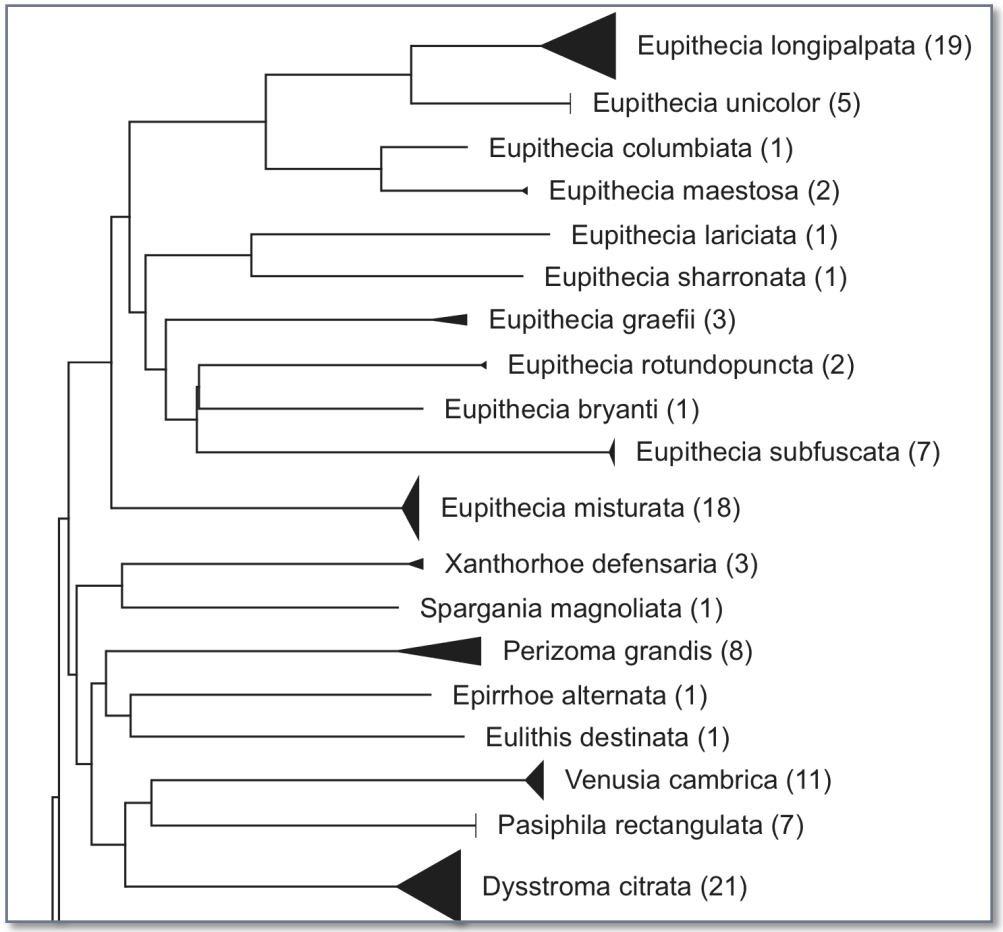
# The 'barcode approach' to combining molecular datasets: Reconstructing the phylogeny of North America's Geometridae as an example

Jeremy deWaard, Charlie Mitter,
Axel Hausmann and Leland Humble

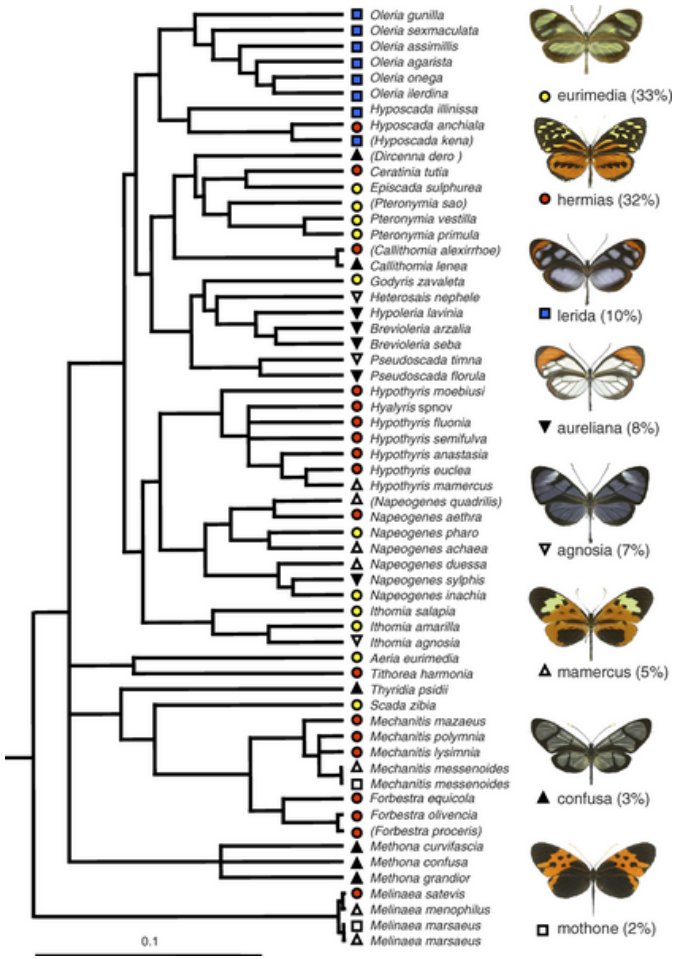
# A little background…

- exploring two novel 'tools' for monitoring forest Lepidoptera diversity:
  - barcodes
  - molecular phylogenies




COI tree used to sort, identify and inventory nocturnal Lepidoptera in Stanley Park, Vancouver (Fig.1 (in part), deWaard et al 2009).

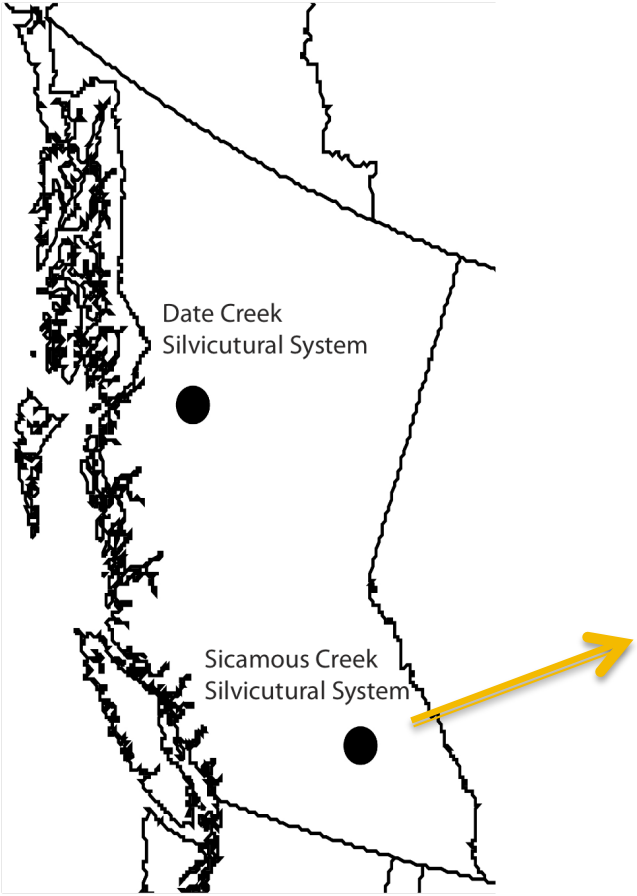
# Application of phylogenies

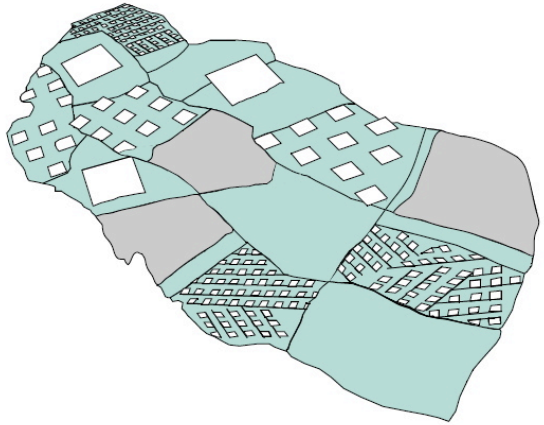


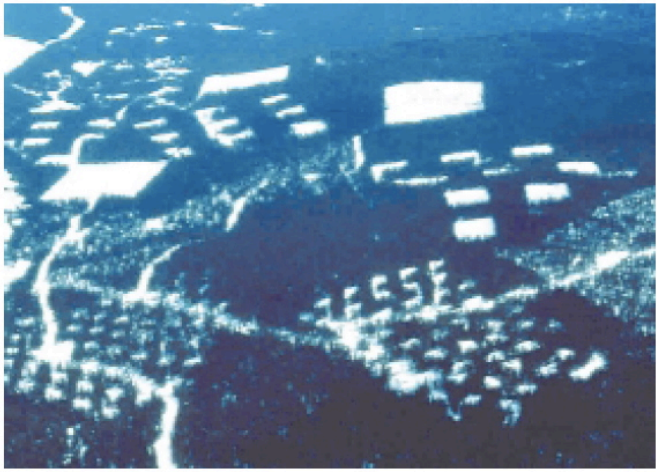

A community phylogeny of ithomiine butterflies found in the Anagu region of Napo, Ecuador (Elias et al 2008).

- can construct community phylogenies:
  - understanding the composition/origin of ecological communities (e.g. Webb et al. 2006)
  - estimating phylogenetic diversity (PD)

# The effect of harvest types on multiple levels of moth diversity



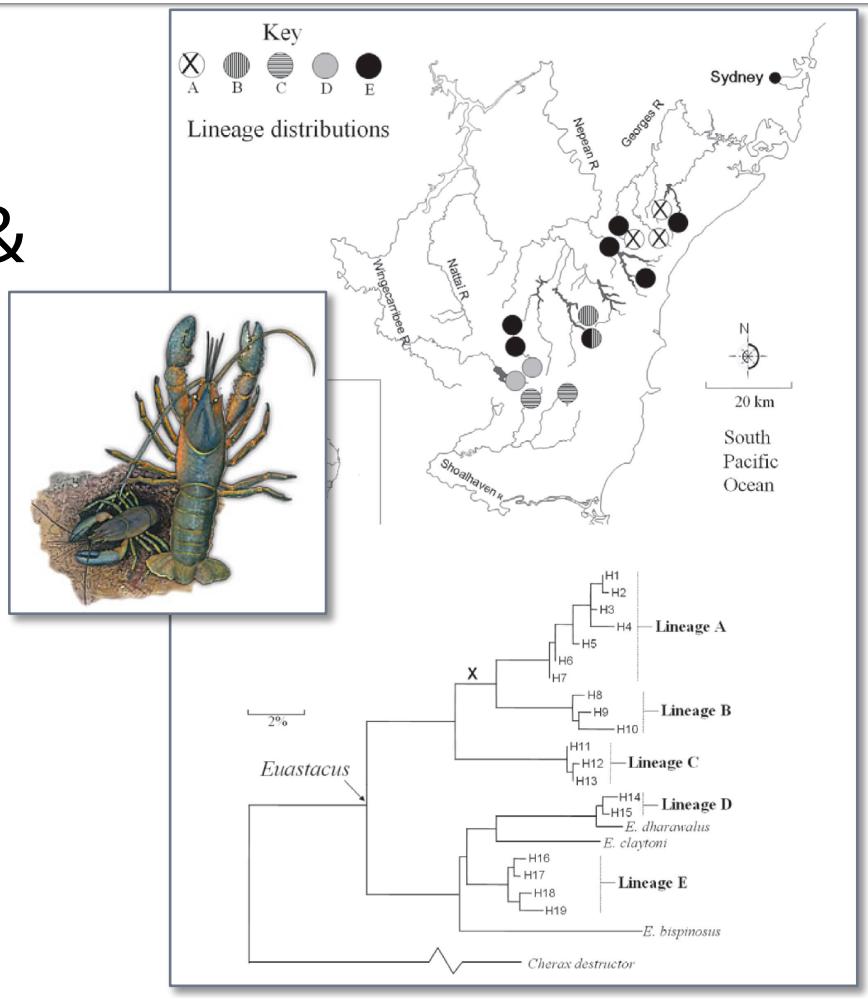





Sicamous Creek Silvicultural System (Huggard & Vyse 2002)

