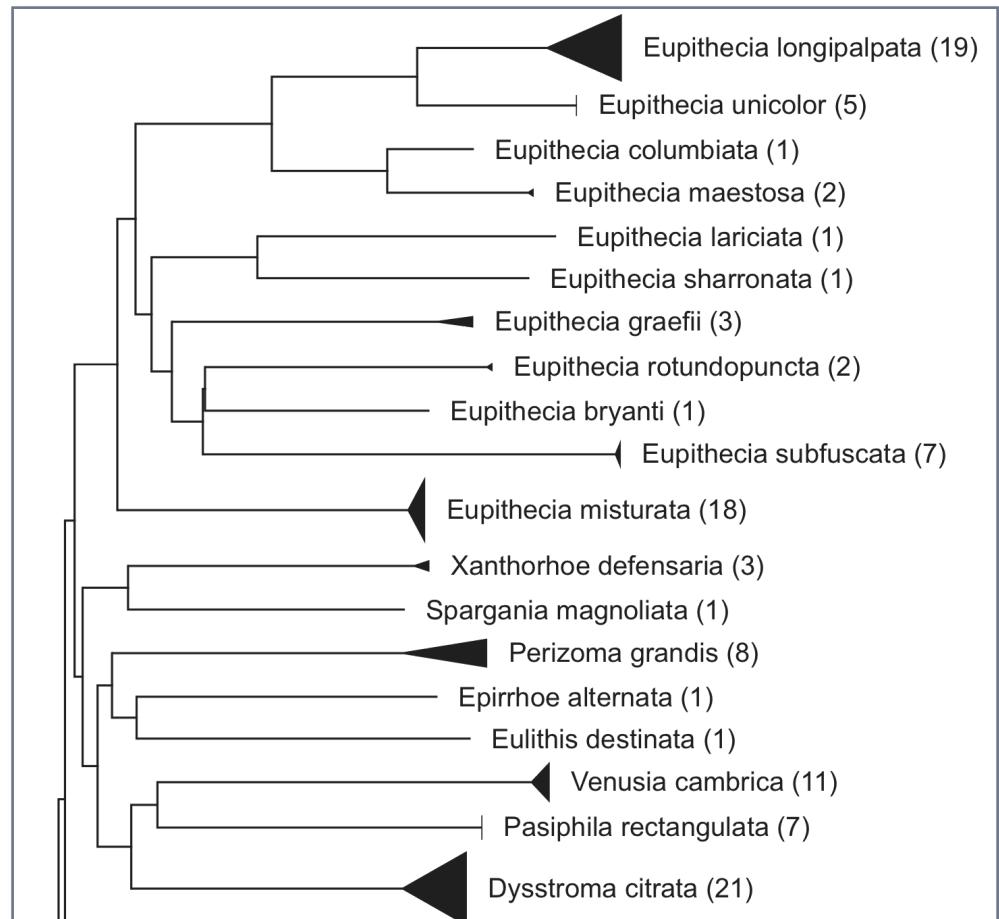


# The 'barcode approach' to combining molecular datasets: Reconstructing the phylogeny of North America's Geometridae as an example

Jeremy deWaard, Charlie Mitter,  
Axel Hausmann and Leland Humble

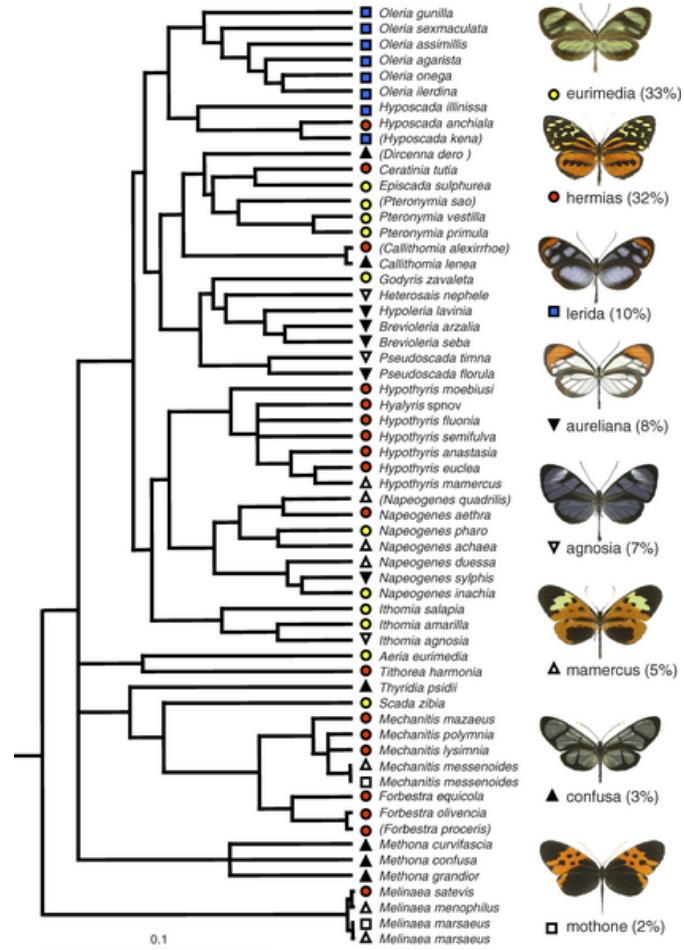
# A little background...

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



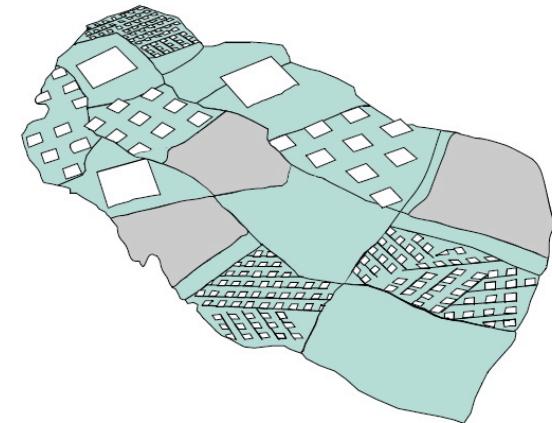
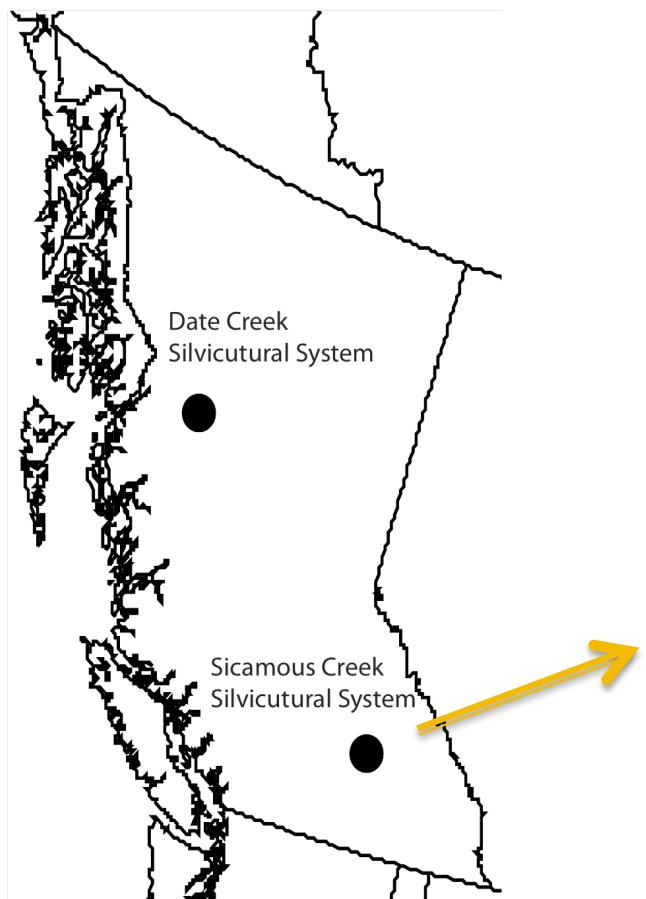
COI tree used to sort, identify and inventory nocturnal Lepidoptera in Stanley Park, Vancouver (Fig.1 (in part), deWaard et al 2009).

