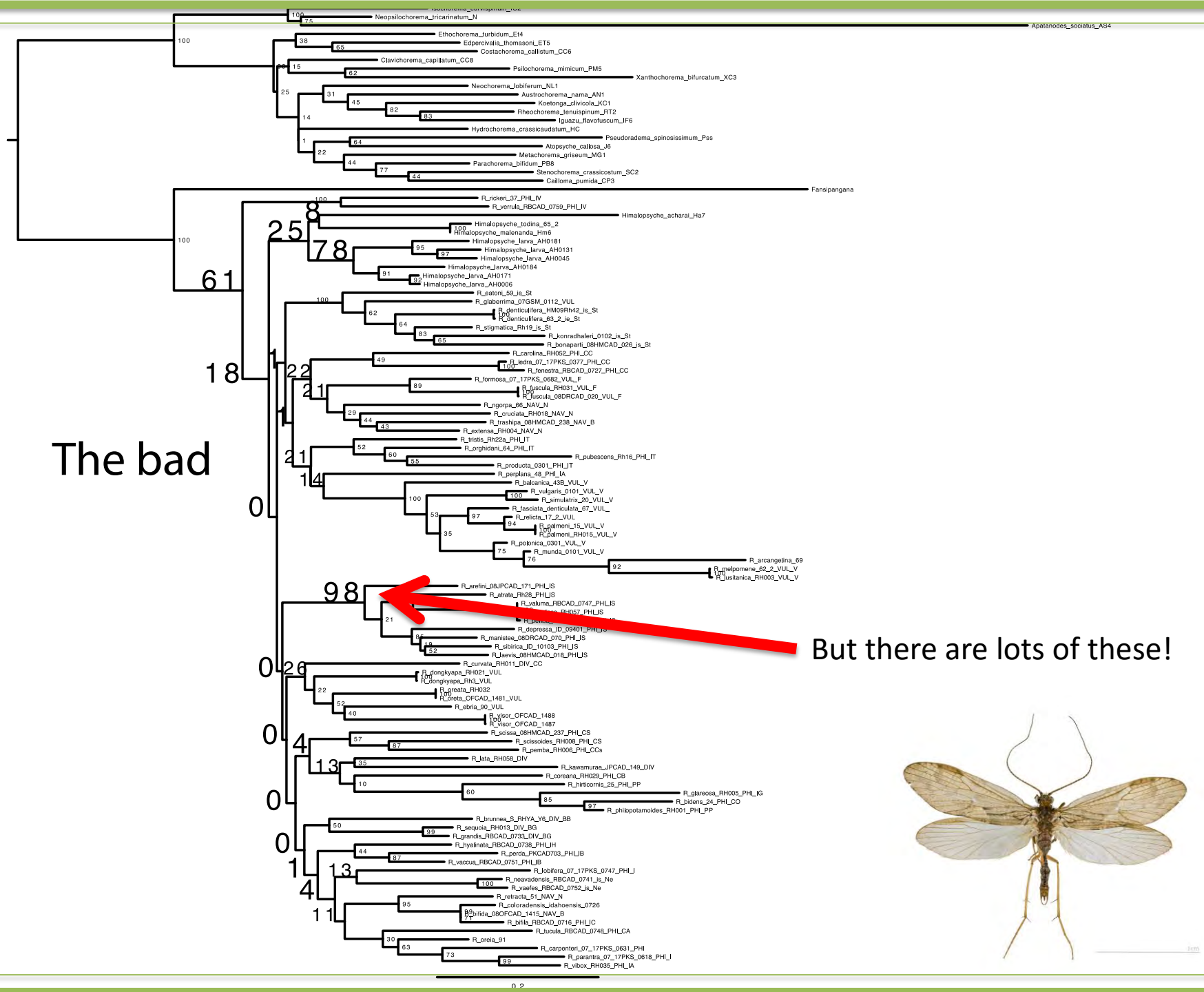
700 + species

Second only to *Chimarra*


Isochorema_curvispinum_IC2
Neopsilochorema_tricarinatum_N
Apatanodes_sociatus_AS4
Ethochorema_turbidum_Et4
Edpercivalia_thomasoni_ET5
Costachorema_callistum_CC6
Clavichorema_capillatum_CC8
Psilochorema_mimicum_PM5
Xanthochorema_bifurcatum_XC3
Neochorema_lobiferum_NL1
Austrochorema_nama_AN1
Koetonga_clivicola_KC1
Rheochorema_tenuispinum_RT2
Iguazu_flavofuscum_IF6
Hydrochorema_crassicaudatum_HC
Pseudoradema_spinosissimum_Pss
Atopsyche_callosa_J6
Metachorema_griseum_MG1
Parachorema_bifidum_PB8
Stenochorema_crassicostum_SC2
Cailloma_pumida_CP3
Hydrobiosidae
Rhyacophilidae
Fansipangana
Rickeri group
R_rickeri_37_PHI_IV
R_verrula_RBCAD_0759_PHI_IV
Himalopsyche_acharai_Ha7
Himalopsyche_todina_65_2
Himalopsyche_malenanda_Hm6
Himalopsyche_larva_AH0181
Himalopsyche_larva_AH0131
Himalopsyche_larva_AH0045
Himalopsyche_larva_AH0184
Himalopsyche_larva_AH0171
Himalopsyche_larva_AH0006
Himalopsyche
R_eatoni_59_ie_St
R_glaberrima_07GSM_0112_VUL
R_denticulifera_HM09Rh42_is_St
R_denticulifera_83_2_ie_St
R_stigmatica_Rh19_is_St
R_konradhaleri_0102_is_St
R_bonaparti_08HMCAD_026_is_St
R_carolina_RH052_PHI_CC
R_ledra_07_17PKS_0377_PHI_CC
R_fenestra_RBCAD_0727_PHI_CC
R_formosa_07_17PKS_0682_VUL_F
R_fuscula_RH031_VUL_F
R_fuscula_08DRCAD_020_VUL_F
R_ngorpa_66_NAV_N
R_cruciata_RH018_NAV_N
R_trashipa_08HMCAD_238_NAV_B
R_extensa_RH004_NAV_N
R_tristis_Rh22a_PHI_IT
R_orghidani_64_PHI_IT
R_pubescens_Rh16_PHI_IT
R_producta_0301_PHI_IT
R_perplana_48_PHI_IA
R_balcanica_43B_VUL_V
R_vulgaris_0101_VUL_V
R_simulatrix_20_VUL_V
R_fasciata_denticulata_67_VUL_
R_relicta_17_2_VUL
R_palmeni_15_VUL_V
R_palmeni_RH015_VUL_V
R_polonica_0301_VUL_V
R_munda_0101_VUL_V
R_arcangelina_69
R_melpomene_62_2_VUL_V
R_lusitanica_RH003_VUL_V
Rhyacophila
The good news
R_arefini_08JPCAD_171_PHI_IS
R_atrata_Rh28_PHI_IS
R_valuma_RBCAD_0747_PHI_IS
R_pelissa_RH057_PHI_IS
R_pellisa_RBCAD_0745_PHI_IS
R_depressa_ID_09401_PHI_IS
R_manistee_08DRCAD_070_PHI_IS
R_sibirica_JD_10103_PHI_IS
R_laevis_08HMCAD_018_PHI_IS
R_curvata_RH011_DIV_CC
R_dongkyapa_RH021_VUL
R_dongkyapa_Rh3_VUL
R_oreata_RH032
R_oreta_OFCAD_1481_VUL
R_ebria_90_VUL
R_visor_OFCAD_1488
R_visor_OFCAD_1487
R_scissa_08HMCAD_237_PHI_CS
R_scissoides_RH008_PHI_CS
R_pemba_RH006_PHI_CCs
R_lata_RH058_DIV
R_kawamurae_JPCAD_149_DIV
R_coreana_RH029_PHI_CB
R_hirticornis_25_PHI_PP
R_glareosa_RH005_PHI_IG
R_bidens_24_PHI_CO
R_philopotamoides_RH001_PHI_PP
R_brunnea_S_RHYA_Y6_DIV_8B
R_sequoia_RH013_DIV_BG
R_grandis_RBCAD_0733_DIV_BG
R_hyalinata_RBCAD_0738_PHI_IH
R_perda_PKCAD703_PHI_IB
R_vaccua_RBCAD_0751_PHI_IB
R_lobifera_07_17PKS_0747_PHI_I
R_neavadensis_RBCAD_0741_is_Ne
R_vaefes_RBCAD_0752_is_Ne
R_retracta_51_NAV_N
R_coloradensis_idahoensis_0726
R_bifida_08OFCAD_1415_NAV_B
R_bifila_RBCAD_0716_PHI_IC
R_tucula_RBCAD_0748_PHI_CA
R_oreia_91
R_carpenteri_07_17PKS_0631_PHI
R_parantra_07_17PKS_0618_PHI_I
R_vibox_RH035_PHI_IA
0.2