# The problem

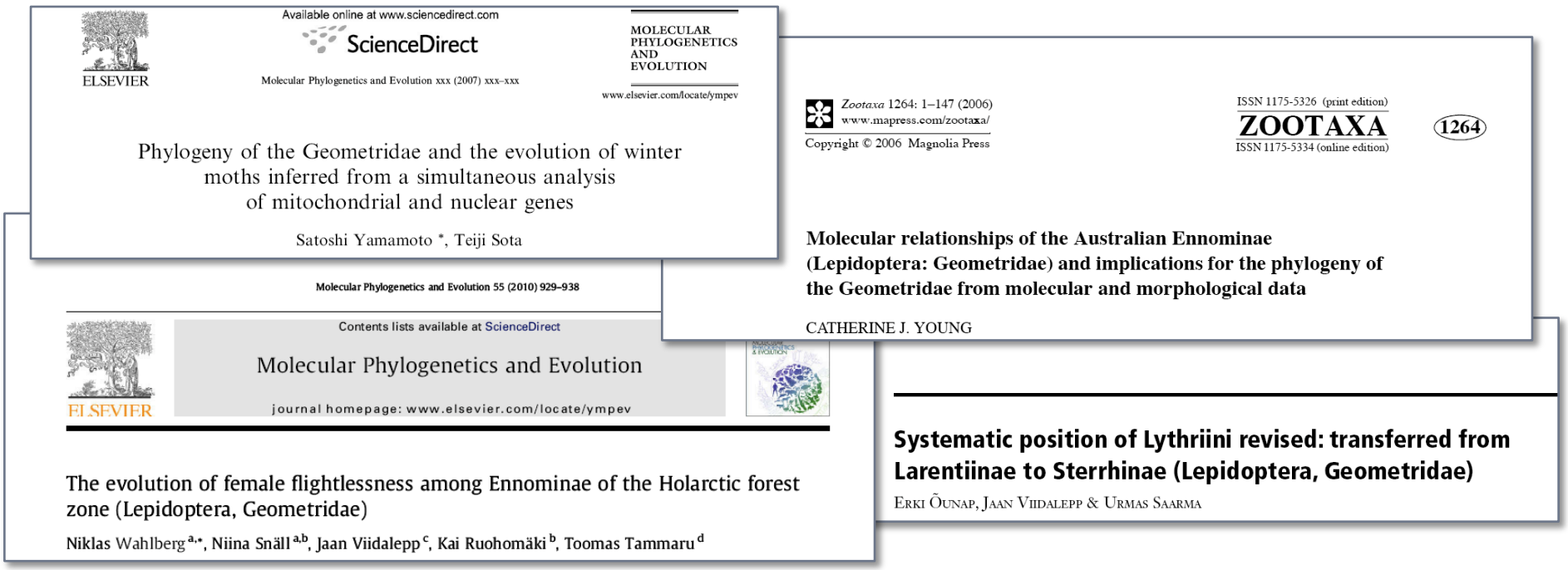
- COI alone can be a good proxy for PD (see Faith & Baker 2006)
- untested at higher taxonomic levels where COI has less signal (e.g. Wilson 2010)




Example of PD calculated with COI: spiny crayfish in SE Australia (Baker et al. 2004)

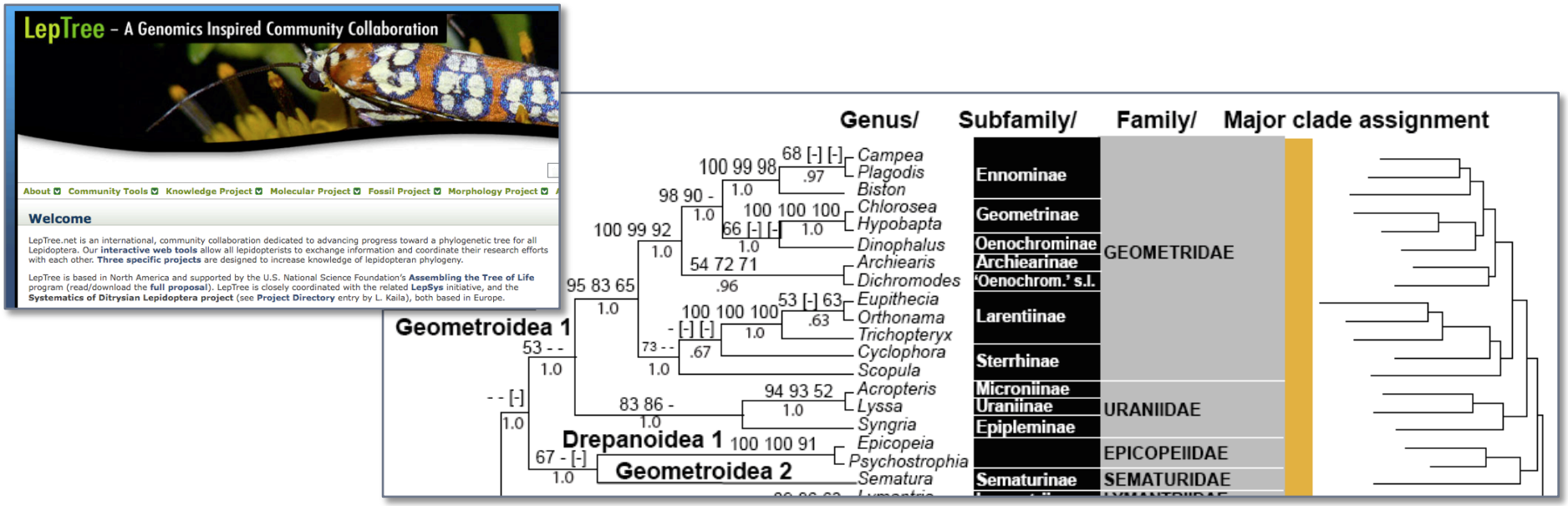
# The objective

- multi-gene, tribe-level phylogeny of NA geometrids: scaffold to constrain COI trees at BC collection sites
  - contribute a North American perspective to deep geometrid systematics



# LepTree collaboration

- build upon large LepTree dataset
  - but RT-PCR of nuclear protein-coding genes from fresh/frozen specimens

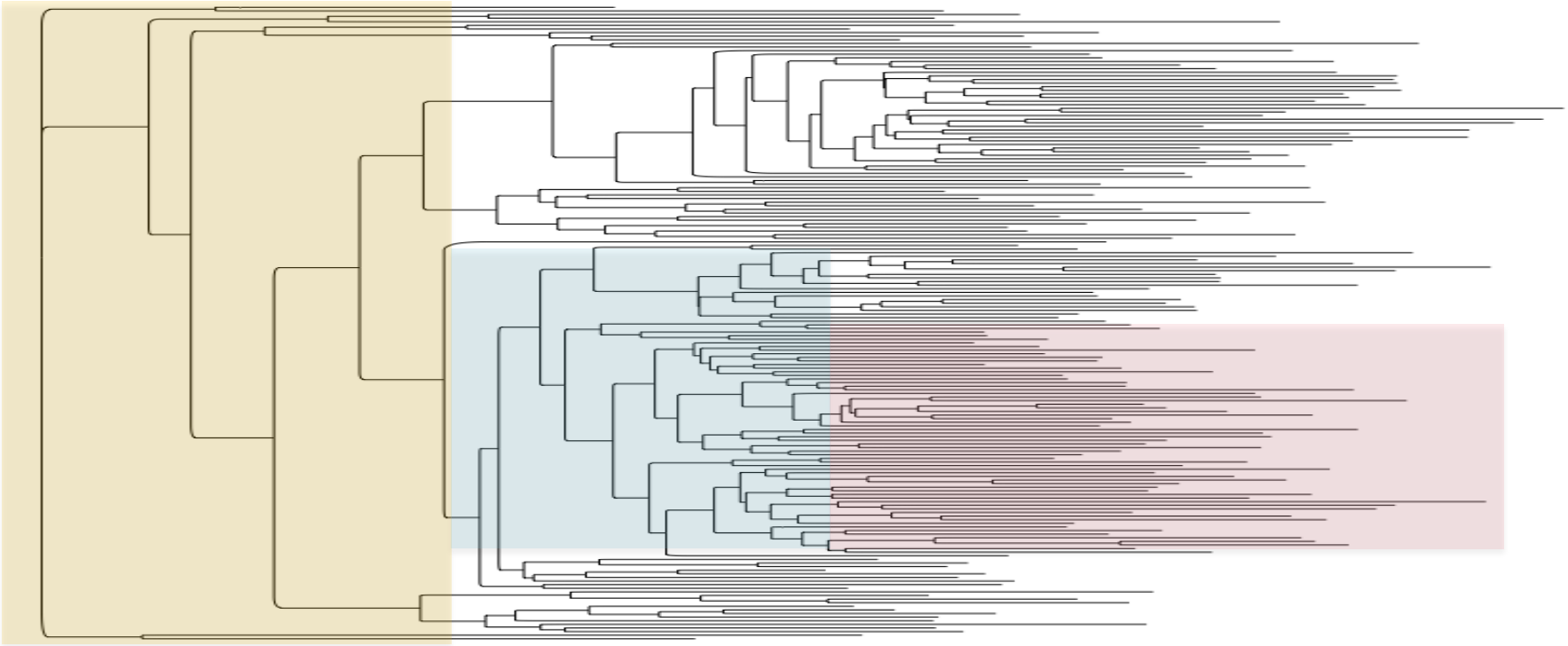



LepTree website (www.leptree.net) and the geometroid section of the ML tree from Regier et al 2009.

# Reconstructing the smaller branches and tips

deep mid shallow

decrease in the availability of recently collected exemplar taxa



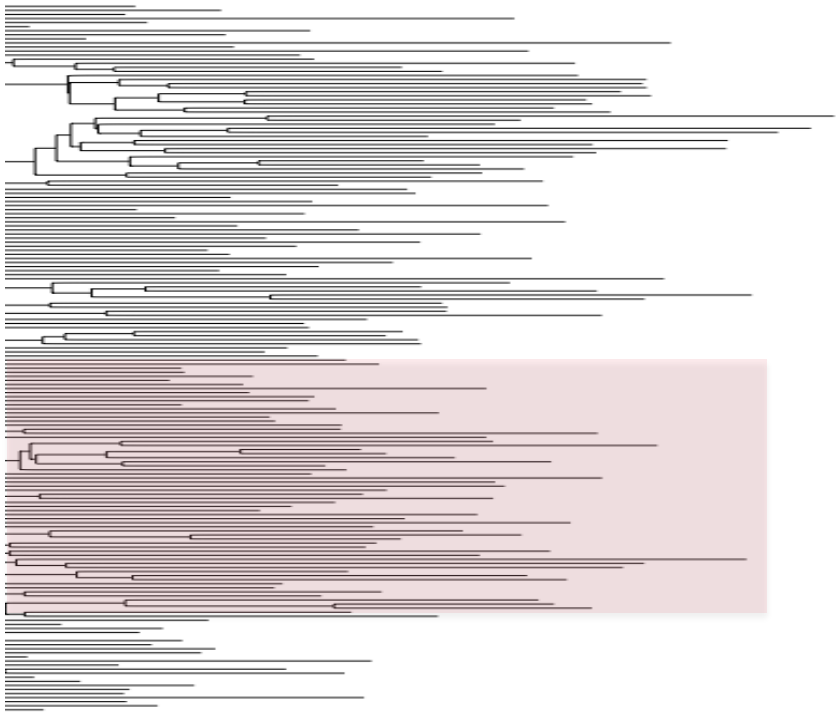
# Reconstructing the smaller branches and tips

deep mid shallow

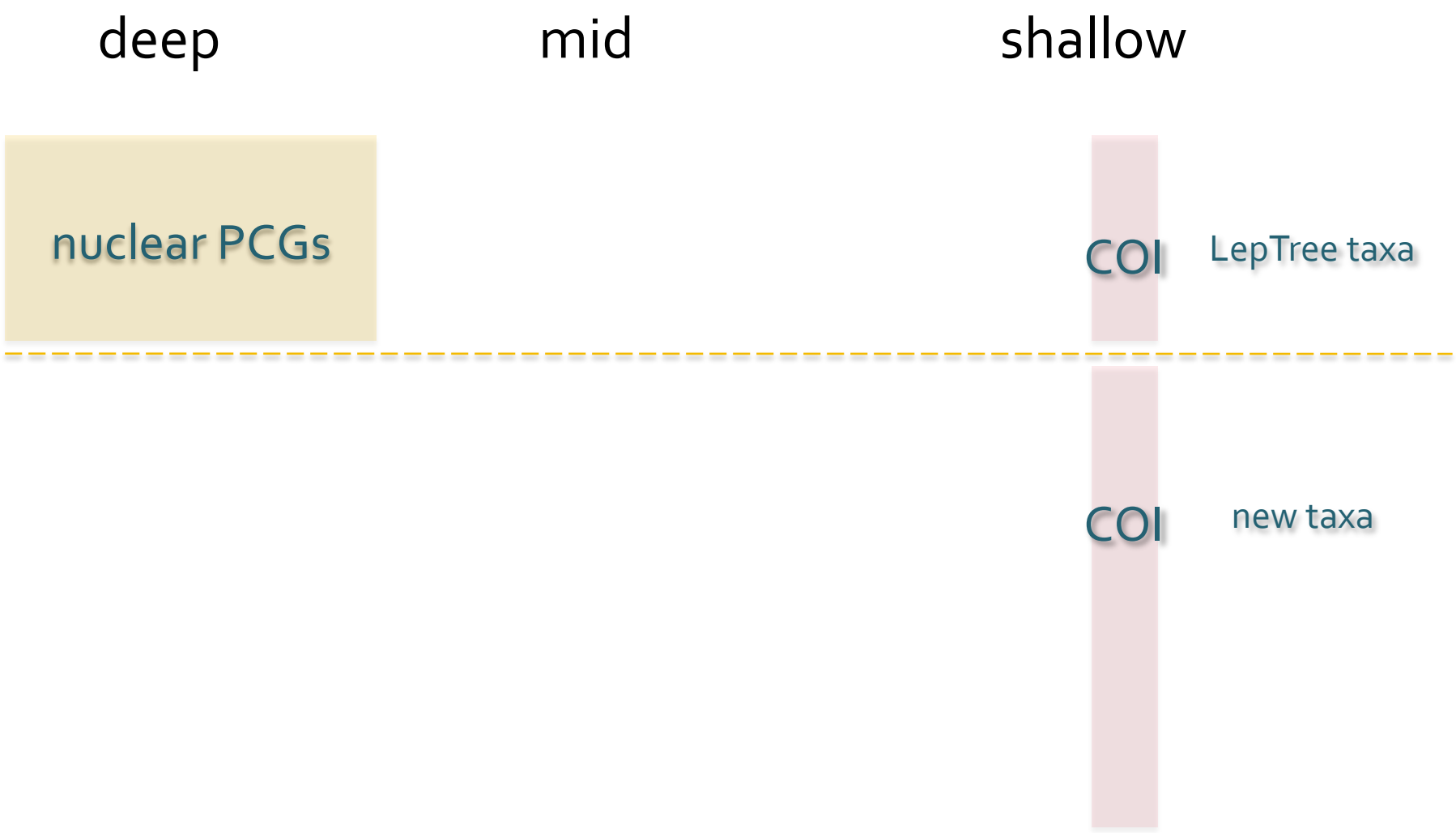
decrease in the availability of fresh exemplar taxa

except: barcoding projects are analyzing heaps of relatively fresh species (particularly in geometrids!)