# Application of phylogenies



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

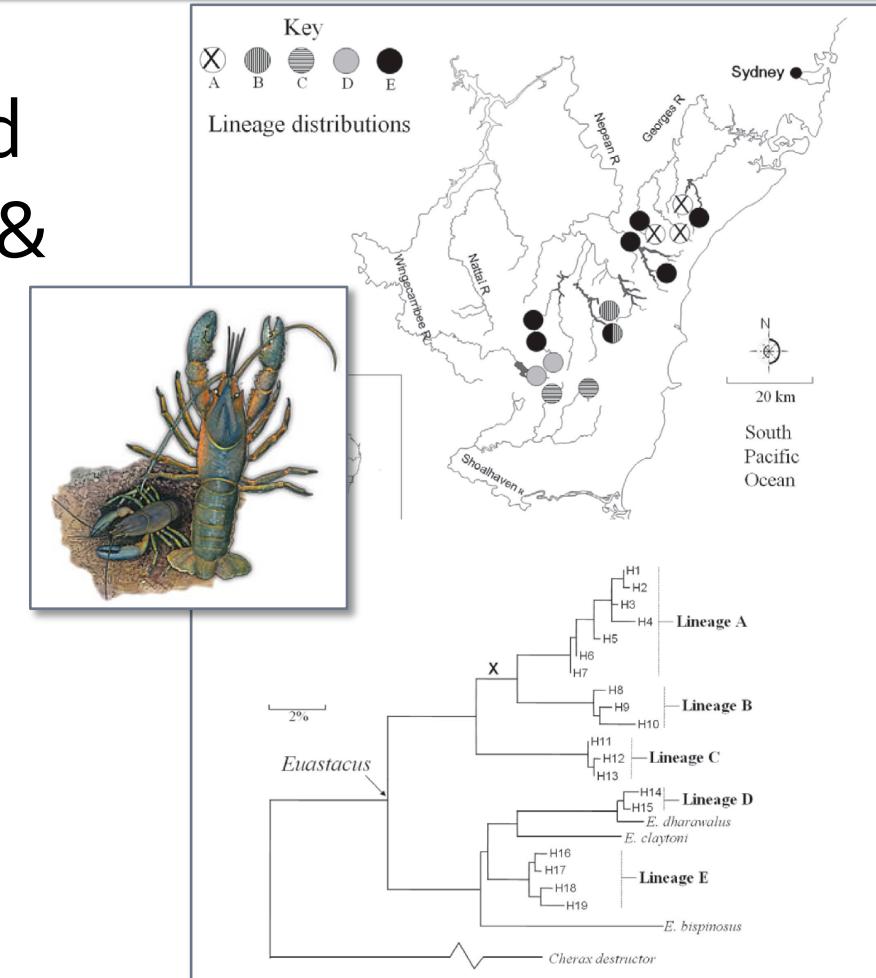
# The effect of harvest types on multiple levels of moth diversity



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

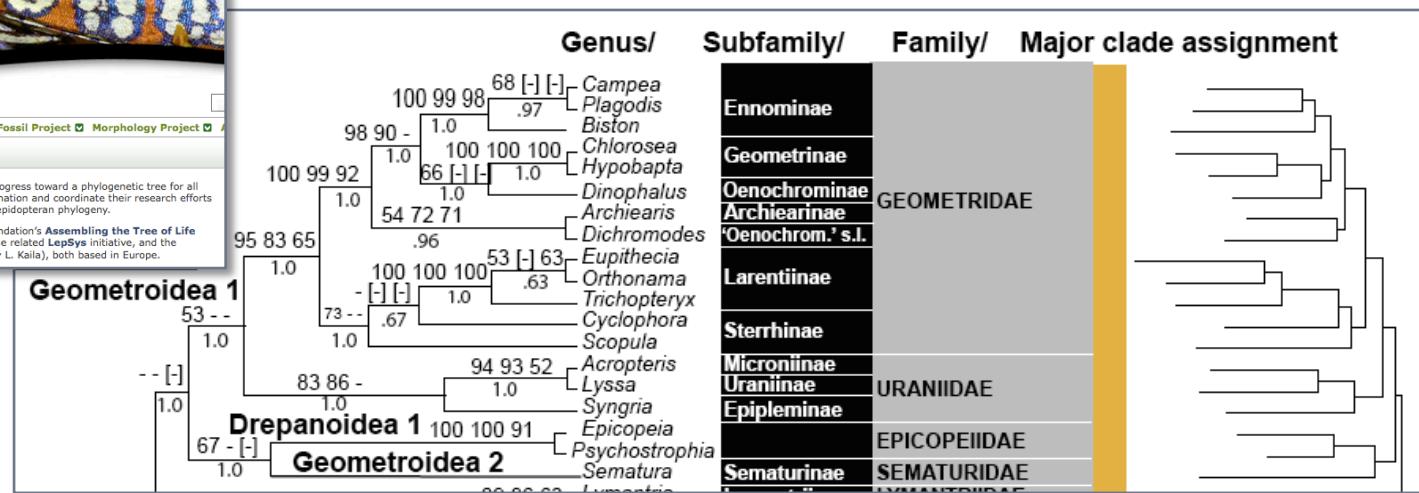
- multi-gene, tribe-level phylogeny of NA geometrids:  
scaffold to constrain COI trees at BC collection sites
  - contribute a North American perspective to deep geometrid systematics

The collage consists of five separate journal covers arranged in a grid-like fashion:

- Top Left:** Molecular Phylogenetics and Evolution (Elsevier).
  - Available online at [www.sciencedirect.com](http://www.sciencedirect.com)
  - ScienceDirect logo
  - Molecular Phylogenetics and Evolution xxx (2007) xxx-xxx
  - MOLECULAR PHYLOGENETICS AND EVOLUTION
  - [www.elsevier.com/locate/ympev](http://www.elsevier.com/locate/ympev)
  - Phylogeny of the Geometridae and the evolution of winter moths inferred from a simultaneous analysis of mitochondrial and nuclear genes
  - Satoshi Yamamoto \*, Teiji Sota
  - Molecular Phylogenetics and Evolution 55 (2010) 929-938
  - Contents lists available at ScienceDirect
  - Molecular Phylogenetics and Evolution journal homepage: [www.elsevier.com/locate/ympev](http://www.elsevier.com/locate/ympev)
- Top Middle:** Zootaxa (Magnolia Press).
  - Zootaxa 1264: 1–147 (2006)
  - [www.mapress.com/zootaxa/](http://www.mapress.com/zootaxa/)
  - Copyright © 2006 Magnolia Press
  - ISSN 1175-5326 (print edition)
  - ZOOTAXA**
  - ISSN 1175-5334 (online edition)
  - (1264)
- Bottom Left:** Molecular Phylogenetics and Evolution (Elsevier).
  - The evolution of female flightlessness among Ennominae of the Holarctic forest zone (Lepidoptera, Geometridae)
  - Niklas Wahlberg <sup>a,\*</sup>, Niina Snäll <sup>a,b</sup>, Jaan Viidalepp <sup>c</sup>, Kai Ruohomäki <sup>b</sup>, Toomas Tammaru <sup>d</sup>
- Bottom Middle:** Molecular Relationships of the Australian Ennominae (Lepidoptera: Geometridae) and implications for the phylogeny of the Geometridae from molecular and morphological data (Catherine J. Young).
  - CATHERINE J. YOUNG
- Bottom Right:** Systematic position of Lythriini revised: transferred from Larentiinae to Sterrhinae (Lepidoptera, Geometridae) (Erki Õunap, Jaan Viidalepp & Urmas Saarma).
  - Systematic position of Lythriini revised: transferred from Larentiinae to Sterrhinae (Lepidoptera, Geometridae)
  - ERKI ÕUNAP, JAAN VIIDALEPP & URMAS SAARMA

# LepTree collaboration

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



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

# Reconstructing the smaller branches and tips

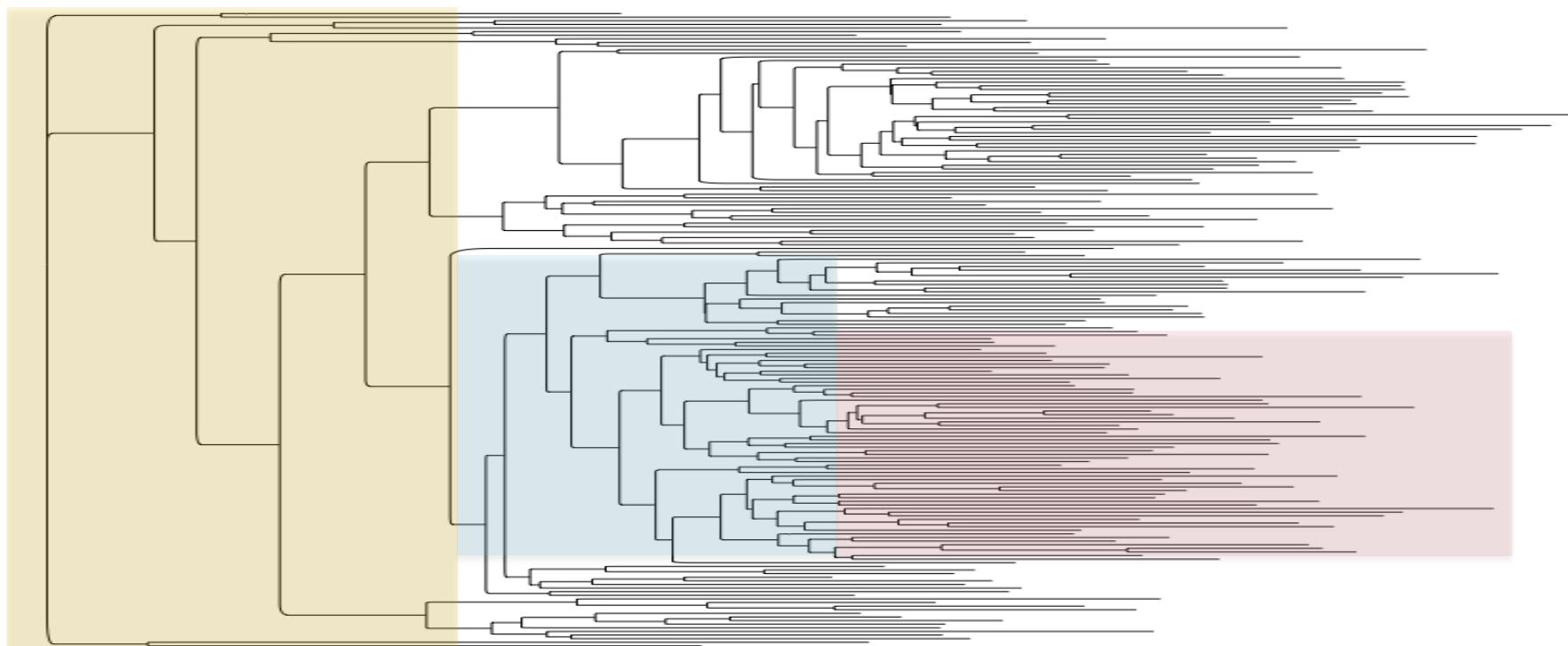
deep

mid

shallow



decrease in the availability of recently collected exemplar taxa



# Reconstructing the smaller branches and tips

deep

mid

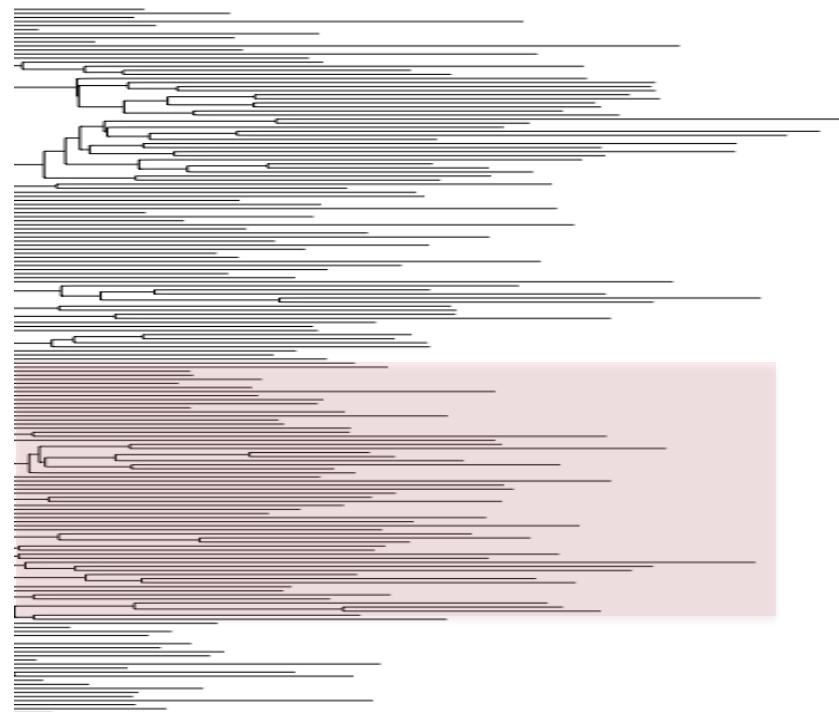
shallow



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

deep

mid

shallow

nuclear PCGs

COI LepTree taxa

COI new taxa

# Sampling strategy

deep

mid

shallow

nuclear PCGs



new

COI

LepTree taxa

If we work with barcoded specimens, COI can 'glue' deep LepTree dataset with new dataset, even if no other genes overlap