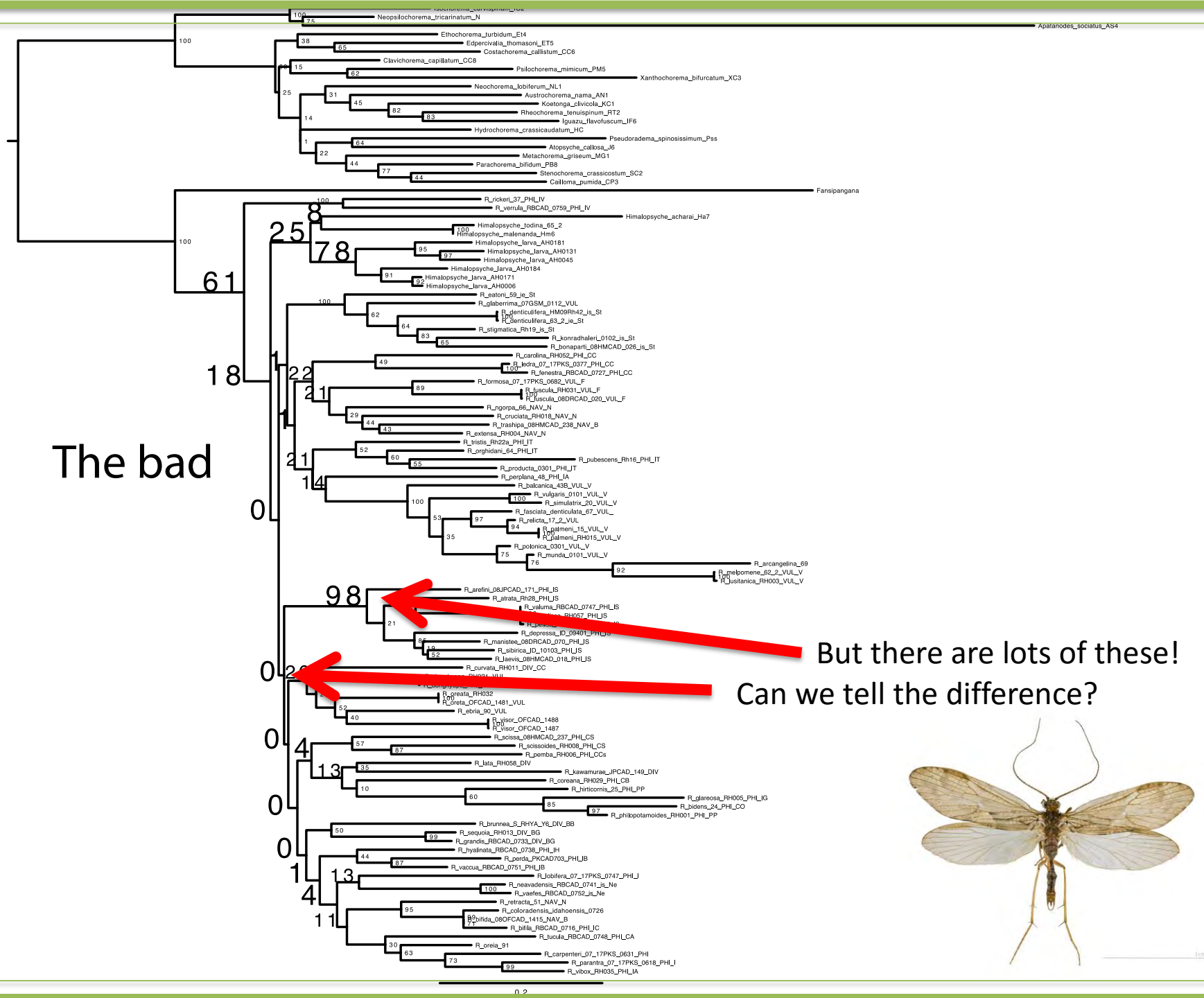

Neopsilochorema_tricarinatum_N
Apatanodes_sociatus_AS4
Ethochorema_turbidum_Et4
Edpercivalia_thomasoni_ET5
Costachorema_callistum_CC6
Clavichorema_capillatum_CC8
Psilochorema_mimicum_PM5
Xanthochorema_bifurcatum_XC3
Neochorema_lobiferum_NL1
Austrochorema_nama_AN1
Koetonga_clivicola_KC1
Rheochorema_tenuispinum_RT2
Iguazu_flavofuscum_IF6
Hydrochorema_crassicaudatum_HC
Pseudoradema_spinosissimum_Pss
Atopsyche_callosa_J6
Metachorema_griseum_MG1
Parachorema_bifidum_PB8
Stenochorema_crassicostum_SC2
Cailloma_pumida_CP3
Fansipangana
R_rickeri_37_PHI_IV
R_verrula_RBCAD_0759_PHI_IV
Himalopsyche_acharai_Ha7
Himalopsyche_todina_65_2
Himalopsyche_malenanda_Hm6
Himalopsyche_larva_AH0181
Himalopsyche_larva_AH0131
Himalopsyche_larva_AH0045
Himalopsyche_larva_AH0184
Himalopsyche_larva_AH0171
Himalopsyche_larva_AH0006
R_eatoni_59_ie_St
R_glaberrima_07GSM_0112_VUL
R_denticulifera_HM09Rh42_is_St
R_denticulifera_83_2_ie_St
R_stigmatica_Rh19_is_St
R_konradhaleri_0102_is_St
R_bonaparti_08HMCAD_026_is_St
R_carolina_RH052_PHI_CC
R_ledra_07_17PKS_0377_PHI_CC
R_fenestra_RBCAD_0727_PHI_CC
R_formosa_07_17PKS_0682_VUL_F
R_fuscula_RH031_VUL_F
R_fuscula_08DRCAD_020_VUL_F
R_ngorpa_66_NAV_N
R_cruciata_RH018_NAV_N
R_trashipa_08HMCAD_238_NAV_B
R_extensa_RH004_NAV_N
R_tristis_Rh22a_PHI_IT
R_orghidani_64_PHI_IT
R_pubescens_Rh16_PHI_IT
R_producta_0301_PHI_IT
R_perplana_48_PHI_IA
R_balcanica_43B_VUL_V
R_vulgaris_0101_VUL_V
R_simulatrix_20_VUL_V
R_fasciata_denticulata_67_VUL_
R_relicta_17_2_VUL
R_palmeni_15_VUL_V
R_palmeni_RH015_VUL_V
R_polonica_0301_VUL_V
R_munda_0101_VUL_V
R_arcangelina_69
R_melpomene_62_2_VUL_V
R_lusitanica_RH003_VUL_V
R_arefini_08JPCAD_171_PHI_IS
R_atrata_Rh28_PHI_IS
R_valuma_RBCAD_0747_PHI_IS
R_pelissa_RH057_PHI_IS
R_pellisa_RBCAD_0745_PHI_IS
R_depressa_ID_09401_PHI_IS
R_manistee_08DRCAD_070_PHI_IS
R_sibirica_JD_10103_PHI_IS
R_laevis_08HMCAD_018_PHI_IS
R_curvata_RH011_DIV_CC
R_dongkyapa_RH021_VUL
R_dongkyapa_Rh3_VUL
R_oreata_RH032
R_oreta_OFCAD_1481_VUL
R_ebria_90_VUL
R_visor_OFCAD_1488
R_visor_OFCAD_1487
R_scissa_08HMCAD_237_PHI_CS
R_scissoides_RH008_PHI_CS
R_pemba_RH006_PHI_CCs
R_lata_RH058_DIV
R_kawamurae_JPCAD_149_DIV
R_coreana_RH029_PHI_CB
R_hirticornis_25_PHI_PP
R_glareosa_RH005_PHI_IG
R_bidens_24_PHI_CO
R_philopotamoides_RH001_PHI_PP
R_brunnea_S_RHYA_Y6_DIV_BB
R_sequoia_RH013_DIV_BG
R_grandis_RBCAD_0733_DIV_BG
R_hyalinata_RBCAD_0738_PHI_IH
R_perda_PKCAD703_PHI_IB
R_vaccua_RBCAD_0751_PHI_IB
R_lobifera_07_17PKS_0747_PHI_I
R_neavadensis_RBCAD_0741_is_Ne
R_vaefes_RBCAD_0752_is_Ne
R_retracta_51_NAV_N
R_coloradensis_idahoensis_0726
R_bifida_08OFCAD_1415_NAV_B
R_bifila_RBCAD_0716_PHI_IC
R_tucula_RBCAD_0748_PHI_CA
R_oreia_91
R_carpenteri_07_17PKS_0631_PHI
R_parantra_07_17PKS_0618_PHI_I
R_vibox_RH035_PHI_IA
0.2

The bad



The bad

But there are lots of these!



The bad

But there are lots of these!

Can we tell the difference?

The bad

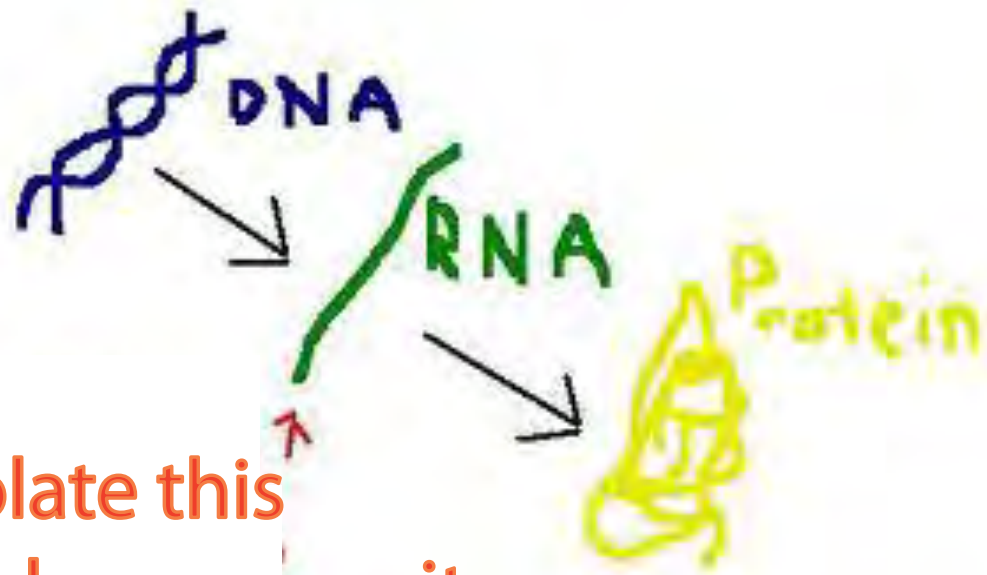
If so, then this is useful

(…*PLEASE* start looking at support)

But there are lots of these!

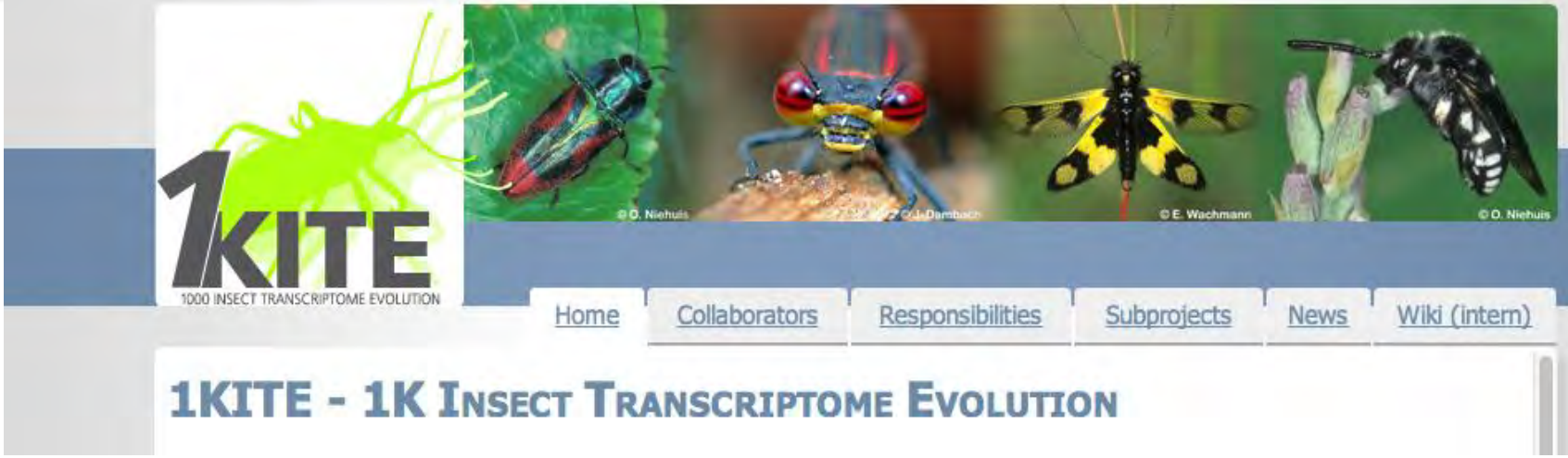
Can we tell the difference?