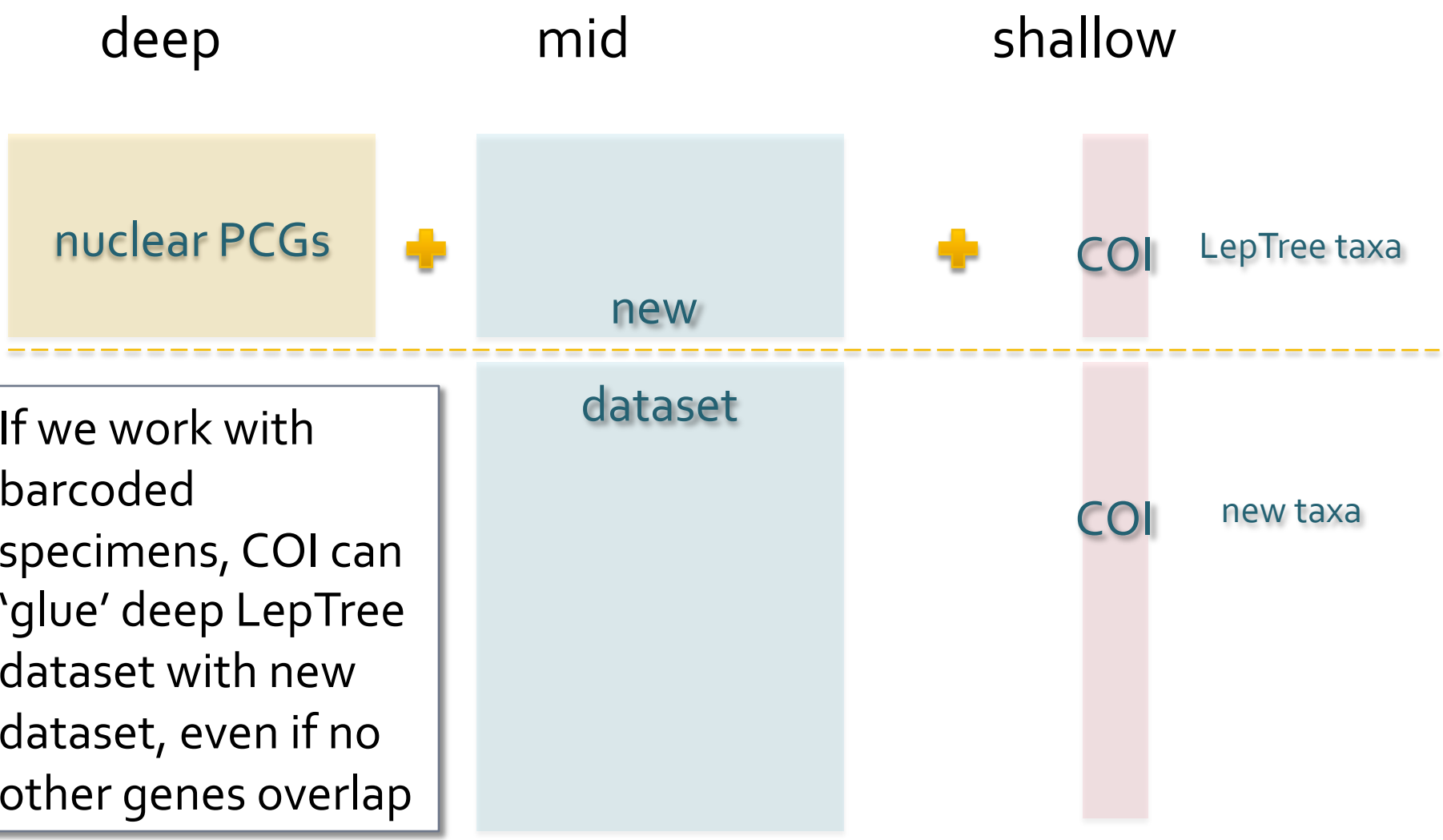
and: many deep TOL projects are now barcoding too



# Sampling strategy

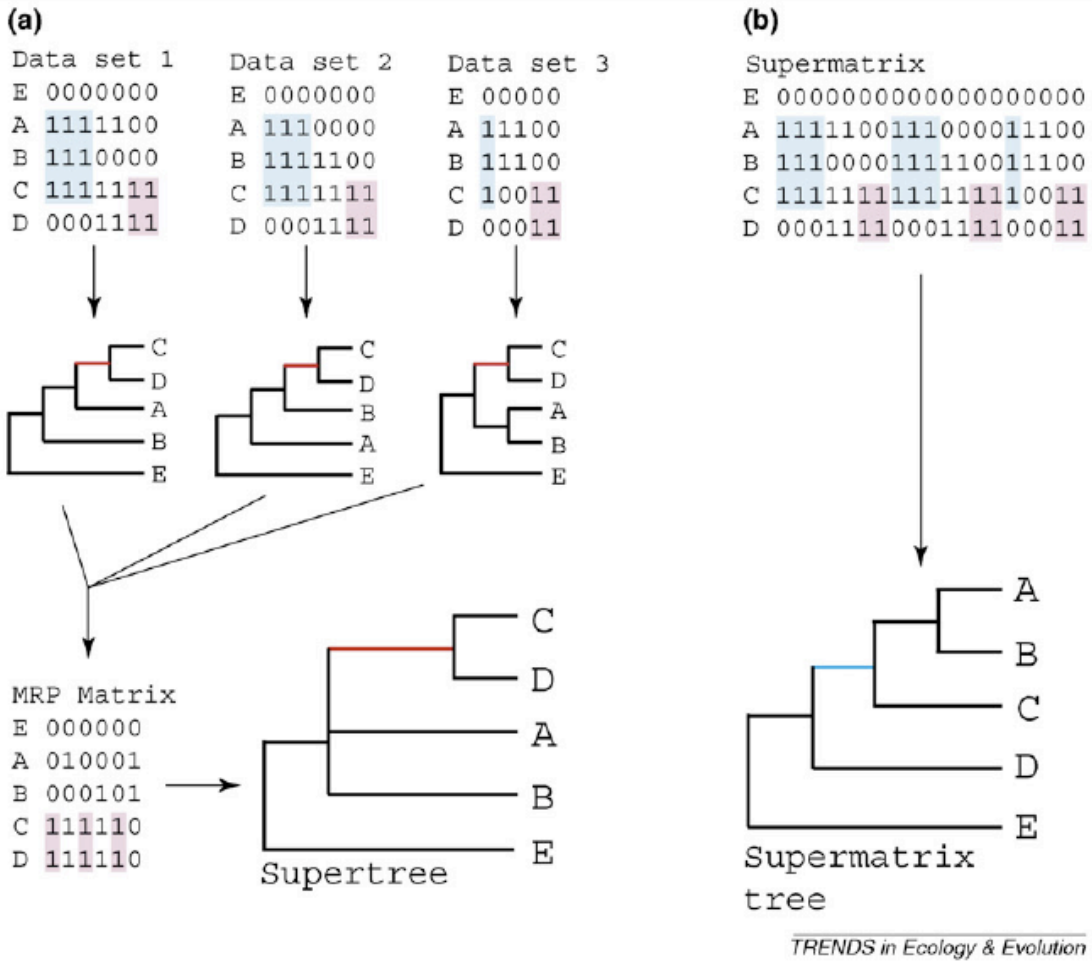


# Sampling strategy



# 'Barcode approach' for combining molecular datasets

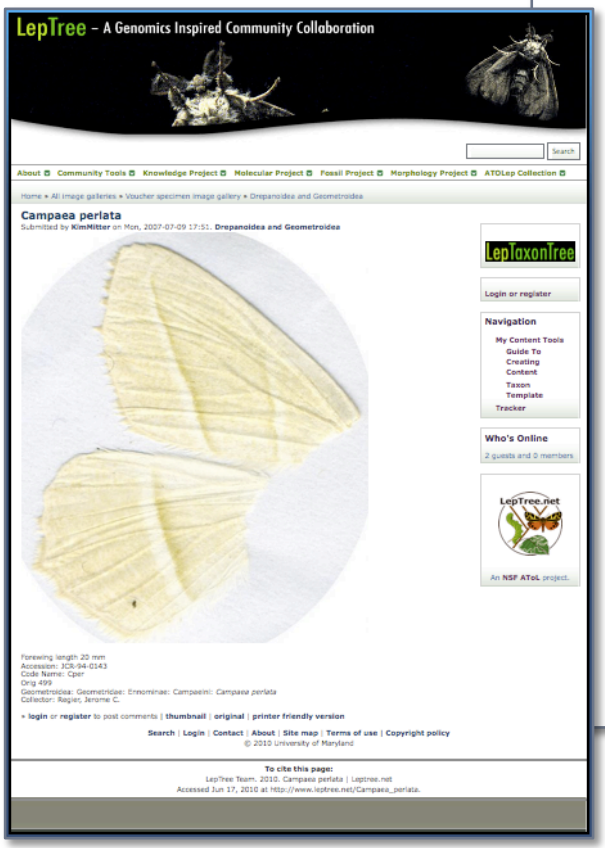
- instance of the supermatrix approach
  - allows calculation of branch lengths




(a) Supertree and (b) supermatrix approaches, shown combining binary data (de Queiroz and Gatesy 2007).

# Breakdown of samples

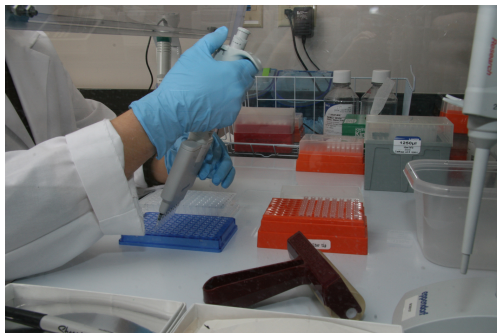
- most from Alcohol Tubes of Lepidoptera (ATOLep) collection
- rare NA tribes from Dan Janzen, ZSM, CNC, SI
- included non-NA taxa
- 284 samples

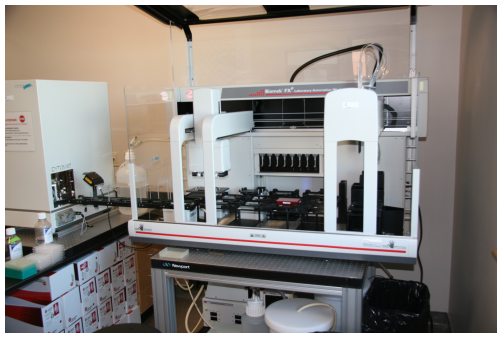


# Laboratory methods

- Canadian Centre of DNA Barcoding in Guelph

  - 15 primer sets (+ repeated COI-5p)
    - COI-3p-COII
    - 16S
    - 18S
    - 28S-D2
    - EF-1α (4)
    - Wingless
    - CAD (4)
    - DDC
    - IDH

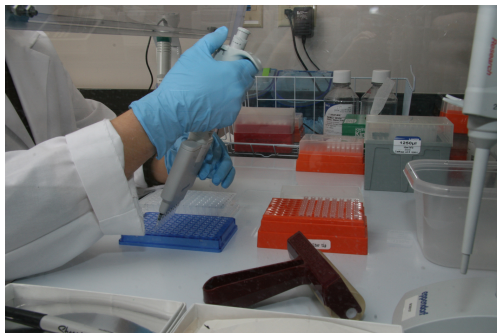


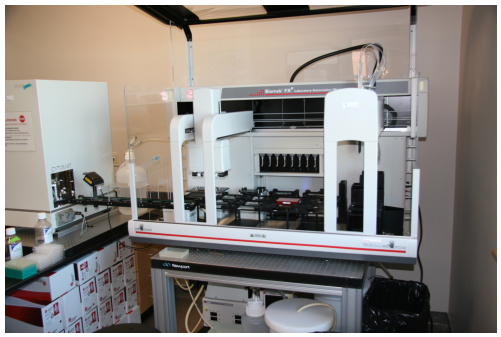


# Laboratory methods

- Canadian Centre of DNA Barcoding in Guelph

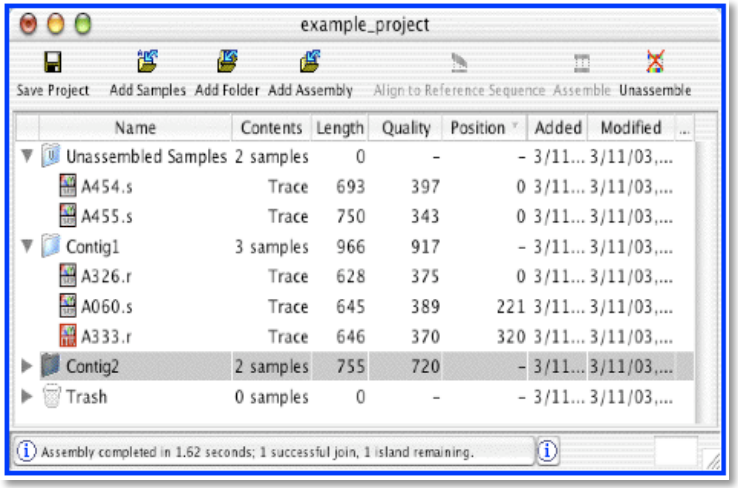
  - 15 primer sets (+ repeated COI-5p)
    - COI-3p-COII (919bp)
    - 16S (340bp)
    - 18S (612bp)
    - 28S-D2 (249bp)
    - EF-1α (924bp)
    - Wingless ✖
    - CAD (4) ✖
    - DDC ✖
    - IDH ✖