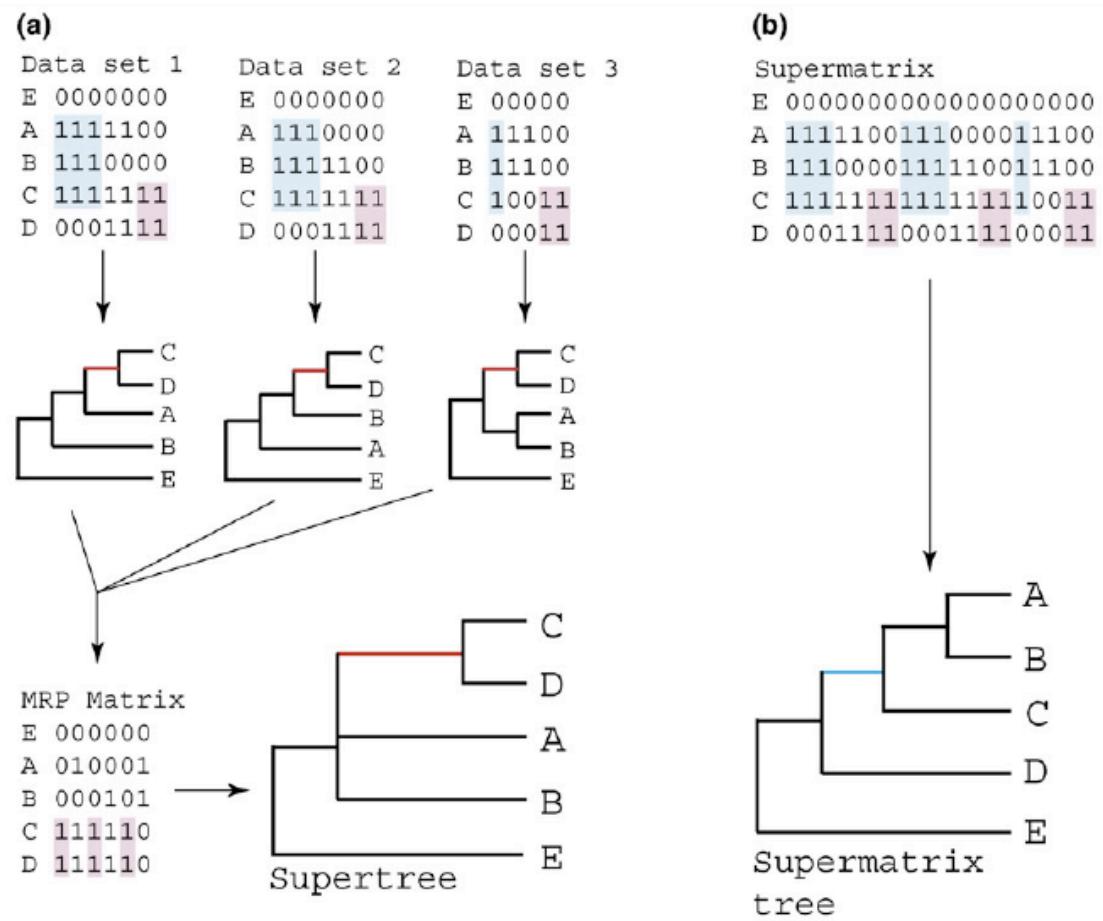
dataset

COI

new taxa

# 'Barcode approach' for combining molecular datasets

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



TRENDS in Ecology & Evolution

(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

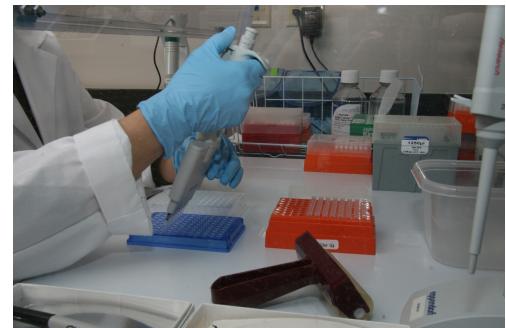


## Specimen Depositories:

- University of Maryland (112)
- Biodiversity Institute of Ontario (25)
- Canadian National Collection of Insects, Arachnids and Nematodes (15)
- Research Collection of Jeremy deWaard (9)
- Pacific Forestry Centre, Canadian Forest Service (8)
- Zoological State Collection, Munich (2)
- Area de Conservacion Guanacaste (2)
- Strickland Museum of Entomology, University of Alberta (1)
- Smithsonian Institution (1)
- James Entomological Collection, Washington State University (1)

# Laboratory methods

- Canadian Centre of DNA Barcoding in Guelph
  - 15 primer sets (+ repeated COI-5p)
    - COI-3p-COII
    - 16S
    - 18S
    - 28S-D2
    - EF-1 $\alpha$  (4)
    - Wingless
    - CAD (4)
    - DDC
    - IDH



# Laboratory methods

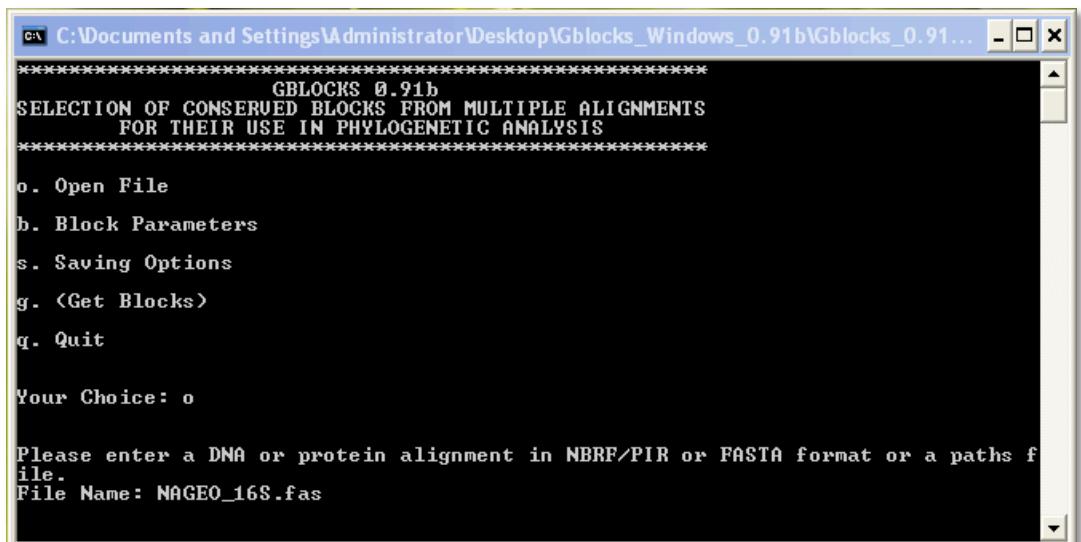
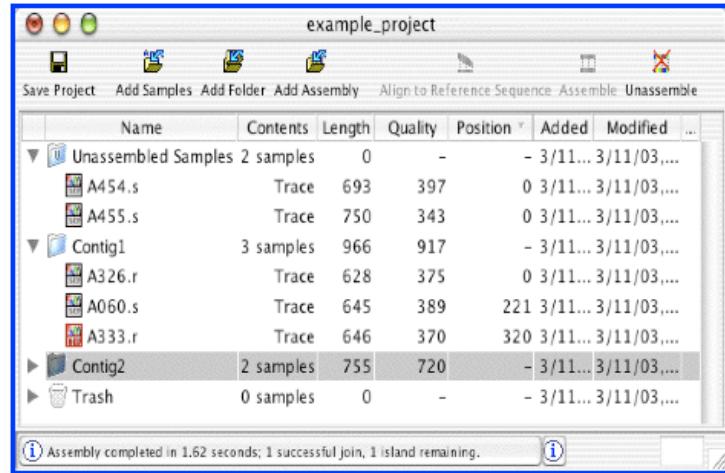
- Canadian Centre of DNA Barcoding in Guelph