On to transcriptomes

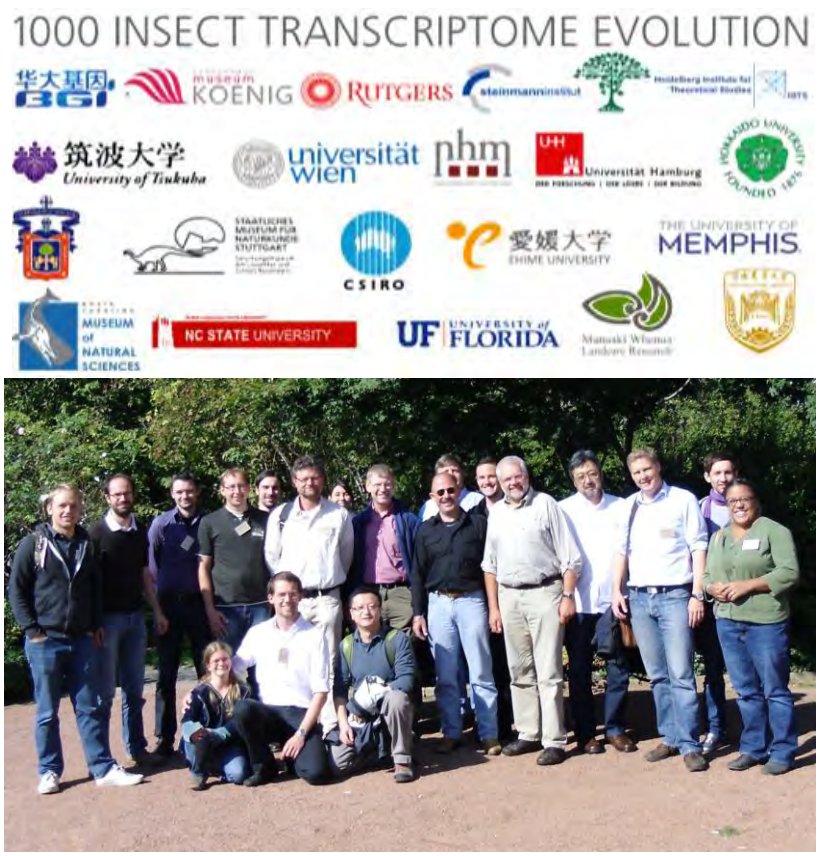


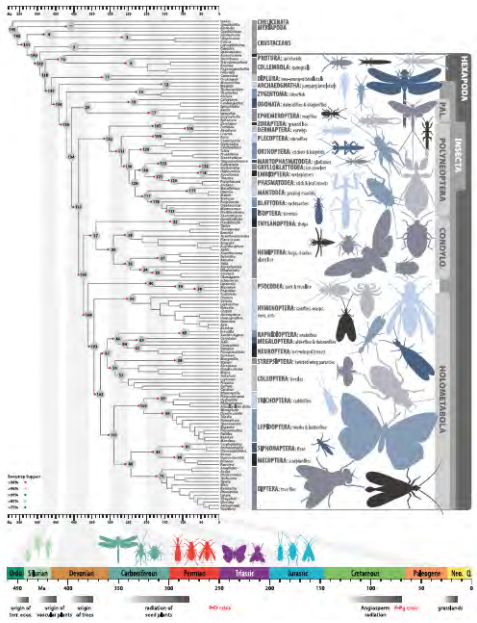
Isolate this
and sequence it

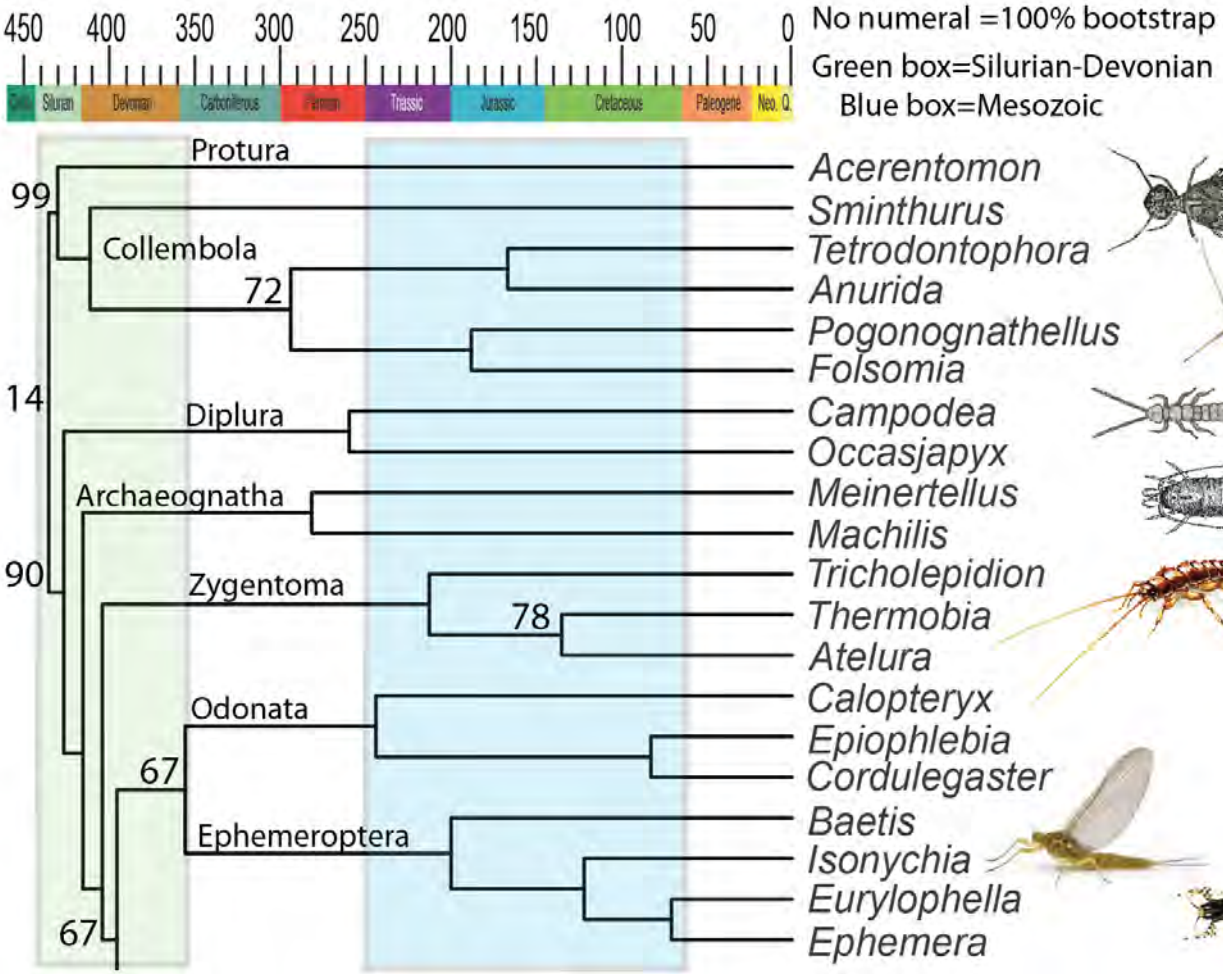
Barcode identification of Transcriptomes

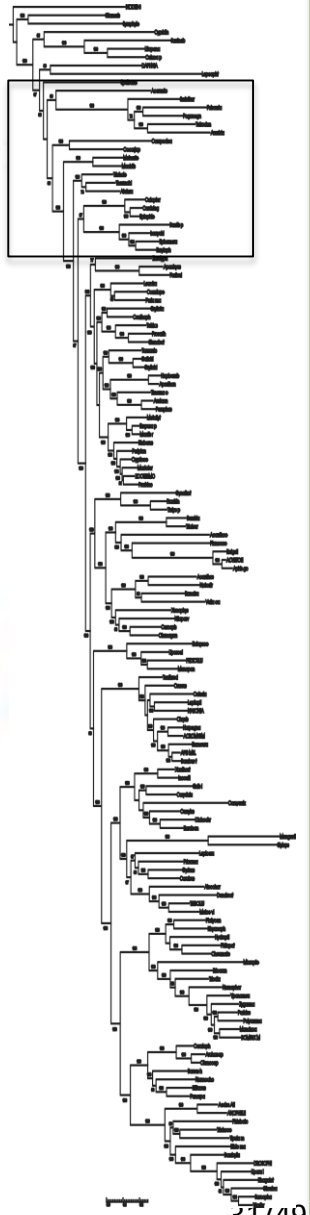

1KITE
1000 INSECT TRANSCRIPTOME EVOLUTION
© O. Niehuis
© J. Dambach
© E. Wachmann
© O. Niehuis
Home
Collaborators
Responsibilities
Subprojects
News
Wiki (intern)
1KITE - 1K INSECT TRANSCRIPTOME EVOLUTION

All aTwitter over an Internet study p. xxx
The extragalactic background's uneven glow pp. xxx & xxx
A cellular target for human norovirus pp. xxx & xxx
Science
$10
7 NOVEMBER 2014
sciencemag.org
AAAS
Insect phylogeny resolved
Molecular insights into insect origins and evolution
p. xxx


1000 INSECT TRANSCRIPTOME EVOLUTION
BGI
museum KOENIG
RUTGERS
Heidelberg Institute for Theoretical Studies
筑波大学
University of Tsukuba
universität wien
phm
Universität Hamburg
HOKKAIDO UNIVERSITY
CSIRO
愛媛大学
EHIME UNIVERSITY
THE UNIVERSITY OF MEMPHIS
MUSEUM of NATURAL SCIENCES
NC STATE UNIVERSITY
UF UNIVERSITY of FLORIDA