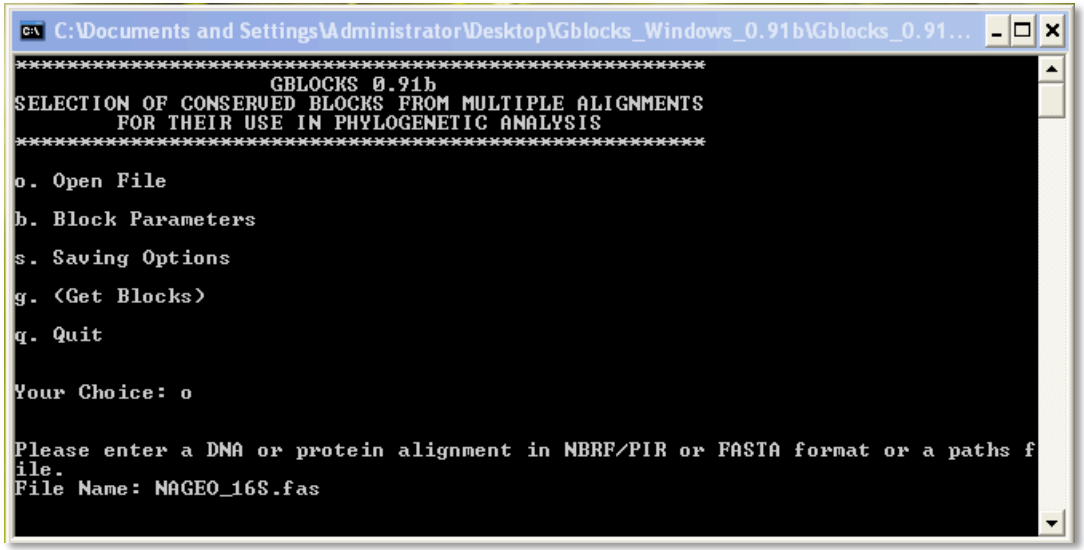


# Alignments

- Sequence editing & alignments with CodonCode Aligner

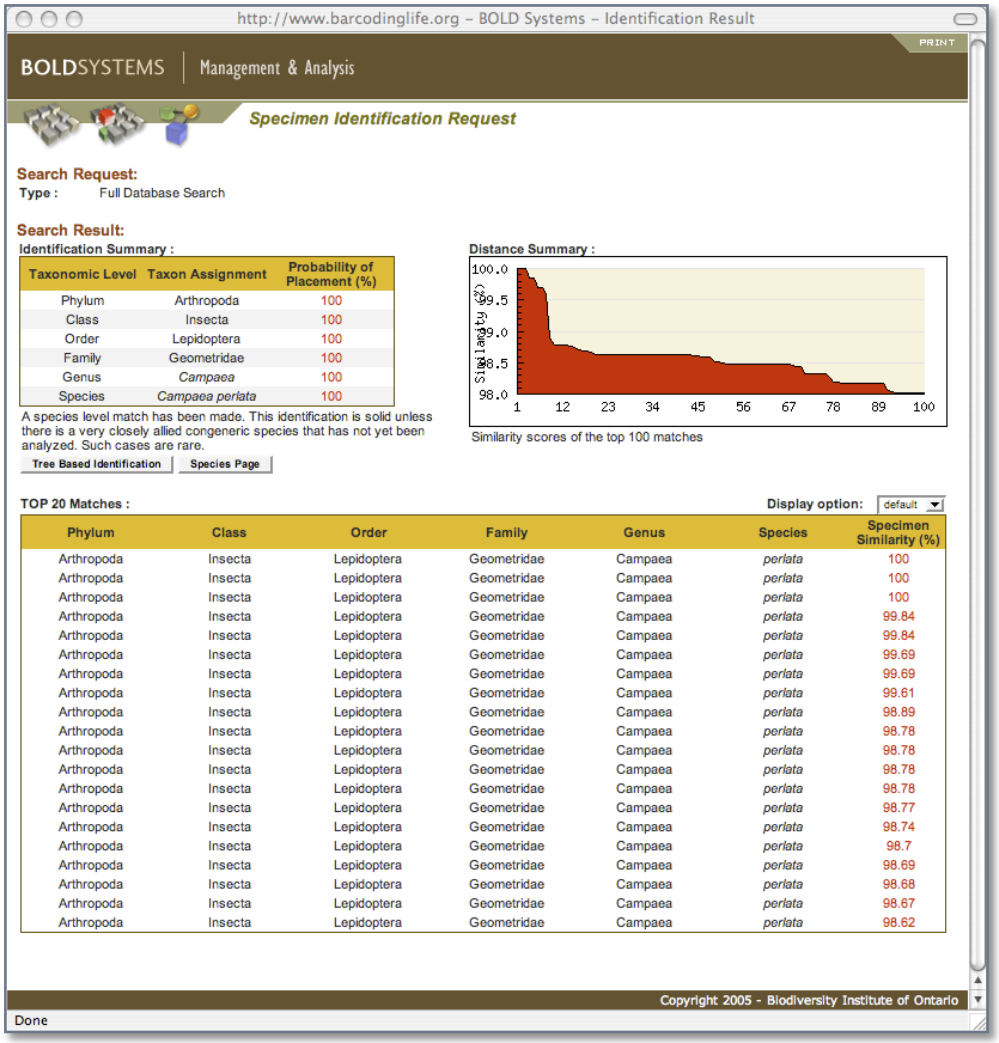



- Gblocks for selecting ambiguous sites to remove from rDNA

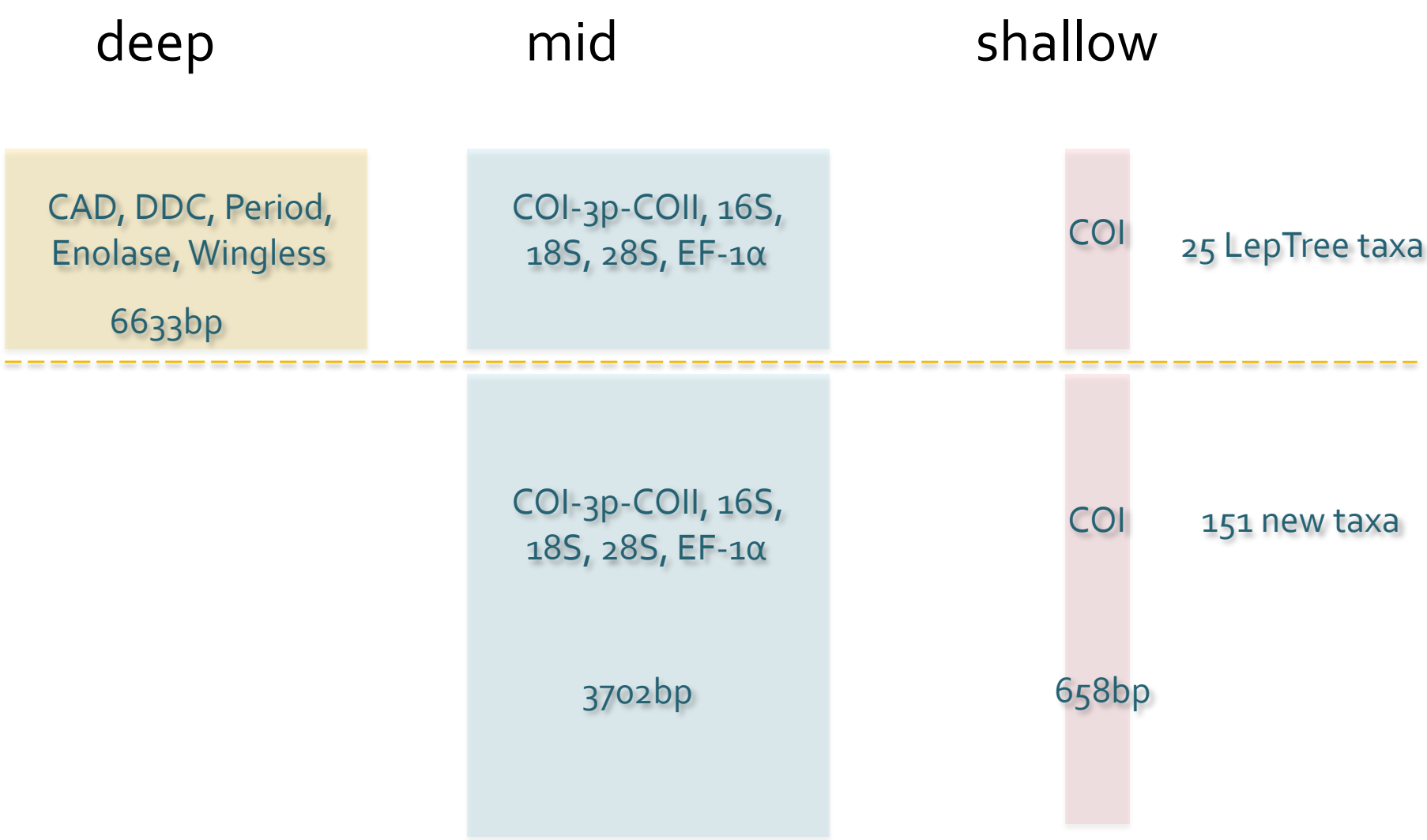


# Confirmation of identifications

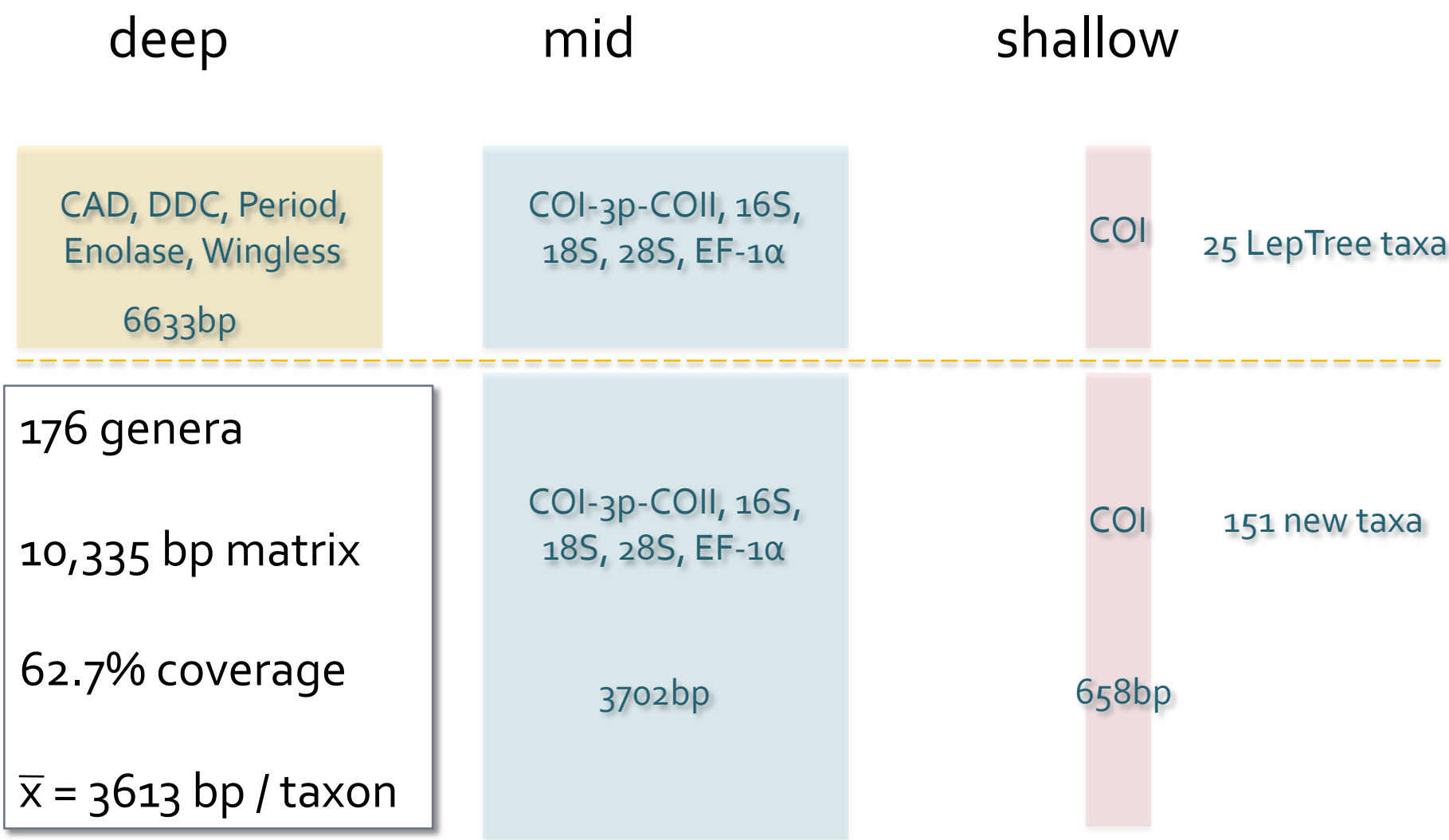
- BOLD-ID to confirm identifications (as in Regier et al. 2009, Wahlberg et al 2010)
- some misIDs from the ATOLep collection (removed)