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



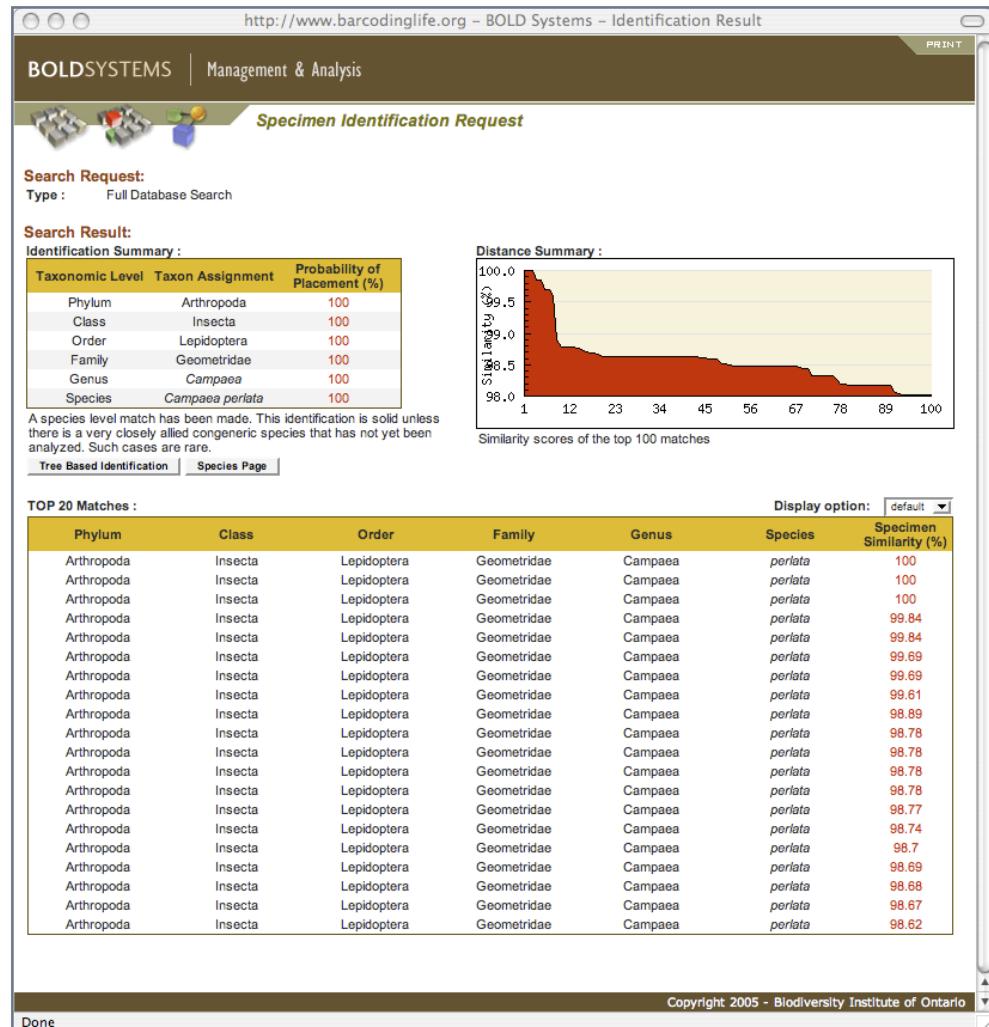
# Alignments

- 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

deep

# CAD, DDC, Period, Enolase, Wingless

6633bp

mid

## COI-3p-COII, 16S, 18S, 28S, EF-1 $\alpha$

3702bp

shallow

COI

## 25 LepTree taxa

COI

151 new taxa

# Final dataset

deep

mid

shallow

CAD, DDC, Period,  
Enolase, Wingless

6633bp

176 genera

10,335 bp matrix

62.7% coverage

$\bar{x} = 3613 \text{ bp / taxon}$

COI-3p-COII, 16S,  
18S, 28S, EF-1 $\alpha$

3702bp

COI

25 LepTree taxa

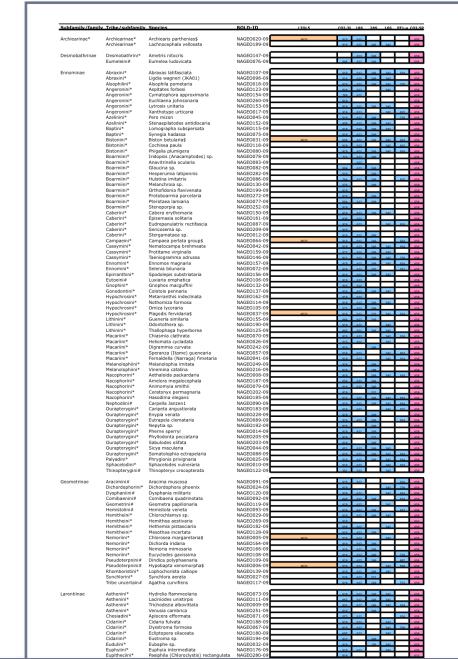
COI

151 new taxa

658bp

# 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



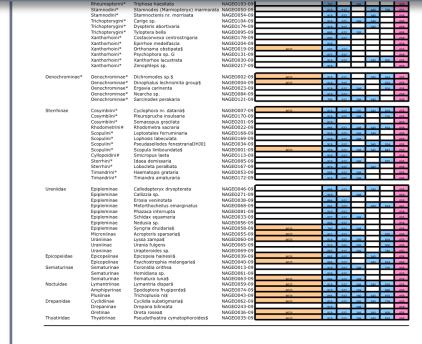
Subfamily/family	Tribe/subfamily	Species	BOLD-ID	LTOL5	COI-3P	18S	28S	16S	EF1-a	COI-5P
Archiearininae*	Archiearininae*	Archiearis parthenias\$	NAGEO020-09	6633	919	612				658
	Archiearininae*	Lachnocephala vellosata	NAGEO189-09		919	612	249	340		658
Desmobathrinae	Desmobathrini*	Ametris nitocris	NAGEO147-09			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