HEXAPODA
INSECTA
PALAEOPTERA
POLYNEOPTERA
CONDYLO.
HOLOMETABOLA





"No" to Nonocculata

Tricholepidion with other Zygentoma

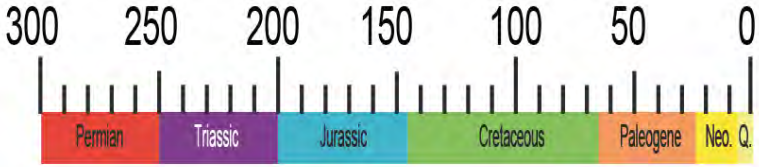
Palaeoptera, weakly supported

www.renderosity
.com

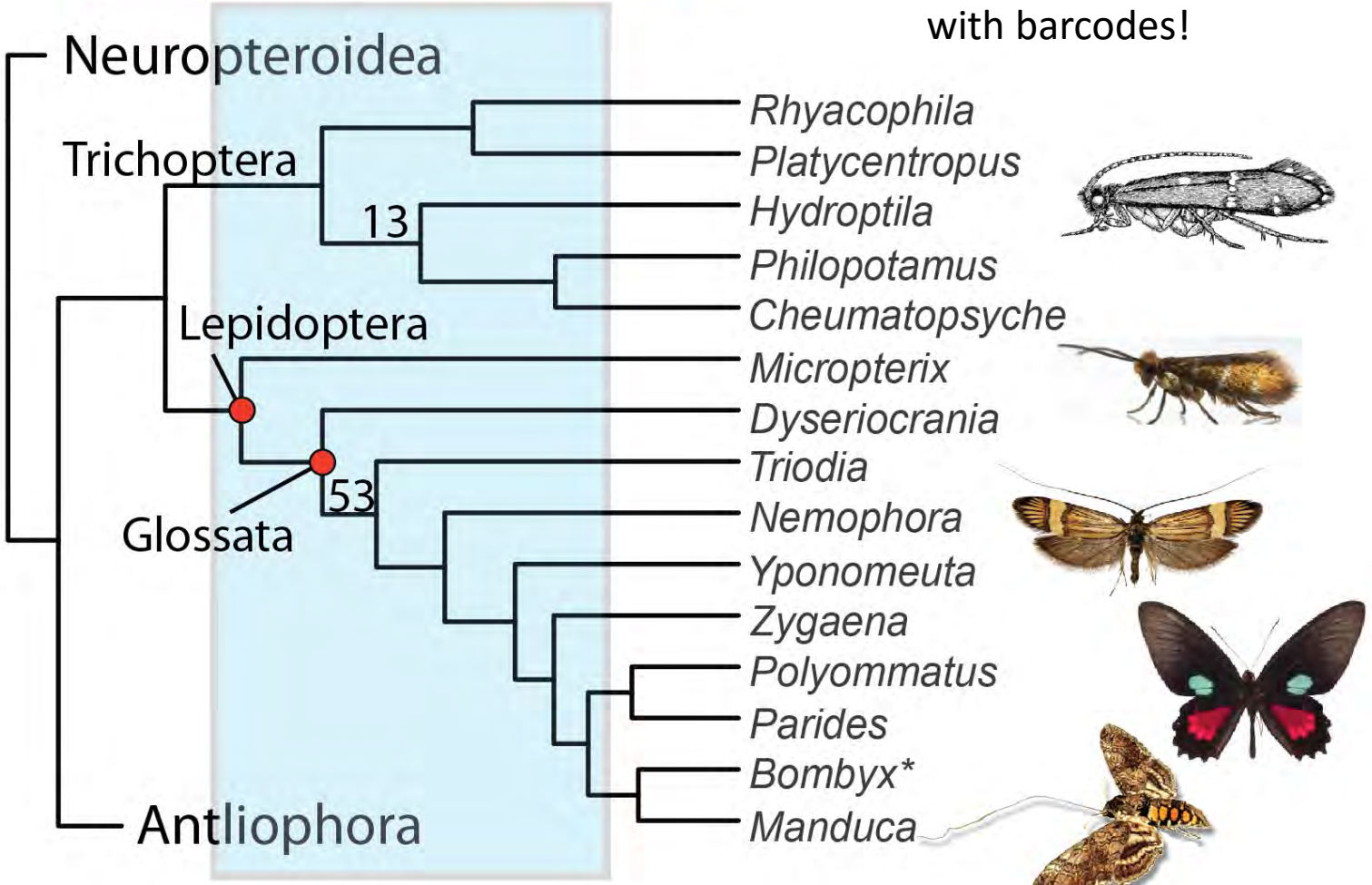
There ***were*** insects flying around the first forests!

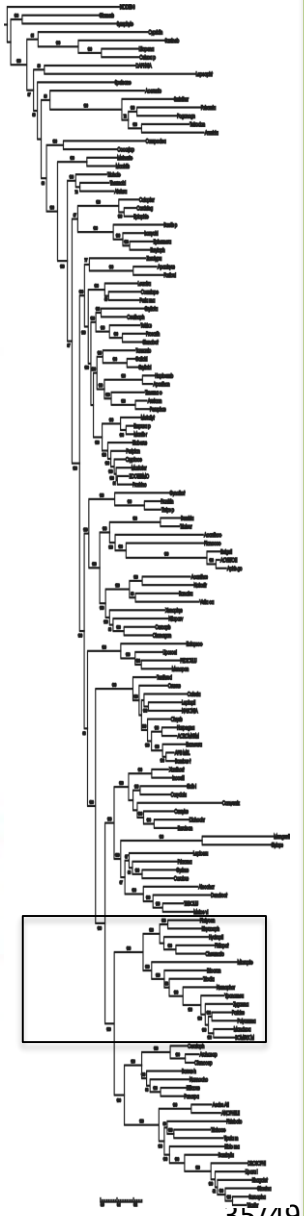
www.renderosity
.com

www.renderosity
.com



All Trichoptera were identified with barcodes!





No surprises here

Our first publication used less than 1% of our data.

There is much more to come, with "big data" setting the backbone for barcode phylogenies.

Hybrid Capture Techniques

Syst. Biol. 61(5):727–744, 2012
© The Author(s) 2012. Published by Oxford University Press, on behalf of the Society of Systematic Biologists. All rights reserved.
For Permissions, please email: journals.permissions@oup.com
DOI:10.1093/sysbio/sys049
Advance Access publication on May 17, 2012

Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics

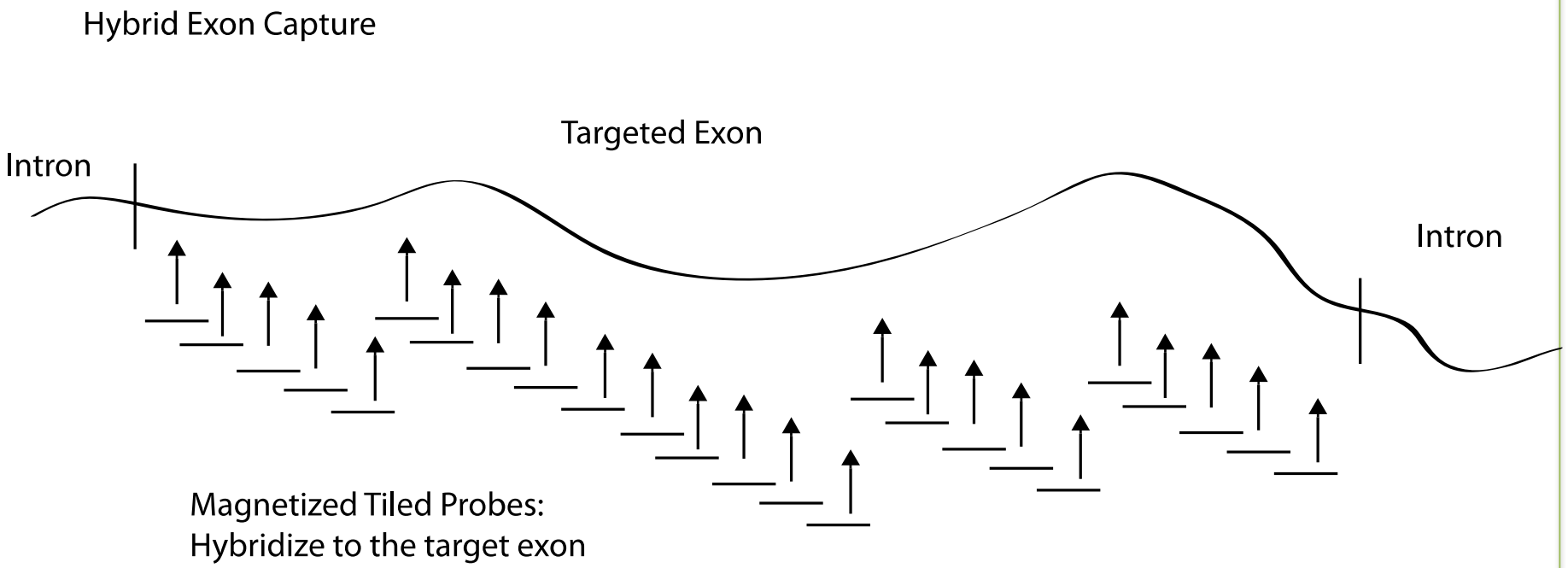
ALAN R. LEMMON[1,*], SANDRA A. EMME[2], AND EMILY MORIARTY LEMMON[2]

[1]*Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306-4102, USA; and* [2]*Department of Biological Science, Florida State University, 319 Stadium Drive, PO Box 3064295, Tallahassee, FL, 32306-4295, USA;*
**Correspondence to be sent to: Department of Scientific Computing, Florida State University, Dirac Science Library, Tallahassee, FL 32306-4102; E-mail: alemmon@fsu.edu.*

Received 1 November 2011; reviews returned 19 January 2012; accepted 7 May 2012
Associate Editor: Bryan Carstens

Abstract.—The field of phylogenetics is on the cusp of a major revolution, enabled by new methods of data collection that leverage both genomic resources and recent advances in DNA sequencing. Previous phylogenetic work has required labor-intensive marker development coupled with single-locus polymerase chain reaction and DNA sequencing on clade-by-clade and locus-by-locus basis. Here, we present a new, cost-efficient, and rapid approach to obtaining data from hundreds of loci for potentially hundreds of individuals for deep and shallow phylogenetic studies. Specifically, we designed probes for target enrichment of >500 loci in highly conserved anchor regions of vertebrate genomes (flanked by less conserved regions) from five model species and tested enrichment efficiency in nonmodel species up to 508 million years divergent from the nearest model. We found that hybrid enrichment using conserved probes (anchored enrichment) can recover a large number of unlinked loci that are useful at a diversity of phylogenetic timescales. This new approach has the potential not only to expedite resolution of deep-scale portions of the Tree of Life but also to greatly accelerate resolution of the large number of shallow clades that remain unresolved. The combination of low cost (~1% of the cost of traditional Sanger sequencing and ~3.5% of the cost of high-throughput amplicon sequencing for projects on the scale of 500 loci × 100 individuals) and rapid data collection (~2 weeks of laboratory time) are expected to make this approach tractable even for researchers working on systems with limited or nonexistent genomic resources. [Anchor regions, anchored enrichment, anchored phylogenomics, highly conserved regions, hybrid enrichment, phylogenetics, phylogeography, sequence capture, ultraconserved elements.]