# Final dataset

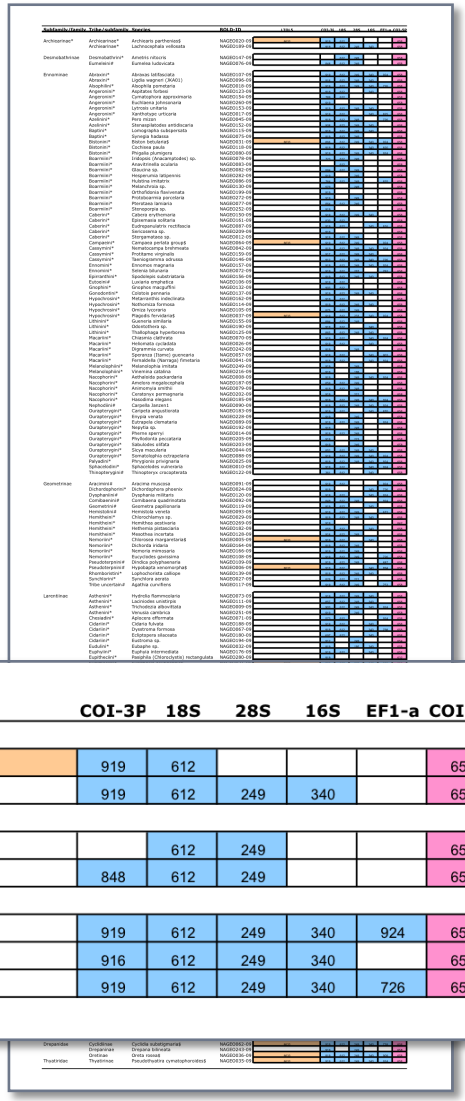


# Final dataset



# Final taxonomic sample

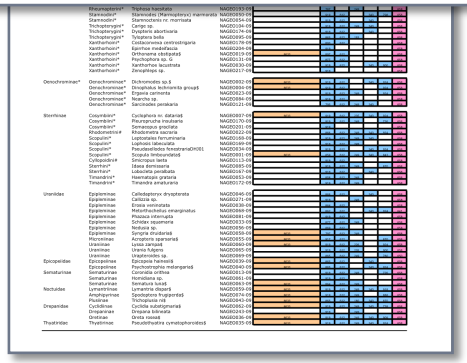
- 176 taxa
- ingroup:
  - 7 subfamilies (no Orthostixinae), 61 tribes, 152 genera
  - all 53 suprageneric taxa in NA
  - 15 of the tribes not found in NA



| Subfamily/family | Tribe/subfamily | Species | BOLD-ID | LTOL5 | COI-3P | 18S | 28S | 16S | EF1-a | COI-5P |
|---|---|---|---|---|---|---|---|---|---|---|
| Archiearinae* | Archiearinae* | Archiearis parthenias$ | NAGEO020-09 | 6633 | 919 | 612 | | | | 658 |
| | Archiearinae* | Lachnocephala vellosata | NAGEO189-09 | | 919 | 612 | 249 | 340 | | 658 |
| Desmobathrinae | Desmobathrini* | Ametris nitocris | NAGEO147-09 | | | 612 | 249 | | | 658 |
| | Eumeleini# | Eumelea ludovicata | NAGEO076-09 | | 848 | 612 | 249 | | | 658 |
| Ennominae | Abraxini* | Abraxas latifasciata | NAGEO107-09 | | 919 | 612 | 249 | 340 | 924 | 658 |
| | Abraxini* | Ligdia wagneri (JKA01) | NAGEO096-09 | | 916 | 612 | 249 | 340 | | 658 |
| | Alsophilini* | Alsophila pometaria | NAGEO018-09 | | 919 | 612 | 249 | 340 | 726 | 658 |

# Final taxonomic sample

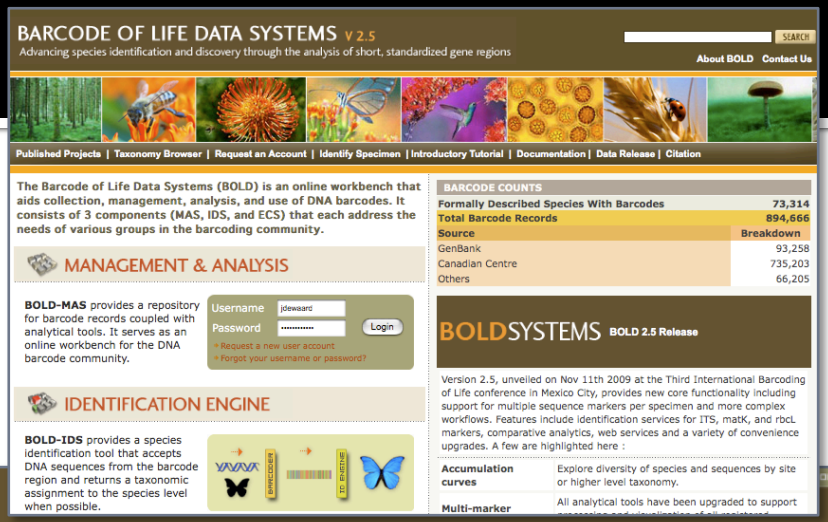
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Uraniidae | Epipleminae | Calledapteryx dryopterata | NAGEO046-09 | | 866 | 612 | | 340 | | 658 |
| | Epipleminae | Callizzia sp. | NAGEO271-09 | | 919 | | 249 | | | 658 |
| | Epipleminae | Erosia veninotata | NAGEO038-09 | | 864 | 612 | | | | 658 |
| | Epipleminae | Metorthocheilus emarginatus | NAGEO068-09 | | 862 | 612 | | 340 | 924 | 641 |
| | Epipleminae | Phazaca interrupta | NAGEO081-09 | | 919 | 612 | | | | 658 |
| | Epipleminae | Schidax squamaria | NAGEO033-09 | | 877 | 612 | 249 | | | 658 |
| | Epipleminae | Nedusia sp. | NAGEO056-09 | | 869 | 612 | | | | 658 |
| | Epipleminae | Syngria druidaria$ | NAGEO058-09 | 6633 | 797 | 612 | 249 | | | 658 |
| | Microniinae | Acropteris sparsaria$ | NAGEO055-09 | 6633 | 917 | 612 | | | 870 | 658 |
| | Uraniinae | Lyssa zampa$ | NAGEO060-09 | 6633 | 919 | 612 | 226 | | 924 | 658 |
| | Uraniinae | Urania fulgens | NAGEO065-09 | | 919 | 612 | 226 | | 900 | 658 |
| | Uraniinae | Urapteroides sp. | NAGEO069-09 | | 865 | 612 | 249 | | 780 | 650 |
| Epicopeiidae | Epicopeiinae | Epicopeia hainesii$ | NAGEO039-09 | 6633 | 865 | 612 | | 340 | | 658 |
| | Epicopeiinae | Psychostrophia melangaria$ | NAGEO040-09 | 6633 | 865 | 612 | | 340 | 924 | 658 |
| Sematuridae | Sematurinae | Coronidia orithea | NAGEO013-09 | | 919 | 612 | 249 | | 726 | 658 |
| | Sematurinae | Homidiana sp. | NAGEO061-09 | | 919 | 612 | | | | 658 |
| | Sematurinae | Sematura luna$ | NAGEO063-09 | 6633 | 864 | 612 | 249 | | | 658 |
| Noctuidae | Lymantriinae | Lymantria dispar$ | NAGEO059-09 | 6633 | 919 | 612 | 249 | 340 | 918 | 658 |
| | Amphipyrinae | Spodoptera frugiperda$ | NAGEO074-09 | 6633 | 919 | 612 | 249 | | 888 | 658 |
| | Plusiinae | Trichoplusia ni$ | NAGEO043-09 | 6633 | 865 | 612 | 180 | 340 | 870 | 658 |
| Drepanidae | Cyclidiinae | Cyclidia substigmaria$ | NAGEO062-09 | 6633 | 919 | 612 | 249 | 340 | 738 | 658 |
| | Drepaninae | Drepana bilineata | NAGEO243-09 | | 919 | | 249 | | | 658 |
| | Oretinae | Oreta rosea$ | NAGEO036-09 | 6633 | 919 | 612 | 249 | 340 | 909 | 658 |
| Thyatiridae | Thyatirinae | Pseudothyatira cymatophoroides$ | NAGEO035-09 | 6633 | 919 | 612 | 249 | 340 | 924 | 658 |

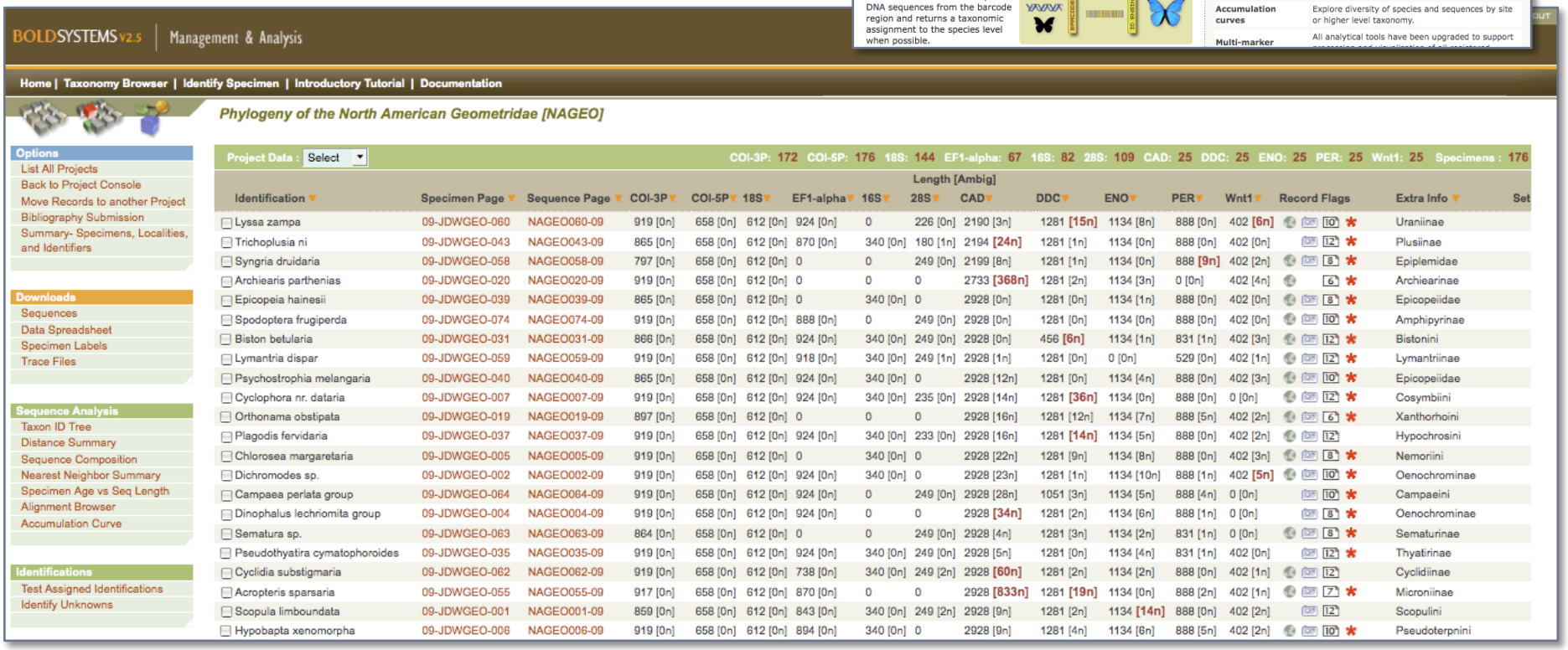


- outgroup (OG):
  - 6 families - 12 subfamilies - 24 genera
  - based on Regier et al 2009 & other studies

# Stored on BOLD

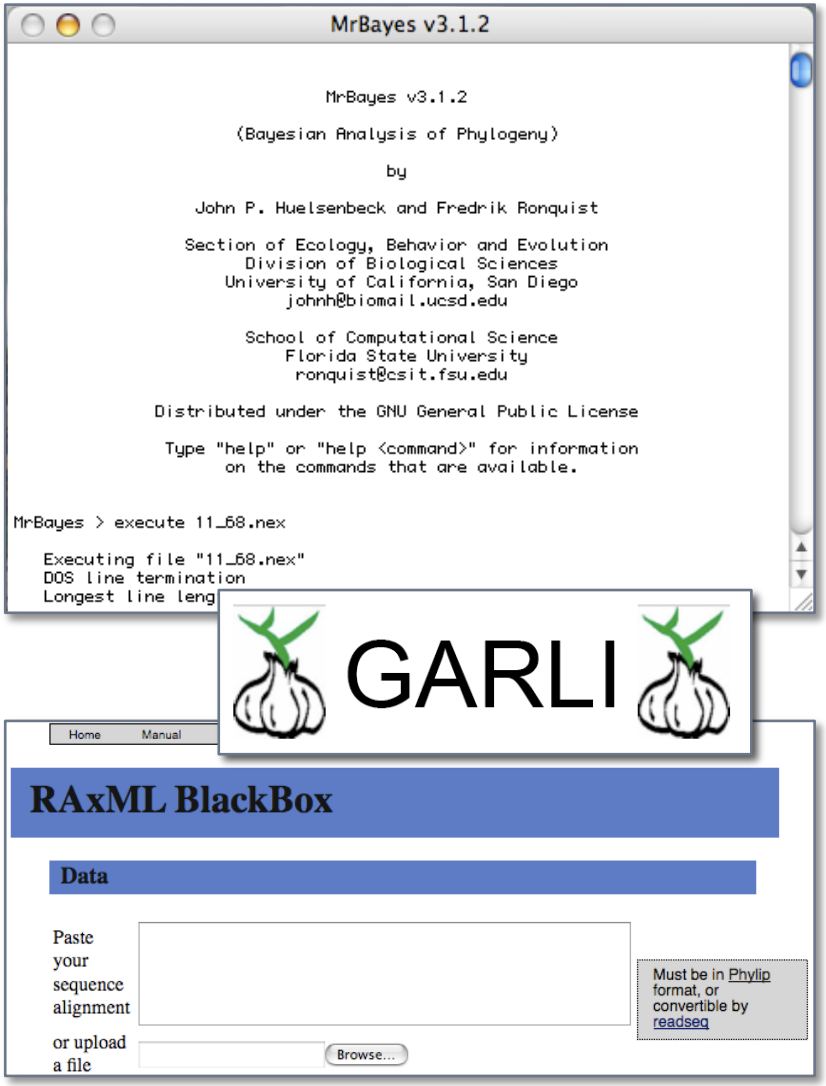
- all 11 genes, trace files, images, collateral data, etc.