					866	612		340		658
Uraniidae	Epipleminae	Calledapteryx dryopterata	NAGEO046-09							
	Epipleminae	Callizzia sp.	NAGEO271-09		919		249			658
	Epipleminae	Erosia veninotata	NAGEO038-09		864	612				658
	Epipleminae	Metorthochelus 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 oritheia	NAGEO013-09		919	612	249		726	658
	Sematurinae	Homidiiana sp.	NAGEO061-09		919	612				658
Noctuidae	Sematurinae	Sematura luna\$	NAGEO063-09	6633	864	612	249			658
	Lymantriinae	Lymantria dispar\$	NAGEO059-09	6633	919	612	249	340	918	658
	Amphipyryinae	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.

**BARCODE OF LIFE DATA SYSTEMS v2.5**  
Advancing species identification and discovery through the analysis of short, standardized gene regions

[About BOLD](#) [Contact Us](#)



[Published Projects](#) | [Taxonomy Browser](#) | [Request an Account](#) | [Identify Specimen](#) | [Introductory Tutorial](#) | [Documentation](#) | [Data Release](#) | [Citation](#)

**BARCODE COUNTS**

Formally Described Species With Barcodes	73,314
Total Barcode Records	894,666
Source	
GenBank	93,258
Canadian Centre	735,203
Others	66,205
<b>Breakdown</b>	

**BOLDSYSTEMS BOLD 2.5 Release**

Version 2.5, unveiled on Nov 11th 2009 at the Third International Barcoding of Life conference in Mexico City, provides new core functionality including support for multiple sequence markers per specimen and more complex workflows. Features include identification services for ITS, matK, and rbcL markers; comparative analytics; web services and a variety of convenience upgrades. A few are highlighted here:

**Management & Analysis**

BOLD-MAS provides a repository for barcode records coupled with analytical tools. It serves as an online workbench for the DNA barcode community.

**Identification Engine**

BOLD-IDS provides a species identification tool that accepts DNA sequences from the barcode region and returns a taxonomic assignment to the species level when possible.

**Accumulation curves** Explore diversity of species and sequences by site or higher level taxonomy.

**Multi-marker** All analytical tools have been upgraded to support

**BOLDSYSTEMS v2.5** | Management & Analysis

[Home](#) | [Taxonomy Browser](#) | [Identify Specimen](#) | [Introductory Tutorial](#) | [Documentation](#)