Probes are tiled across entire exons

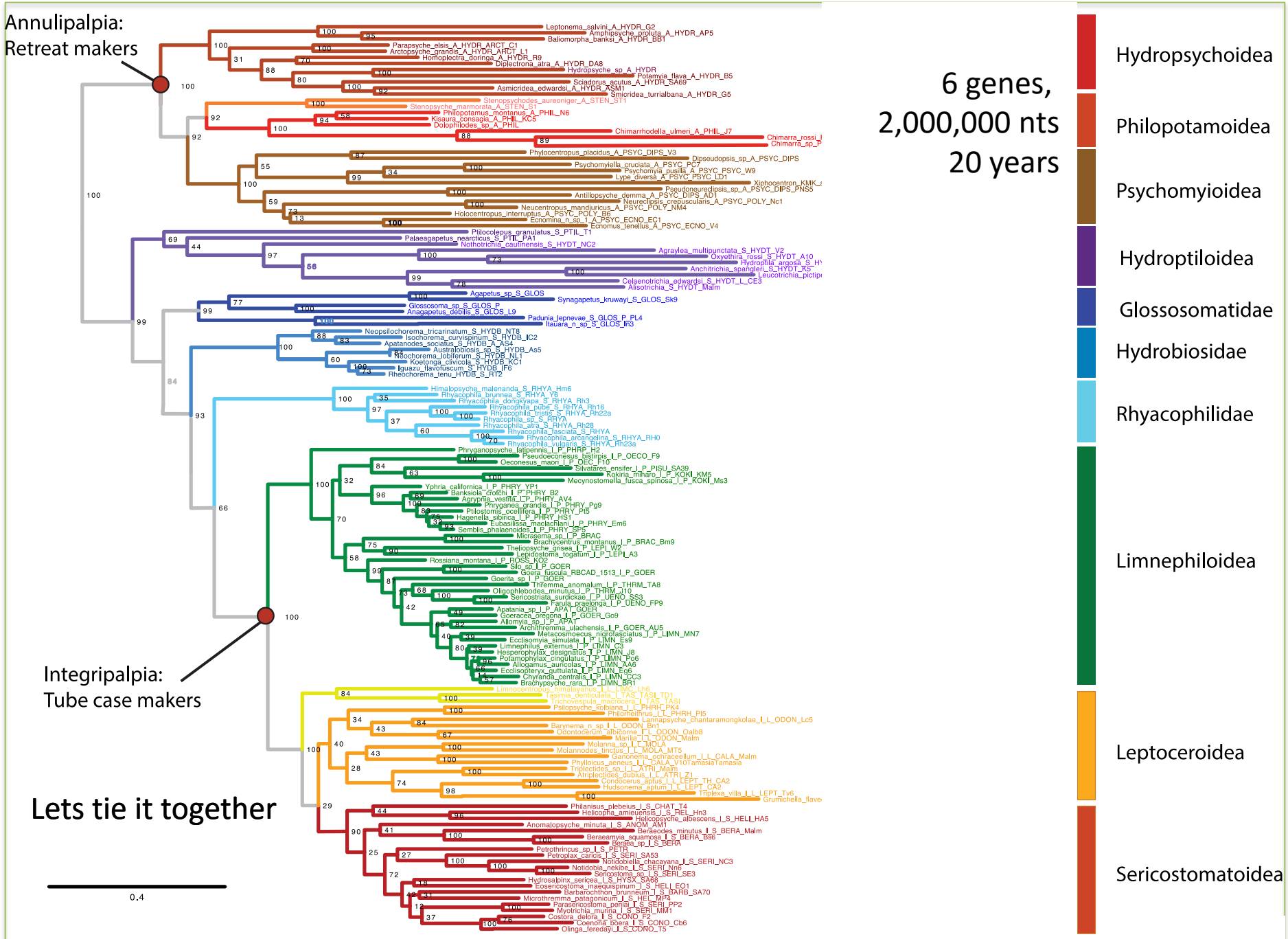


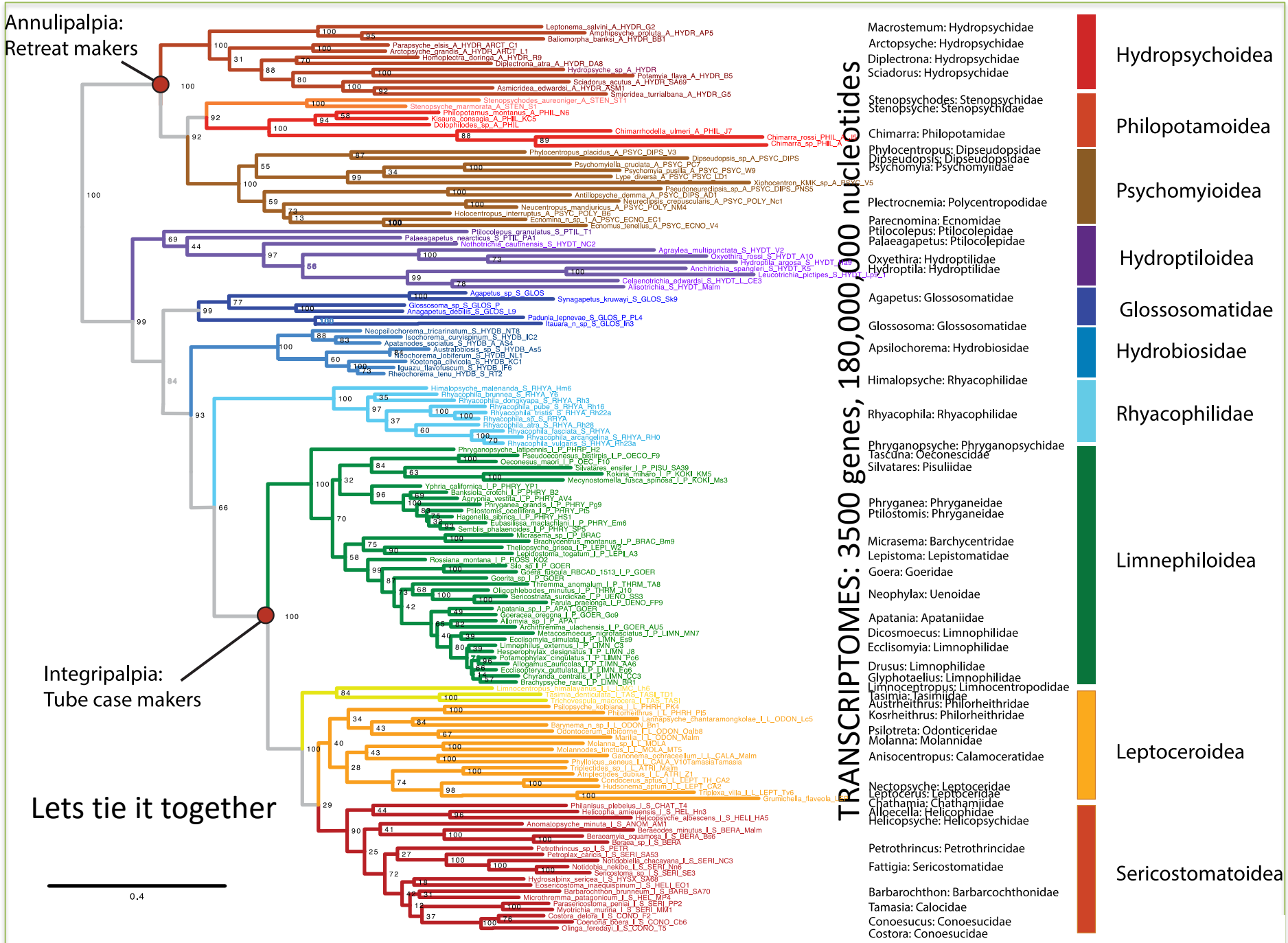
Imagine you target 900 exons this way.
Sequence them all in one run…

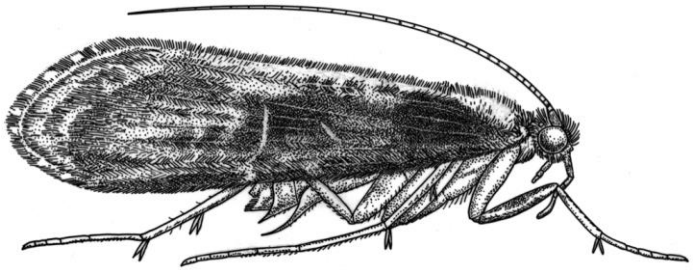
My beloved freezer
(good ol' #5; we've been through a lot together)

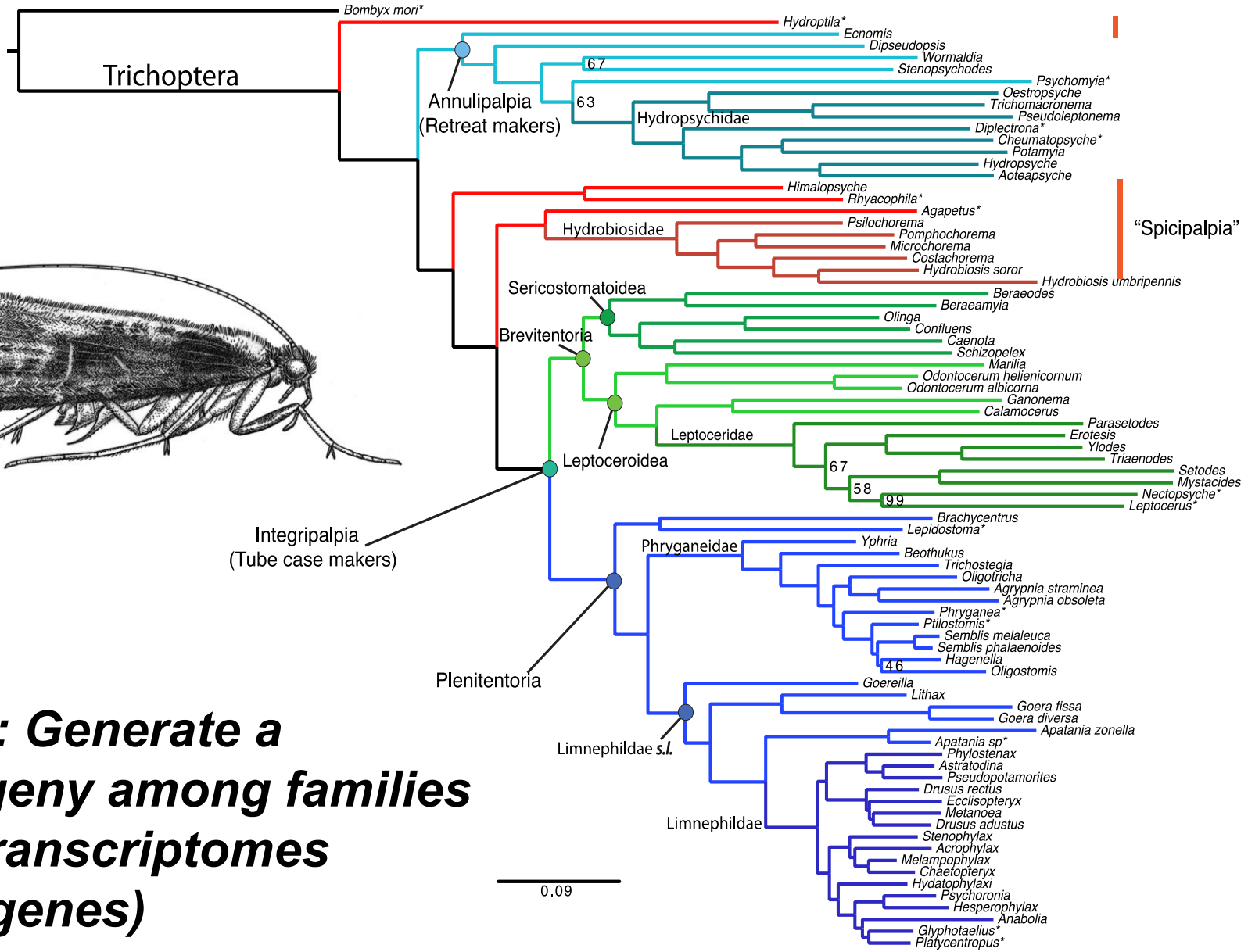
It took 20 years, and 200 person years to collect this...
It is useless for transcriptome work

Perfect for exome capture and barcoding!