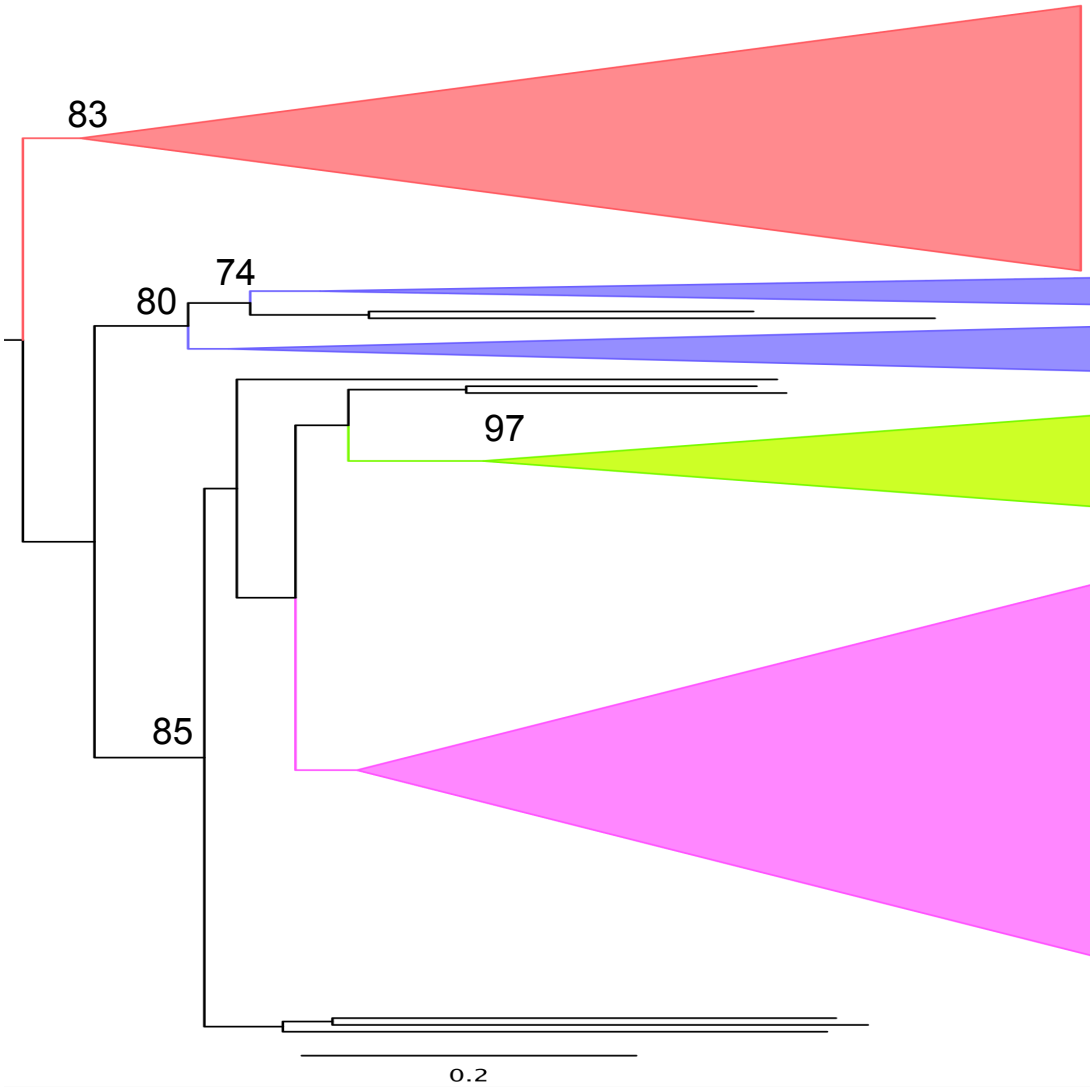



# Analyses completed

- Bayesian inference (BI)
  - MrBayes
- Maximum likelihood (ML)
  - Garli
  - RAxML (online)
- GTR+α+I / partition-specific models when possible
- on reduced dataset (RD) of 68 taxa as well



# ML tree – full dataset (RAxML)



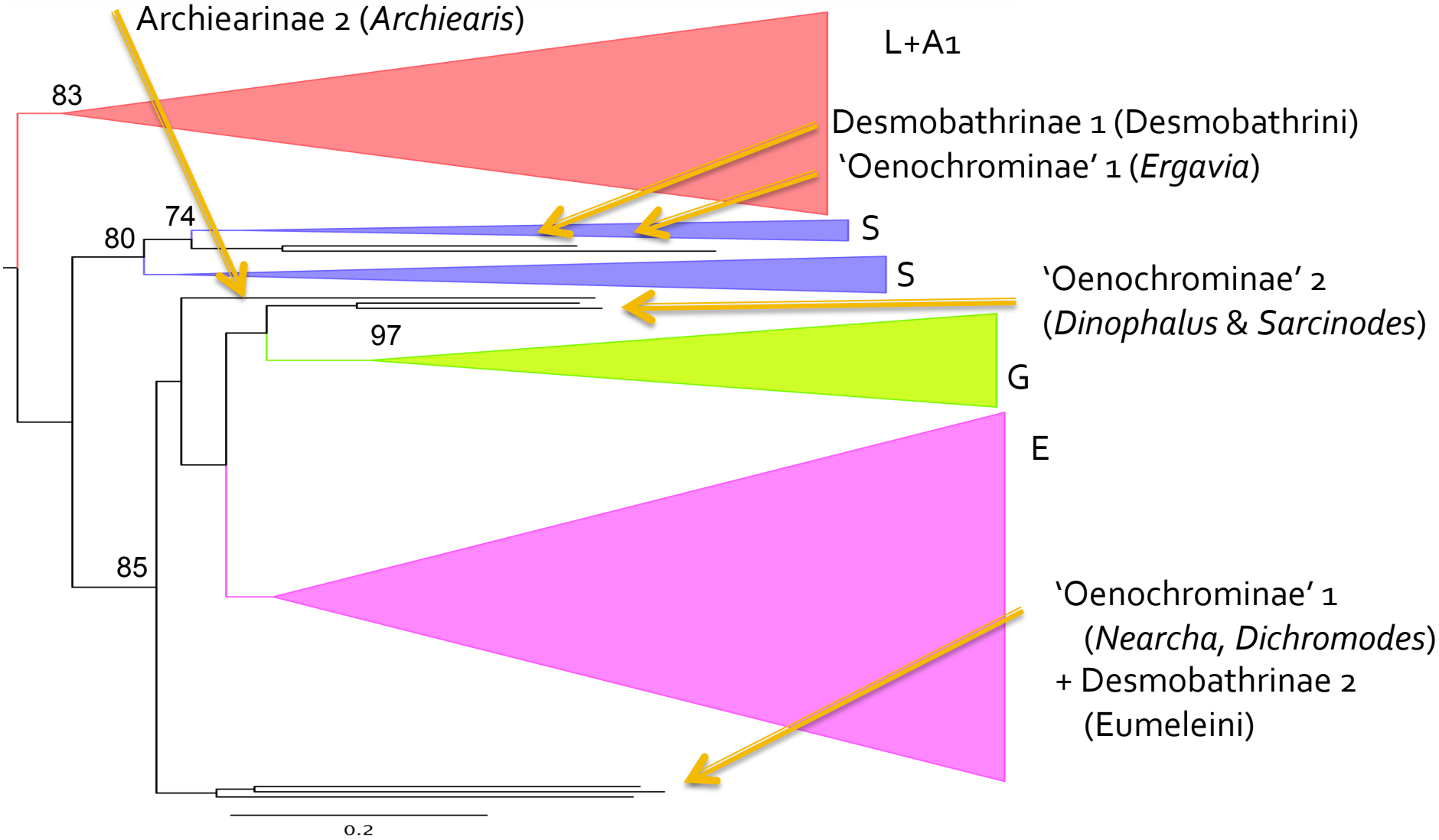

Larentiinae (39 genera/16 tribes)
+ Archiearinae 1 (*Lachnocephala*)

Sterrhinae (13/6)

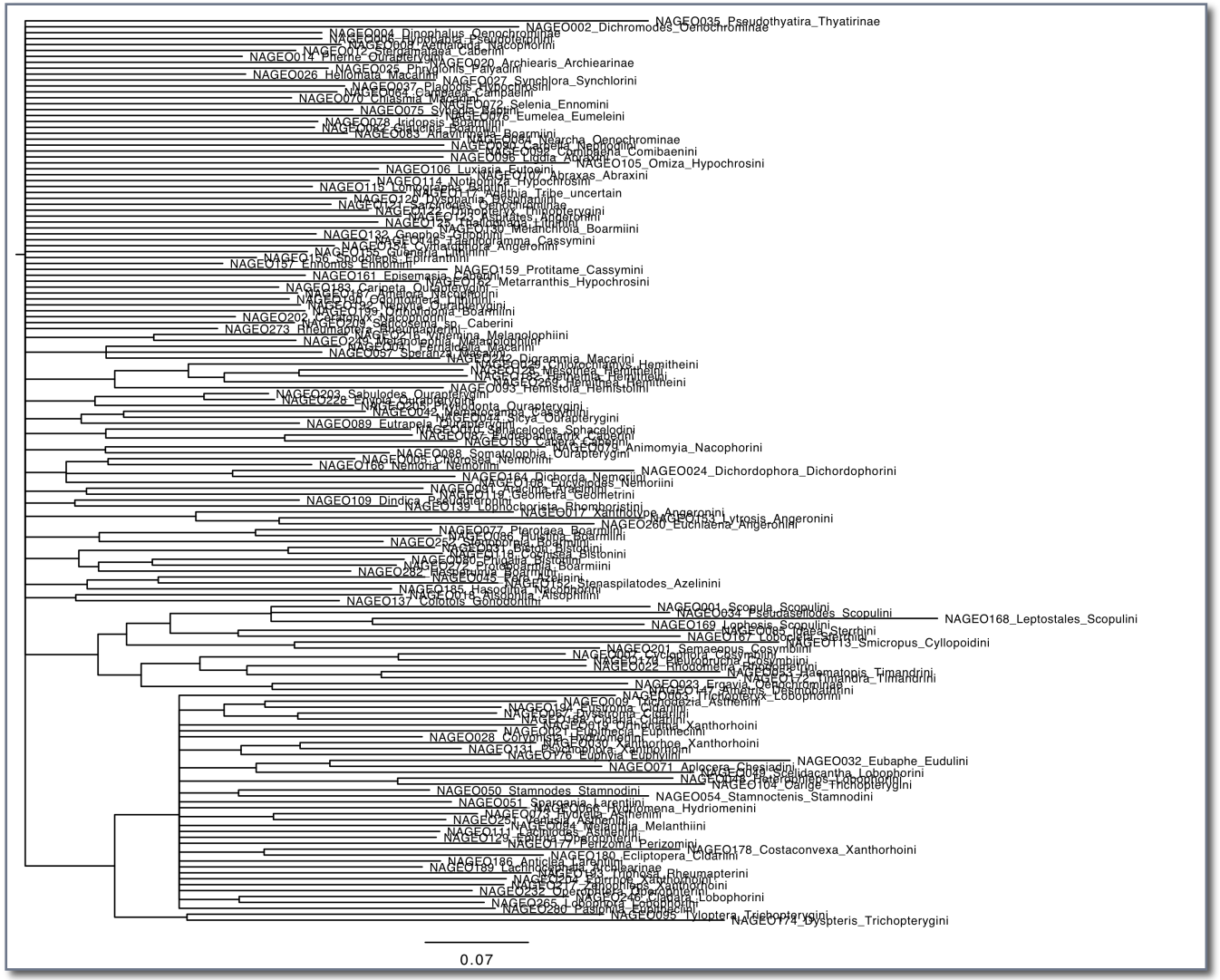
Geometrinae (19/12)

Ennominae (72/25)

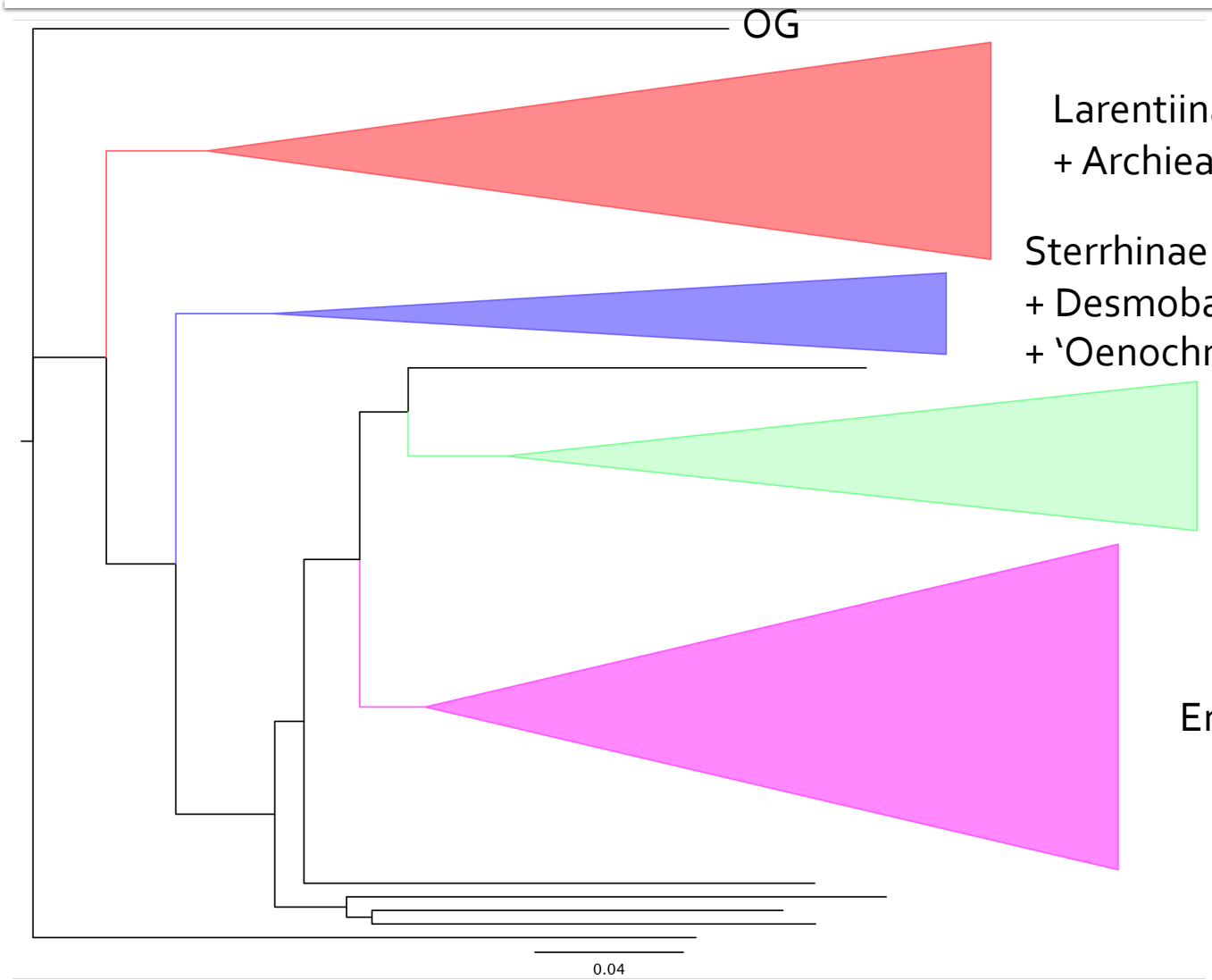
# ML tree – full dataset (RAxML)