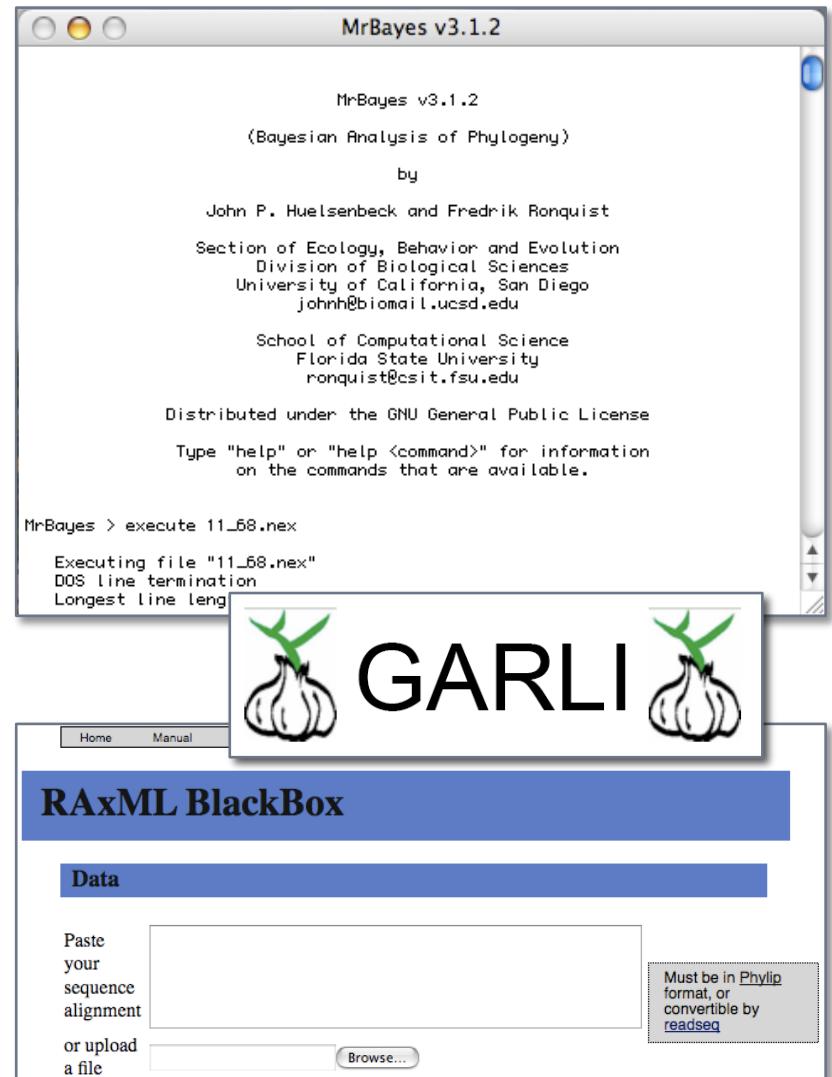


**Phylogeny of the North American Geometridae [NAGEO]**

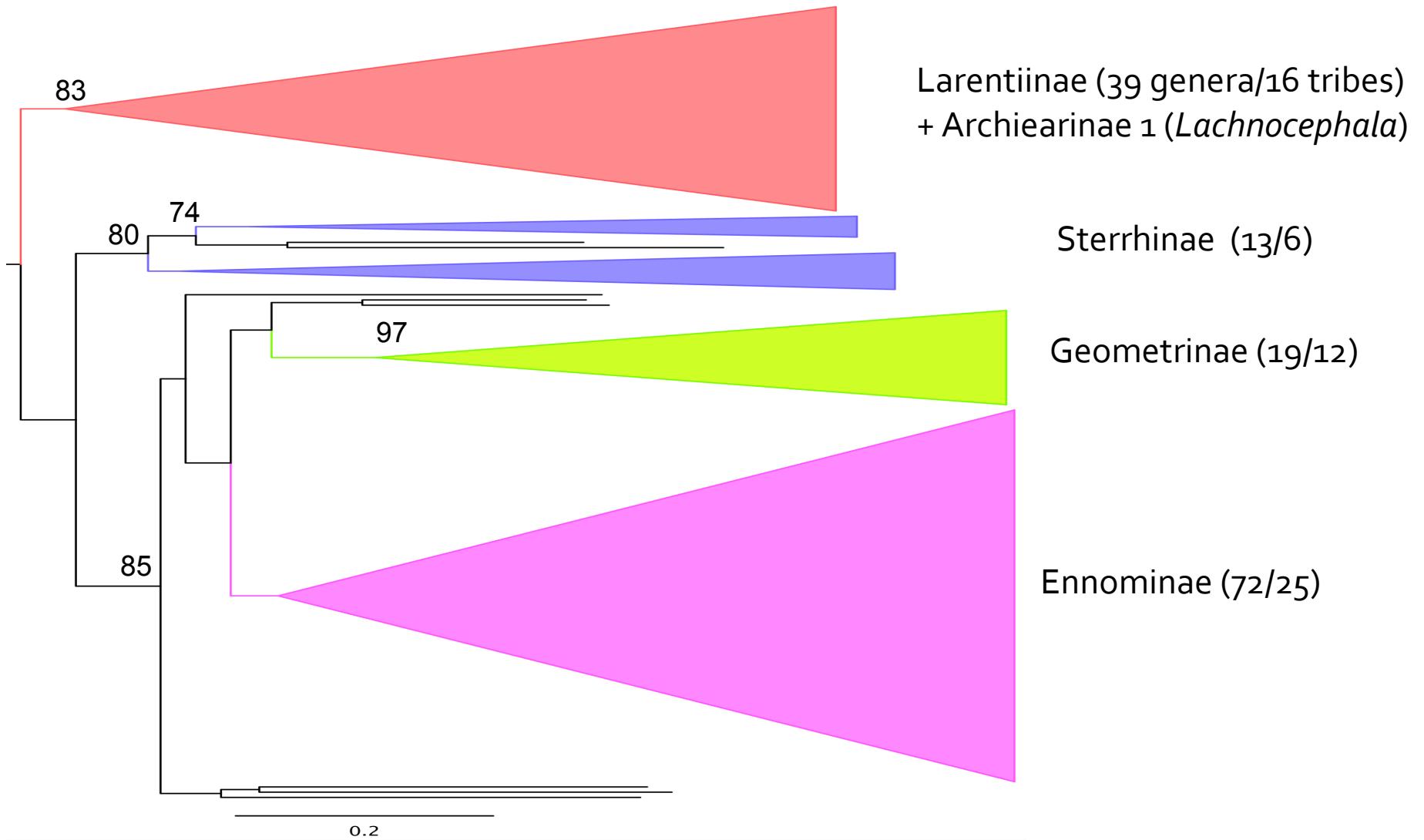
Project Data : Select		COI-3P: 172	COI-5P: 176	18S: 144	EF1-alpha: 67	16S: 82	28S: 109	CAD: 25	DDC: 25	ENO: 25	PER: 25	Wnt1: 25	Specimens : 176			
Identification	Specimen Page	Sequence Page	COI-3P▼	COI-5P▼	18S▼	EF1-alpha▼	16S▼	28S▼	CAD▼	DDC▼	ENO▼	PER▼	Wnt1▼	Record Flags	Extra Info ▼	Set
<input type="checkbox"/> Lyssa zampa	09-JDWGEO-060	NAGEO0060-09	919 [0n]	658 [0n]	612 [0n]	924 [0n]	0	226 [0n]	2190 [3n]	1281 [15n]	1134 [8n]	888 [0n]	402 [6n]	 	Uraniiinae	
<input type="checkbox"/> Trichoplusia ni	09-JDWGEO-043	NAGEO0043-09	865 [0n]	658 [0n]	612 [0n]	870 [0n]	340 [0n]	180 [1n]	2194 [24n]	1281 [1n]	1134 [0n]	888 [0n]	402 [0n]	 	Plusiinae	
<input type="checkbox"/> Syngria druidaria	09-JDWGEO-058	NAGEO0058-09	797 [0n]	658 [0n]	612 [0n]	0	0	249 [0n]	2199 [8n]	1281 [1n]	1134 [0n]	888 [9n]	402 [2n]	 	Epiplemidae	
<input type="checkbox"/> Archiearis parthenias	09-JDWGEO-020	NAGEO0020-09	919 [0n]	658 [0n]	612 [0n]	0	0	2733 [368n]	1281 [2n]	1134 [3n]	0 [0n]	402 [4n]	 	Archiearinae		
<input type="checkbox"/> Epicopeia hainesii	09-JDWGEO-039	NAGEO0039-09	865 [0n]	658 [0n]	612 [0n]	0	340 [0n]	0	2928 [0n]	1281 [0n]	1134 [1n]	888 [0n]	402 [0n]	 	Epicopeiidae	
<input type="checkbox"/> Spodoptera frugiperda	09-JDWGEO-074	NAGEO0074-09	919 [0n]	658 [0n]	612 [0n]	888 [0n]	0	249 [0n]	2928 [0n]	1281 [0n]	1134 [0n]	888 [0n]	402 [0n]	 	Amphipyrinae	
<input type="checkbox"/> Biston betularia	09-JDWGEO-031	NAGEO0031-09	866 [0n]	658 [0n]	612 [0n]	924 [0n]	340 [0n]	249 [0n]	2928 [0n]	456 [6n]	1134 [1n]	831 [1n]	402 [3n]	 	Bistonini	
<input type="checkbox"/> Lymantria dispar	09-JDWGEO-059	NAGEO0059-09	919 [0n]	658 [0n]	612 [0n]	918 [0n]	340 [0n]	249 [1n]	2928 [1n]	1281 [0n]	0 [0n]	529 [0n]	402 [1n]	 	Lymantriinae	
<input type="checkbox"/> Psychostrophia melangaria	09-JDWGEO-040	NAGEO0040-09	865 [0n]	658 [0n]	612 [0n]	924 [0n]	340 [0n]	0	2928 [12n]	1281 [0n]	1134 [4n]	888 [0n]	402 [3n]	 	Epicopeidae	
<input type="checkbox"/> Cyclophora nr. dataria	09-JDWGEO-007	NAGEO0007-09	919 [0n]	658 [0n]	612 [0n]	924 [0n]	340 [0n]	235 [0n]	2928 [14n]	1281 [36n]	1134 [0n]	888 [0n]	402 [0n]	 	Cosymbiini	
<input type="checkbox"/> Orthonauma obstipata	09-JDWGEO-019	NAGEO0019-09	897 [0n]	658 [0n]	612 [0n]	0	0	0	2928 [16n]	1281 [12n]	1134 [7n]	888 [5n]	402 [2n]	 	Xanthorniini	
<input type="checkbox"/> Plagodis fervidaria	09-JDWGEO-037	NAGEO0037-09	919 [0n]	658 [0n]	612 [0n]	924 [0n]	340 [0n]	233 [0n]	2928 [16n]	1281 [14n]	1134 [5n]	888 [0n]	402 [2n]	 	Hypochrosini	
<input type="checkbox"/> Chlorosea margaretaria	09-JDWGEO-005	NAGEO0005-09	919 [0n]	658 [0n]	612 [0n]	0	340 [0n]	0	2928 [22n]	1281 [9n]	1134 [8n]	888 [0n]	402 [3n]	 	Nemorini	
<input type="checkbox"/> Dichromodes sp.	09-JDWGEO-002	NAGEO0002-09	919 [0n]	658 [0n]	612 [0n]	924 [0n]	340 [0n]	0	2928 [23n]	1281 [1n]	1134 [10n]	888 [1n]	402 [5n]	 	Oenochrominae	
<input type="checkbox"/> Campea perlata group	09-JDWGEO-064	NAGEO0064-09	919 [0n]	658 [0n]	612 [0n]	924 [0n]	0	249 [0n]	2928 [28n]	1051 [3n]	1134 [5n]	888 [4n]	0 [0n]	 	Campaeini	
<input type="checkbox"/> Dinophalus lechriomita group	09-JDWGEO-004	NAGEO0004-09	919 [0n]	658 [0n]	612 [0n]	924 [0n]	0	0	2928 [34n]	1281 [2n]	1134 [6n]	888 [1n]	0 [0n]	 	Oenochrominae	
<input type="checkbox"/> Sematura sp.	09-JDWGEO-063	NAGEO0063-09	864 [0n]	658 [0n]	612 [0n]	0	0	249 [0n]	2928 [4n]	1281 [3n]	1134 [2n]	831 [1n]	0 [0n]	 	Sematurinae	
<input type="checkbox"/> Pseudothyatira cymatophoroidea	09-JDWGEO-035	NAGEO0035-09	919 [0n]	658 [0n]	612 [0n]	924 [0n]	340 [0n]	249 [0n]	2928 [5n]	1281 [0n]	1134 [4n]	831 [1n]	402 [0n]	 	Thyatirinae	
<input type="checkbox"/> Cyclidia substigmaria	09-JDWGEO-062	NAGEO0062-09	919 [0n]	658 [0n]	612 [0n]	738 [0n]	340 [0n]	249 [2n]	2928 [60n]	1281 [2n]	1134 [2n]	888 [0n]	402 [1n]	 	Cyclidiinae	
<input type="checkbox"/> Acropteris sparsaria	09-JDWGEO-055	NAGEO0055-09	917 [0n]	658 [0n]	612 [0n]	870 [0n]	0	0	2928 [833n]	1281 [19n]	1134 [0n]	888 [2n]	402 [1n]	 	Microniinae	
<input type="checkbox"/> Scopula limbounbindata	09-JDWGEO-001	NAGEO0001-09	859 [0n]	658 [0n]	612 [0n]	843 [0n]	340 [0n]	249 [2n]	2928 [9n]	1281 [2n]	1134 [14n]	888 [0n]	402 [2n]	 	Scopulini	
<input type="checkbox"/> Hypobapta xenomorpha	09-JDWGEO-006	NAGEO0006-09	919 [0n]	658 [0n]	612 [0n]	894 [0n]	340 [0n]	0	2928 [9n]	1281 [4n]	1134 [8n]	888 [5n]	402 [2n]	 	Pseudoterpini	