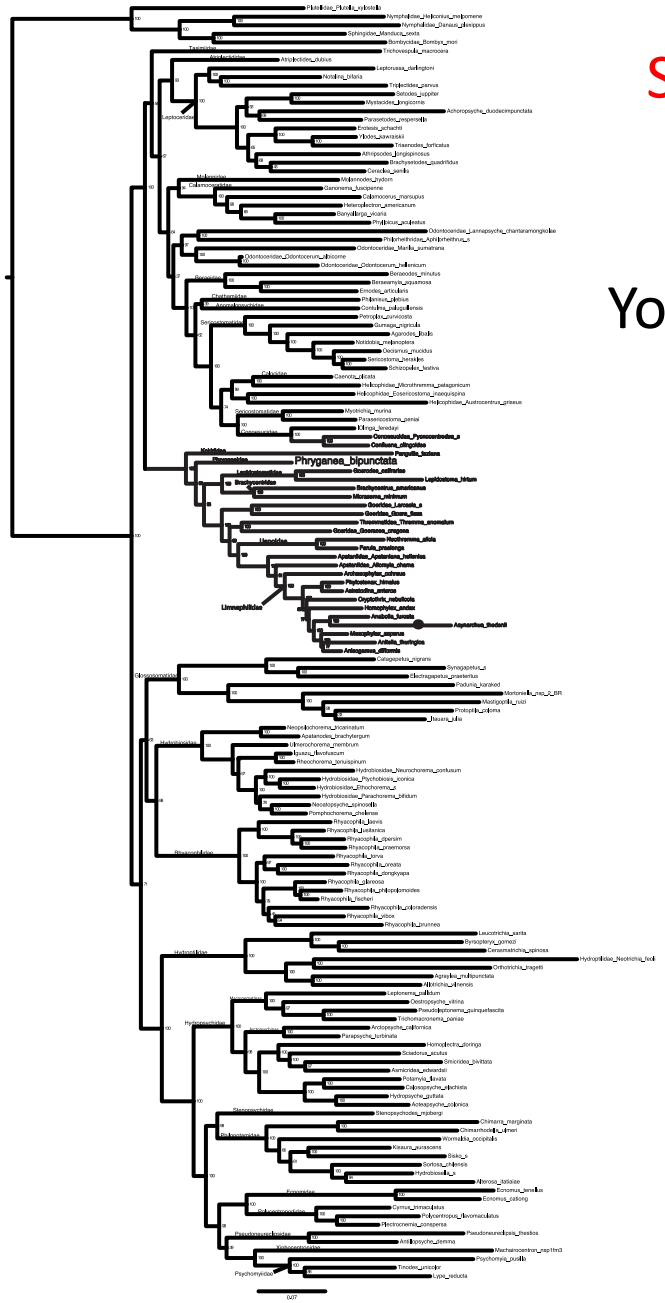

Annulipalpia:
Retreat makers
Integripalpia:
Tube case makers
Lets tie it together
0.4
6 genes,
2,000,000 nts
20 years
Hydropsychoidea
Philopotamoidea
Psychomyioidea
Hydroptiloidea
Glossosomatidae
Hydrobiosidae
Rhyacophilidae
Limnephiloidea
Leptoceroidea
Sericostomatoidea


Annulipalpia:
Retreat makers
Integripalpia:
Tube case makers
Lets tie it together
0.4
TRANSCRIPTOMES: 3500 genes, 180,000,000 nucleotides
Macrostemum: Hydropsychidae
Arctopsyche: Hydropsychidae
Diplectrona: Hydropsychidae
Sciadorus: Hydropsychidae
Stenopsychodes: Stenopsychidae
Stenopsyche: Stenopsychidae
Chimarra: Philopotamidae
Phylocentropus: Dipseudopsidae
Dipseudopsis: Dipseudopsidae
Psychomyia: Psychomyiidae
Plectrocnemia: Polycentropodidae
Parecnomina: Ecnomidae
Ptilocolepus: Ptilocolepidae
Palaeagapetus: Ptilocolepidae
Oxyethira: Hydroptilidae
Hydroptila: Hydroptilidae
Agapetus: Glossosomatidae
Glossosoma: Glossosomatidae
Apsilochorema: Hydrobiosidae
Himalopsyche: Rhyacophilidae
Rhyacophila: Rhyacophilidae
Phryganopsyche: Phryganopsychidae
Tascina: Oeconesidae
Silvatares: Pisuliidae
Phryganea: Phryganeidae
Ptilostomis: Phryganeidae
Micrasema: Barchycentridae
Lepistoma: Lepistomatidae
Goera: Goeridae
Neophylax: Uenoidae
Apatania: Apataniidae
Dicosmoecus: Limnophilidae
Ecclisomyia: Limnophilidae
Drusus: Limnophilidae
Glyphotaelius: Limnophilidae
Limnocentropus: Limnocentropodidae
Tasimia: Tasimiidae
Austrheithrus: Philorheithridae
Kosrheithrus: Philorheithridae
Psilotreta: Odonticeridae
Molanna: Molannidae
Anisocentropus: Calamoceratidae
Nectopsyche: Leptoceridae
Leptocerus: Leptoceridae
Chathamia: Chathamiidae
Alloecella: Helicophidae
Helicopsyche: Helicopsychidae
Petrothrincus: Petrothrincidae
Fattigia: Sericostomatidae
Barbarochthon: Barbarcochthonidae
Tamasia: Calocidae
Conoesucus: Conoesucidae
Costora: Conoesucidae
Hydropsychoidea
Philopotamoidea
Psychomyioidea
Hydroptiloidea
Glossosomatidae
Hydrobiosidae
Rhyacophilidae
Limnephiloidea
Leptoceroidea
Sericostomatoidea



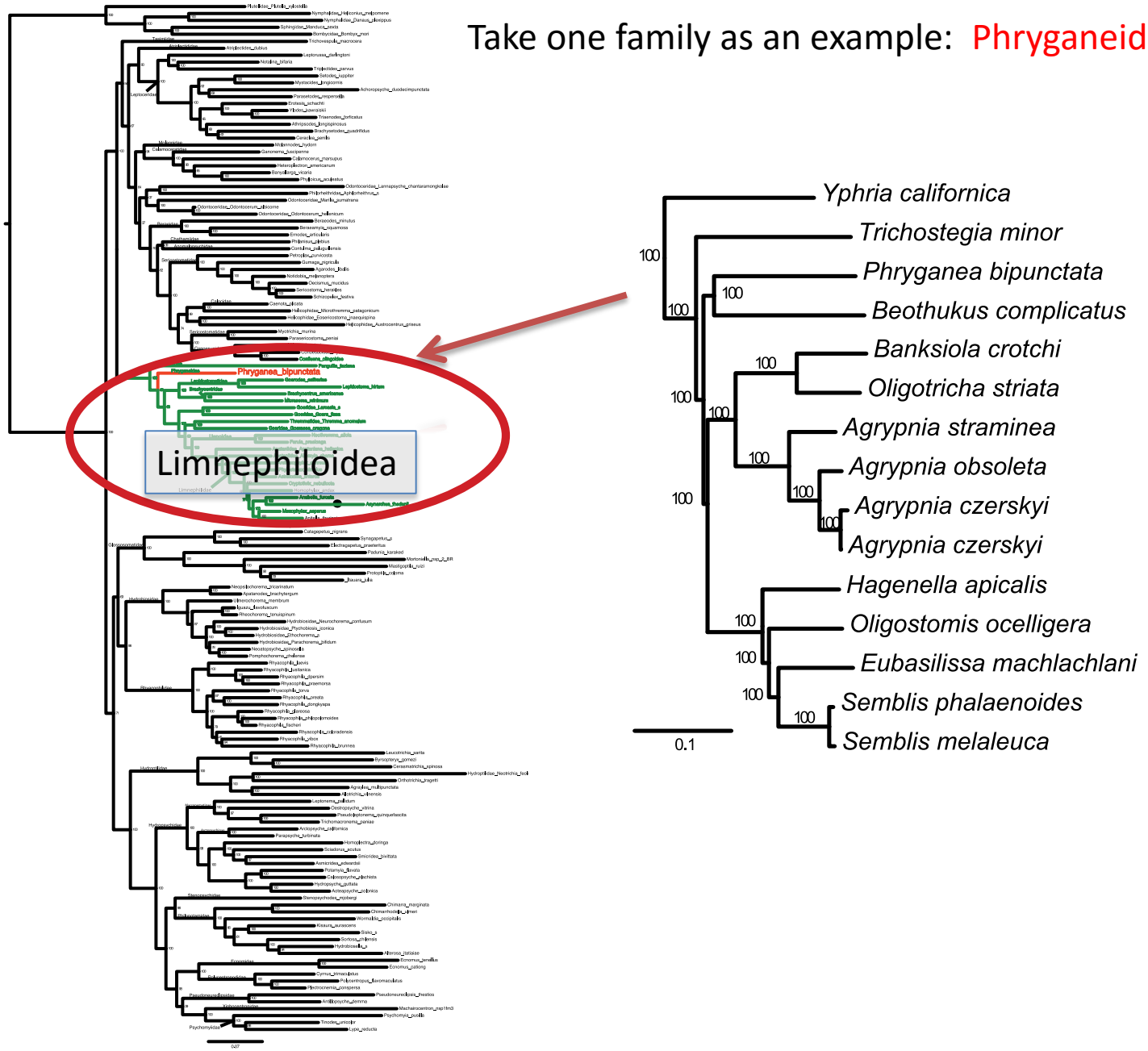


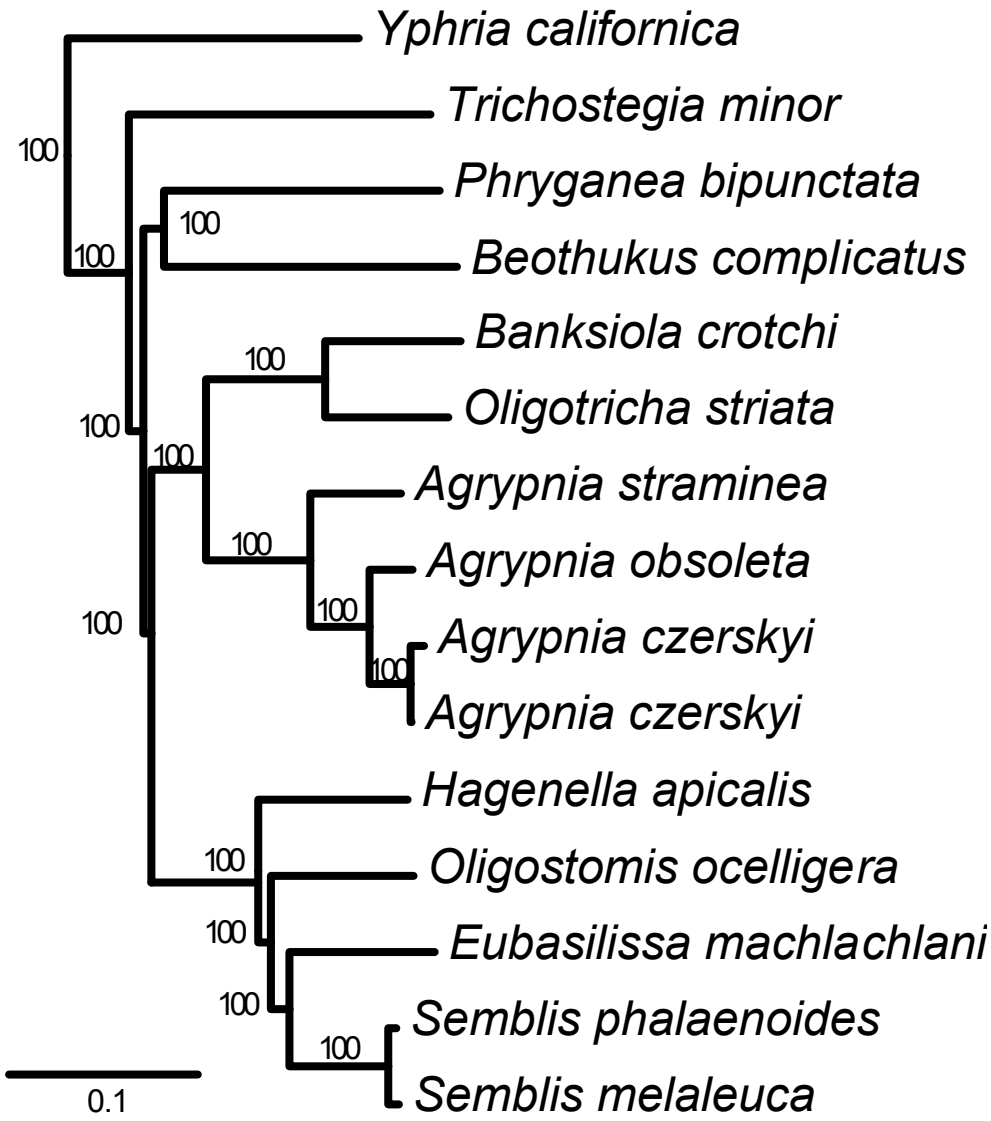
***Step 1*: Generate a phylogeny among families from transcriptomes (3500 genes)**