# BI tree – full dataset



# ML tree – reduced dataset (Garli)



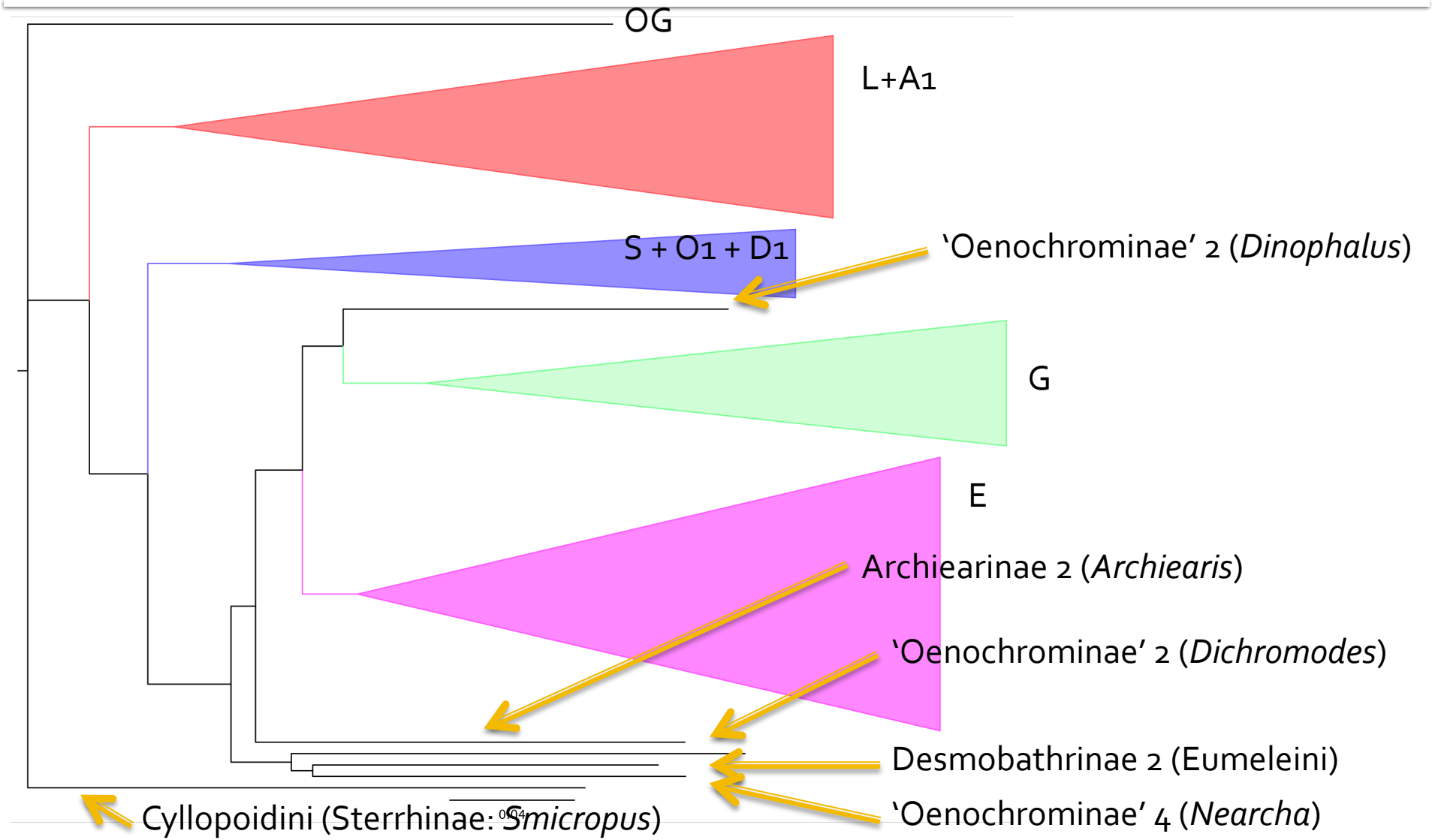

Larentiinae (16)
+ Archiearinae 1 (*Lachnocephala*)

Sterrhinae – *in part* (6)
+ Desmobathrinae 1 (Desmobathrini)
+ 'Oenochrominae' 1 (*Ergavia*)

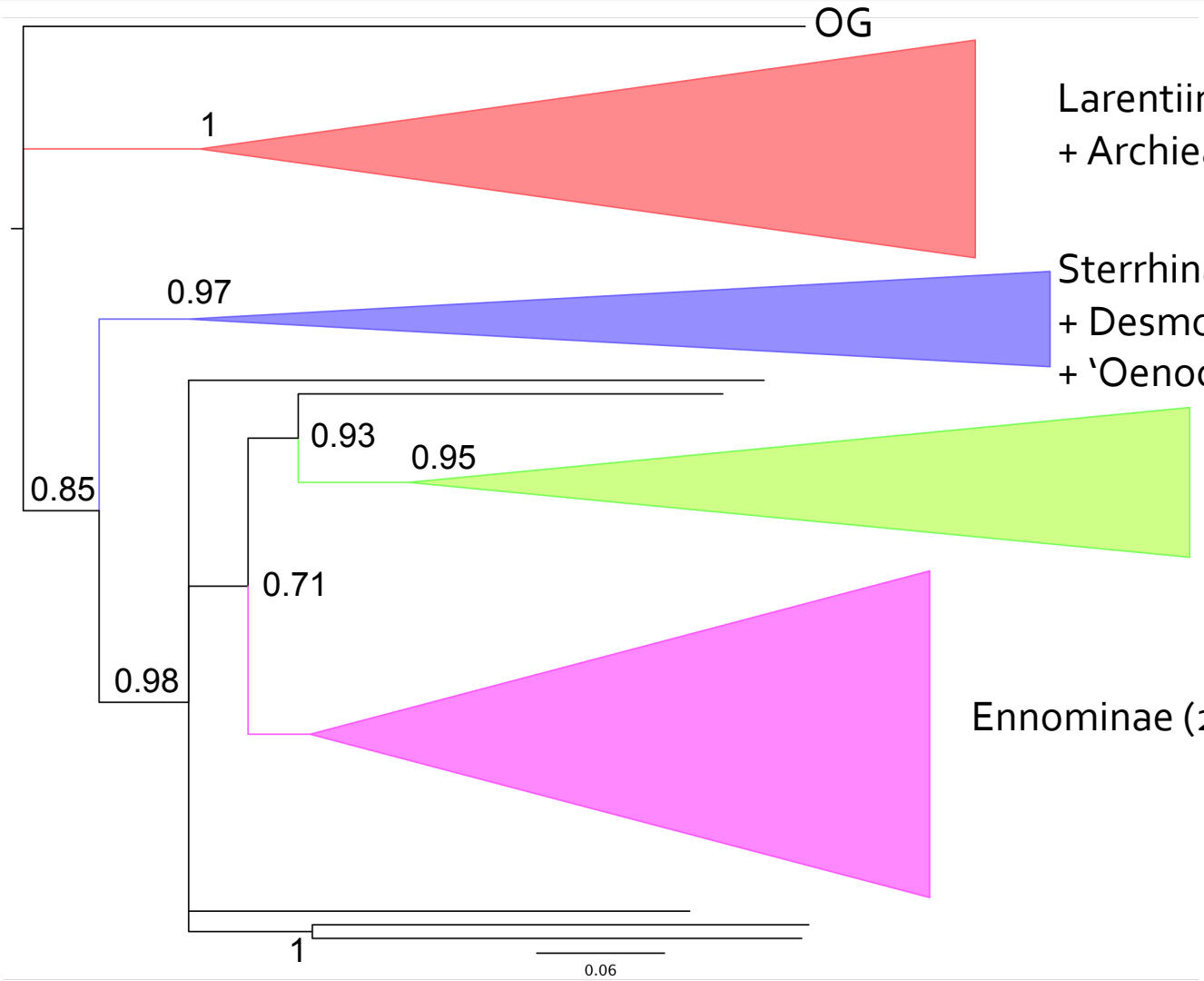
Geometrinae (12)

Ennominae (25)

# ML tree – reduced dataset (Garli)



# Bayesian analysis – reduced dataset



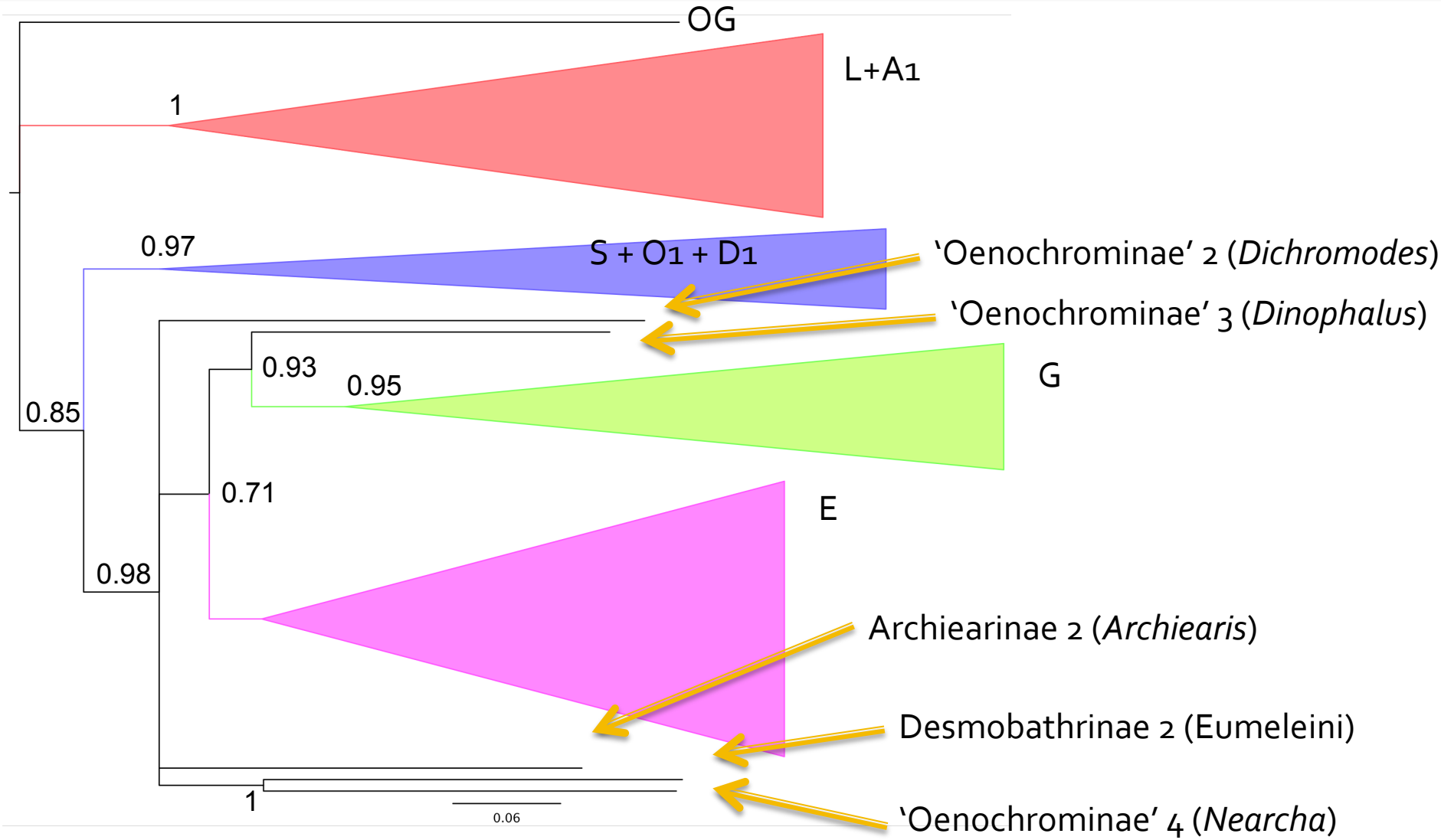

Larentiinae (16)
+ Archiearinae 1 (*Lachnocephala*)

Sterrhinae (6)
+ Desmobathrinae 1 (Desmobathrini)
+ 'Oenochrominae' 1 (*Ergavia*)

Geometrinae (12)

Ennominae (25)