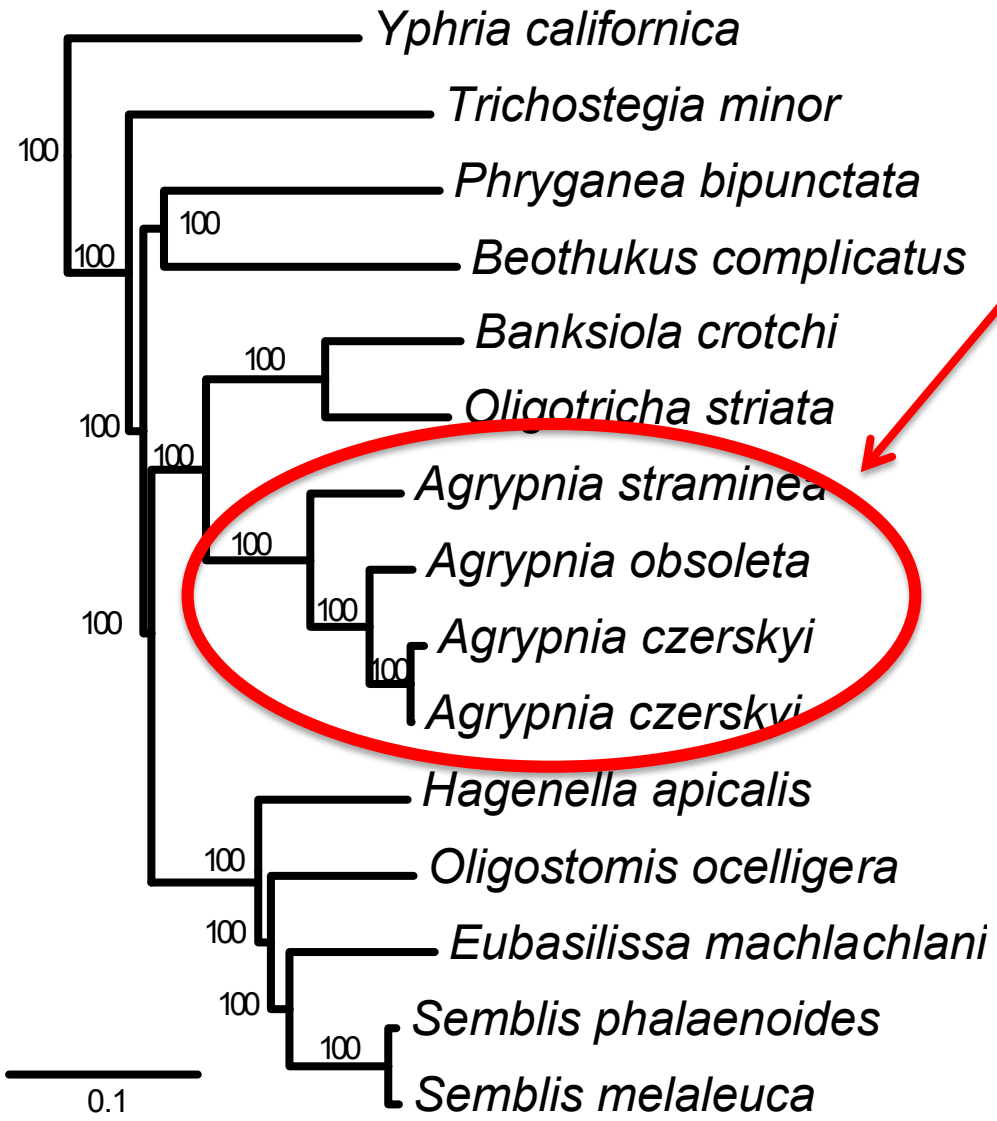
Step 2: Use transcriptomes to design probes to add compatible targeted enrichment techniques (900 genes). You now have a genus level phylogeny with big data.

Take one family as an example: Phryganeidae

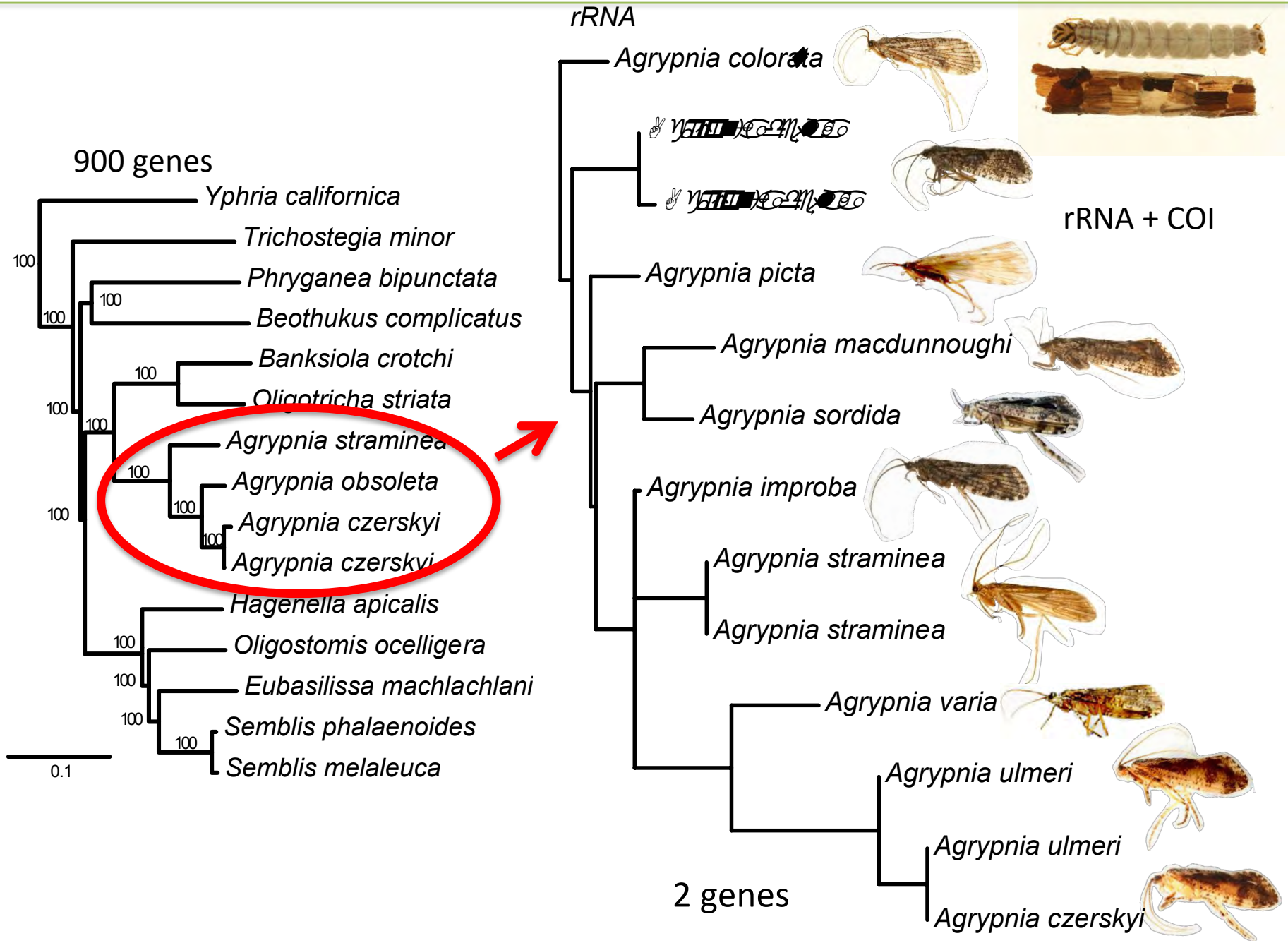


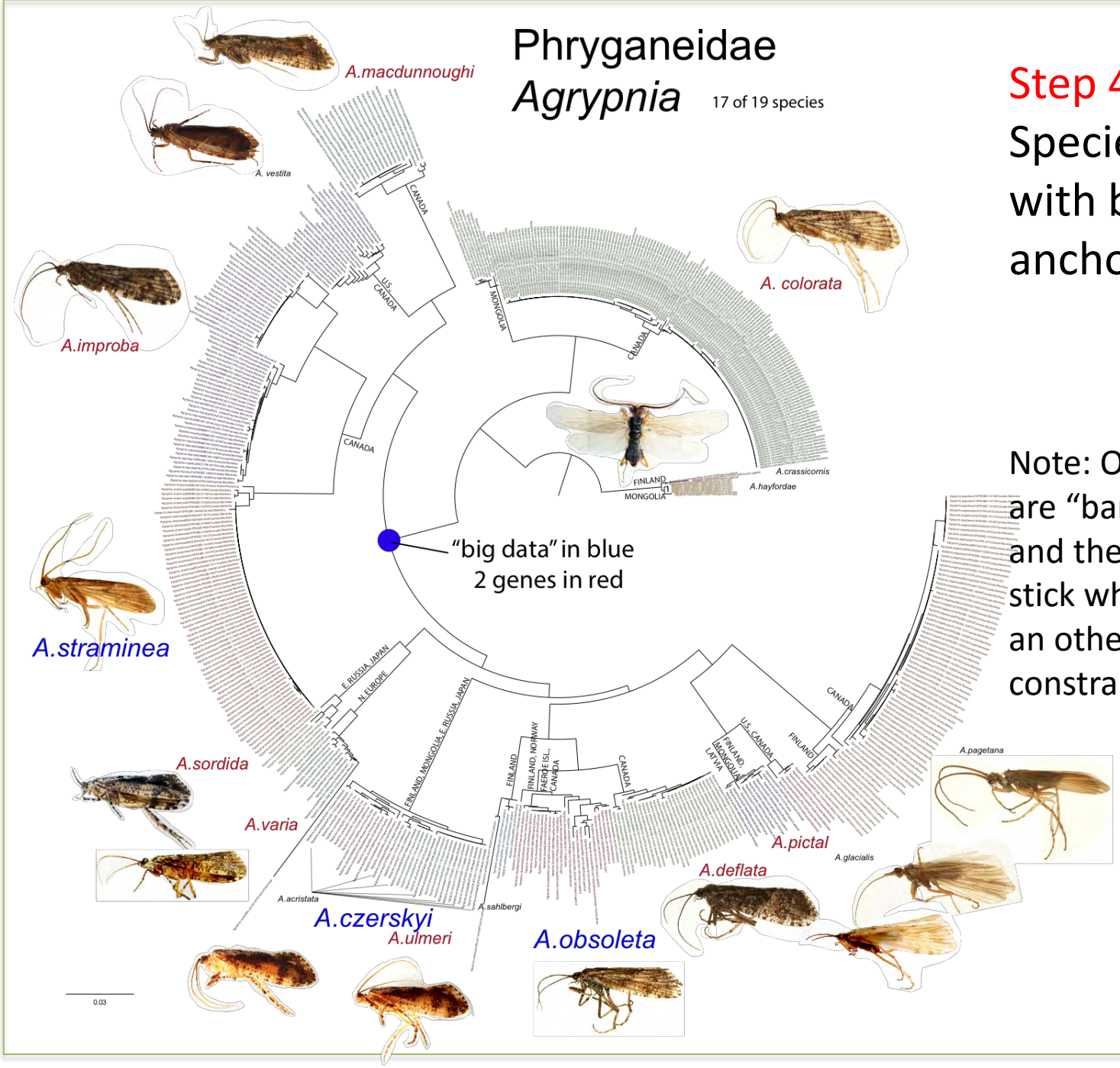


We have 900 genes for all but one of the genera



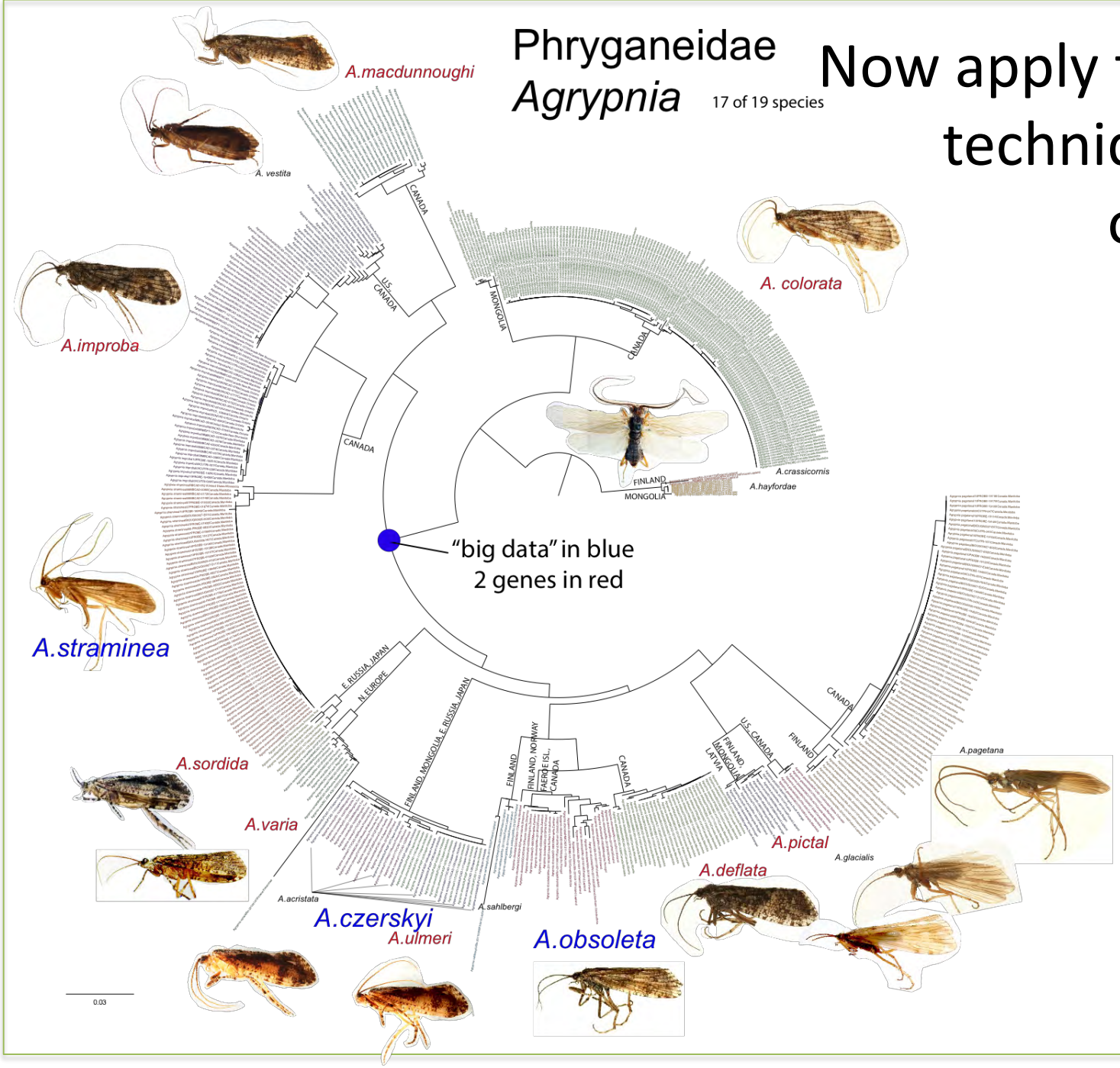
Step 3: look at one genus; *Agrypnia* for example


900 genes
Yphria californica
Trichostegia minor
Phryganea bipunctata
Beothukus complicatus
Banksiola crotchi
Oligotricha striata
Agrypnia straminea
Agrypnia obsoleta
Agrypnia czerskyi
Agrypnia czerskyi
Hagenella apicalis
Oligostomis ocelligera
Eubasilissa machlachlani
Semblis phalaenoides
Semblis melaleuca
100
0.1
rRNA
Agrypnia colorata
rRNA + COI
Agrypnia picta
Agrypnia macdunnoughi
Agrypnia sordida
Agrypnia improba
Agrypnia straminea
Agrypnia straminea
Agrypnia varia
Agrypnia ulmeri
Agrypnia ulmeri
Agrypnia czerskyi
2 genes



Step 4: fill out Species phylogeny with barcode data, anchored by rRNA

Note: Only 5 taxa are "barcode only" and these need only stick where they go in an otherwise constrained tree


Phryganeidae
Agrypnia
17 of 19 species
Now apply this same technique to the other 549 genera
A.macdunnoughi
A. vestita
A.improba
A. colorata
A.straminea
A.crassicornis
A.hayfordae
"big data" in blue
2 genes in red
A.sordida
A.varia
A.pictal
A.pagetana
A.glacialis
A.deflata
A.acristata
A.sahlbergi
A.czerskyi
A.ulmeri
A.obsoleta
CANADA
U.S. CANADA
MONGOLIA
FINLAND
MONGOLIA
E RUSSIA, JAPAN
N EUROPE
FINLAND, MONGOLIA, E RUSSIA, JAPAN
FINLAND
FINLAND, NORWAY
FAEROE ISL., CANADA
CANADA
U.S. CANADA
FINLAND, MONGOLIA, LATVIA
CANADA
FINLAND
0.03

The End

Acknowledgments:
Ralph Holzenthal, and the U of Minnesota Insect Museum

Oliver Flint, and the National Museum of Natural History

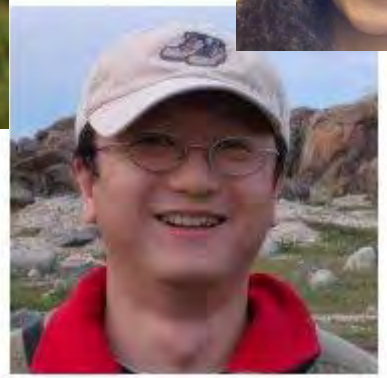
Jessica Ware, former graduate student

Paul Frandsen, former graduate student

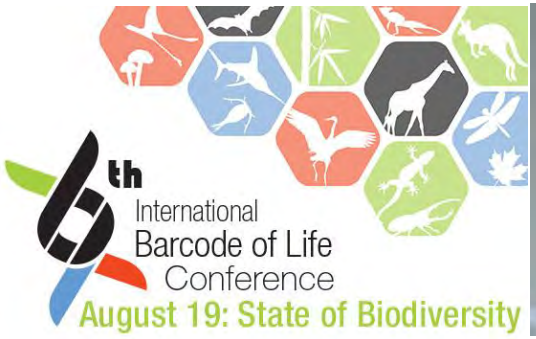
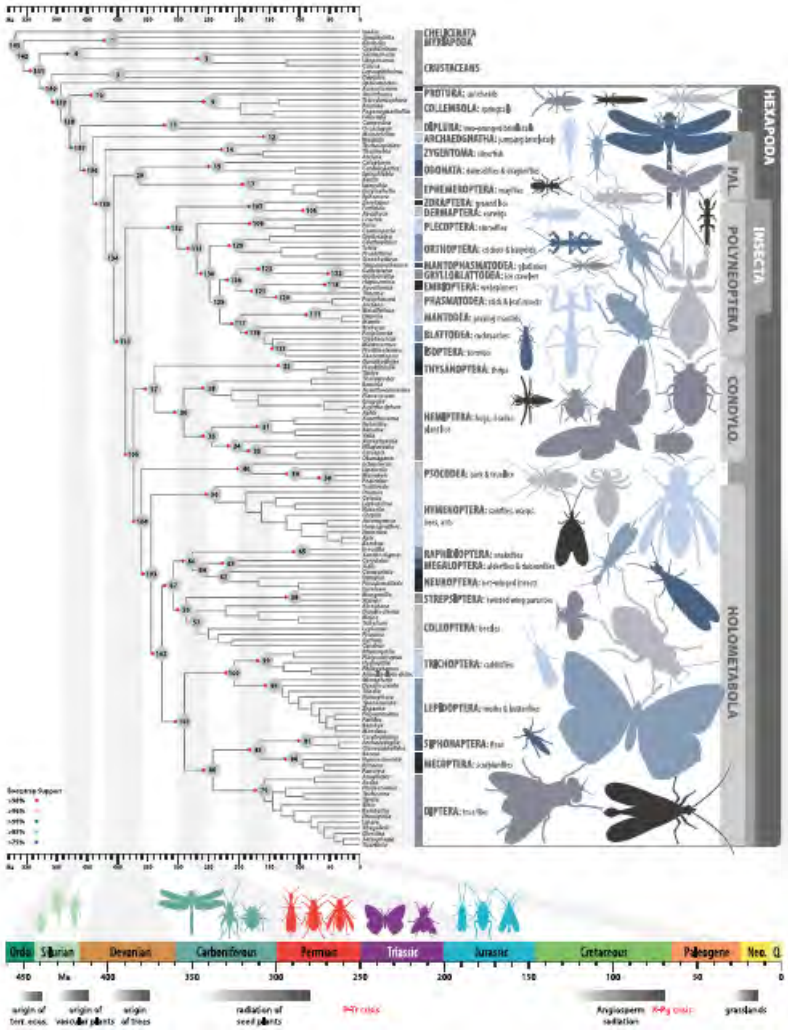
Xin Zhou, former graduate student
once here, now at BGI

Karen Meusemann, 1KITE

Bernhard Misof, and 1KITE



Merhdad Hajibabael and the organizers
for inviting me!



Thank you! …please check out www.1kite.org, www.boldsystems.org/, www.tolweb.org/Trichoptera