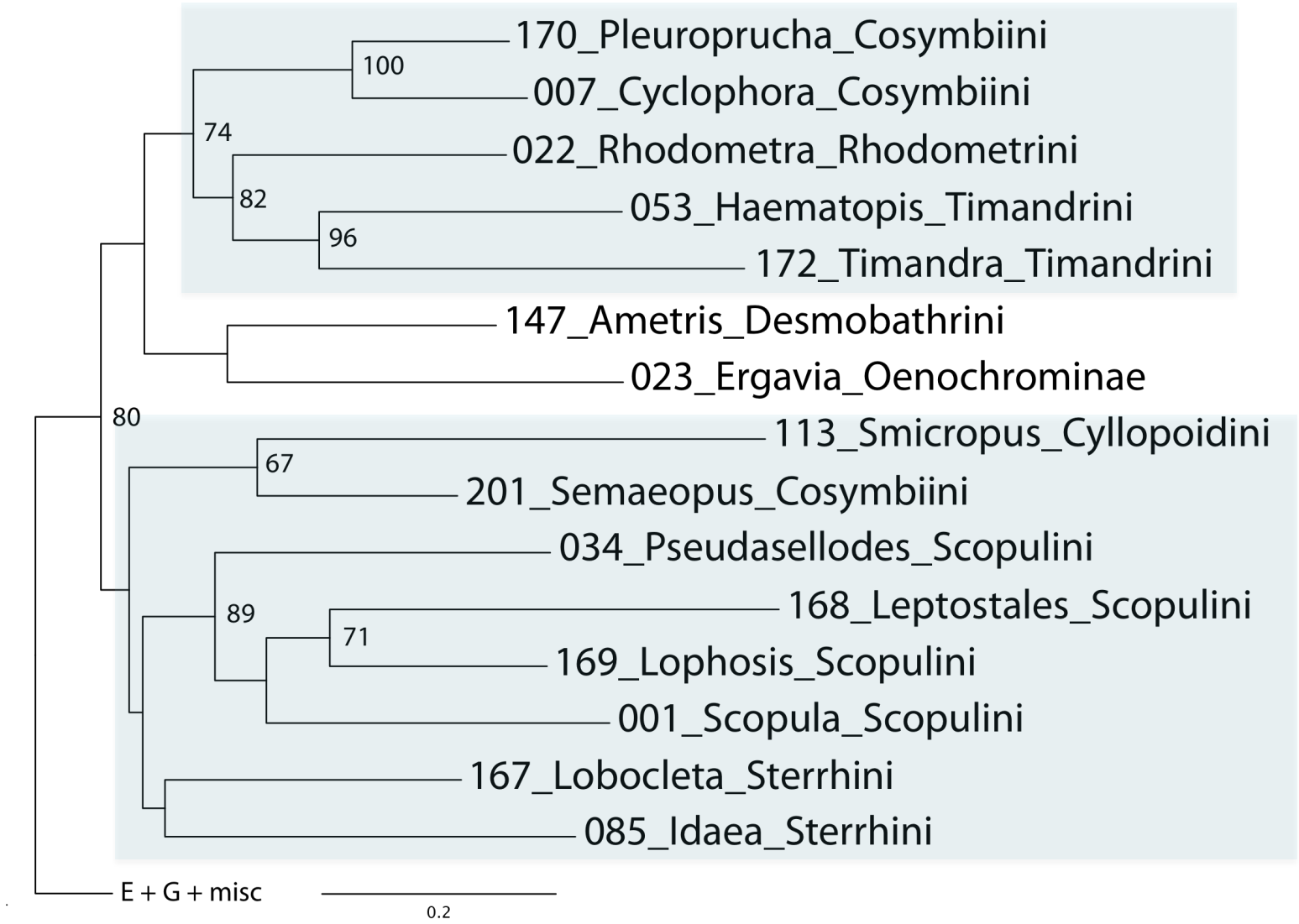
# Bayesian analysis – reduced dataset



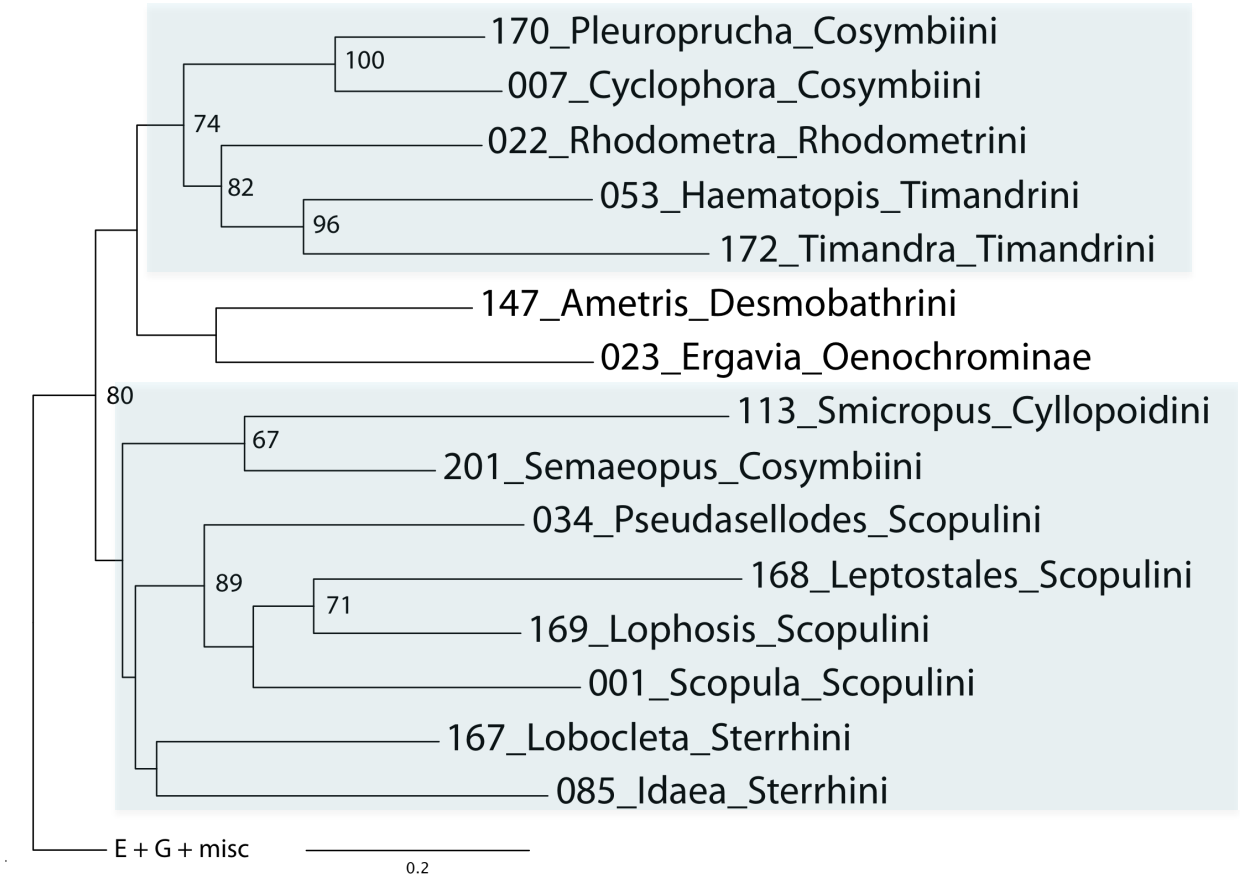
# Recovery of selected relationships

| | ML – full dataset | ML - reduced | BI - reduced |
|---|---|---|---|
| Larentiinae basal | ✓ | ✓ | ✓ |
| Larentiinae + Sterrhinae | ✗ | ✗ | ✗ |
| Ennominae monophyletic | ✓ | ✓ | ✓ |
| Sterrhinae monophyletic (inc. O1 + D1) | ✓ | ✗ | ✓ |
| (Ennominae + Geometrinae) + Archiearinae (*s. str.*) | ✓ | ✓ | ✓ |

# Close-up of Sterrhinae (ML-FD)



# Close-up of Sterrhinae (ML-FD)




Sihvonen & Kaila 2004:

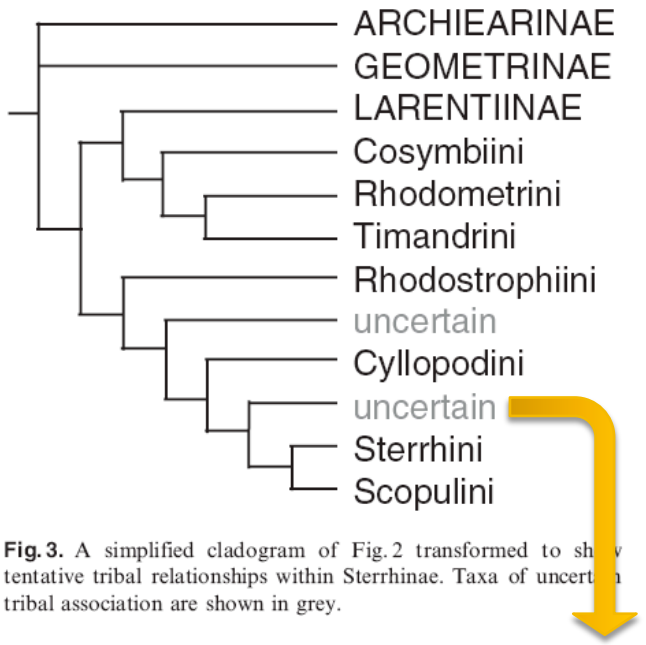



**Fig. 3.** A simplified cladogram of Fig. 2 transformed to sh[illegible] tentative tribal relationships within Sterrhinae. Taxa of uncert[illegible] tribal association are shown in grey.

Includes *Semaeopus*

# Other interesting relationships…

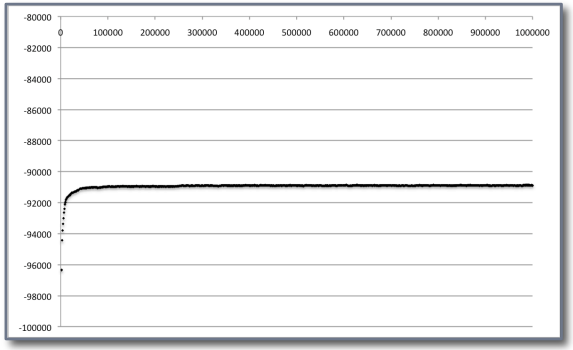
- Alsophilini + Colotoini + Campaeini as in Yamamota & Sota 2007, Wahlberg et al. 2010

- *Lachnocephala* (Neotropical Archiearinae) – Young (2006) suggested affinities with Tasmanian Archiearinae, which are 'probably misplaced'; perhaps they sit within Larentiinae – BOLD-ID supports this as well

- *Ametris* (Desmobathrini) and *Eumelea (Eumeleini)* analyzed for the first time, suggesting Desmobathrinae nonmonophyly

# Next up for analysis

- Continue long runs until convergence




- Further partitioning of data (e.g. codon position, nuclear, mitochondrial)

- Tests of significance for predicted relationships
  - AU (Shimodaira 2002) tests

# Next up for analysis

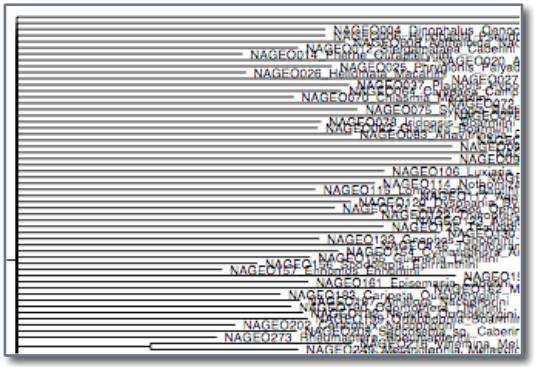
- explore the outgroups
  - root of Geometridae and monophyly of Geometroidea (recently questioned)
- Analyze genes separately and investigate congruence/conflict
  - Partition addition bootstrap alteration (PABA) tests (Struck et al. 2008)
- explore the the effect of missing characters in the matrix

Urania (Uraniidae)

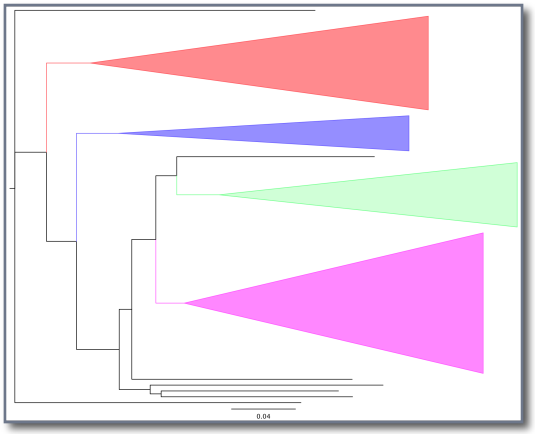
*Epicopeia* (Epicopeiidae)

# Preliminary conclusions

- relatively low branch support + a few discrepancies between methods and datasets = poor resolution
  - poor choice in gene regions?
  - missing data obscuring benefits of increased taxon sampling?



- good support for some instances of nonmonophyly
  - increased sampling of minor subfamilies *before* tackling the majors



# Preliminary conclusions

- Jury is still out on 'barcode approach' of combining molecular datasets
  - could use to 'glue' this and previous studies from Australia, Japan, Northern Europe, etc. for global perspective – and still date tree

# Acknowledgements


- People:
  - Sujeevan Ratnasingham, Stephanie Kirk, Paul Hebert & staff of the CCDB (U. of Guelph)
  - Kim Mitter (UMD)
  - Dan Janzen (UPenn) & Chris Schmidt (CNC)
- Funding:
  - Canadian Barcode of Life Network [Genome Canada, Ontario Genomics Institute, NSERC]
  - NSERC Graduate Scholarship Program
  - Canadian Forest Service – Pacific
  - Forest Investment Account – Forest Science Program