# Analyses completed

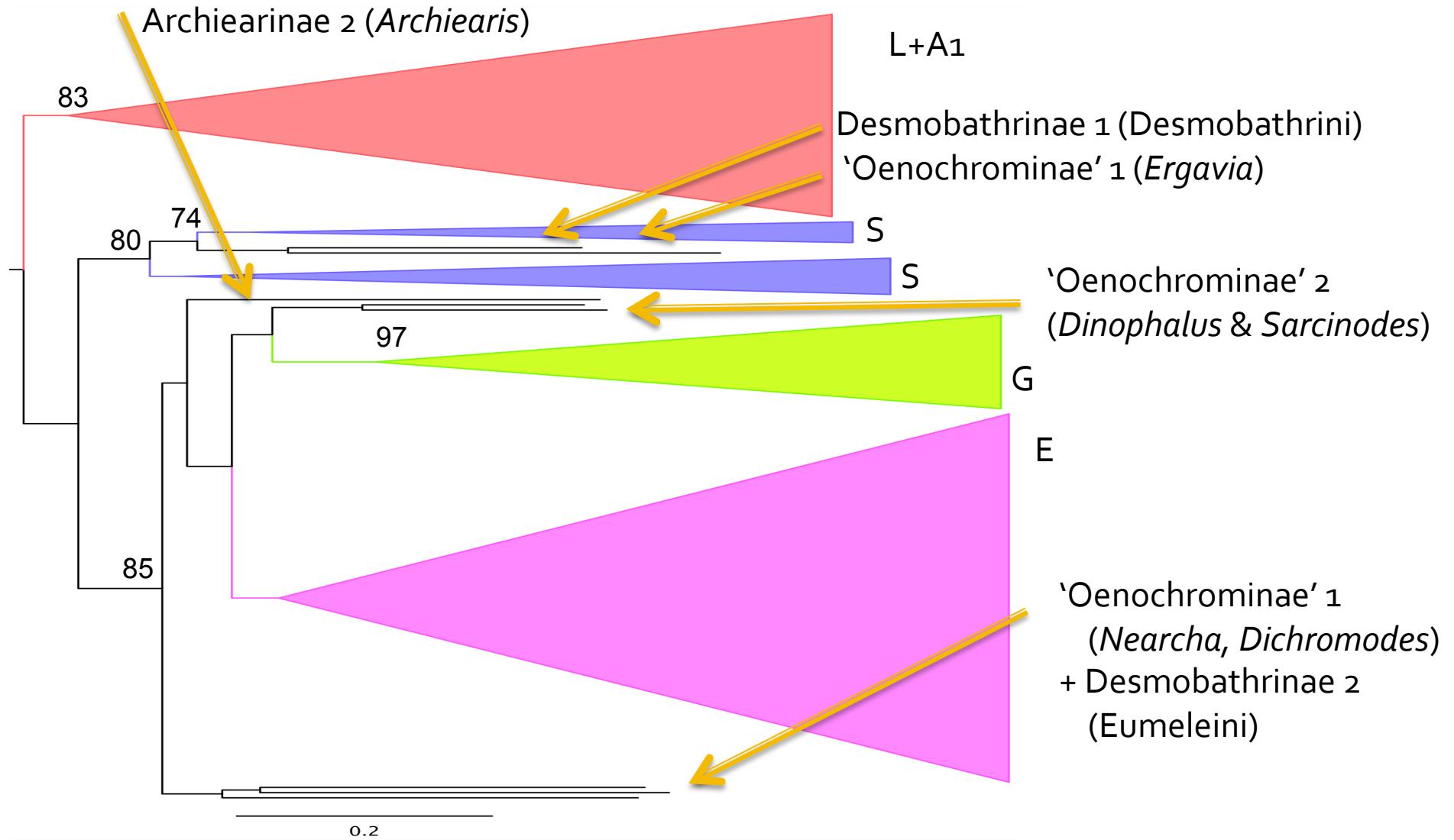
- Bayesian inference (BI)
  - MrBayes
- Maximum likelihood (ML)
  - Garli
  - RAxML (online)
- GTR+ $\alpha$ +I / partition-specific models when possible
- on reduced dataset (RD) of 68 taxa as well



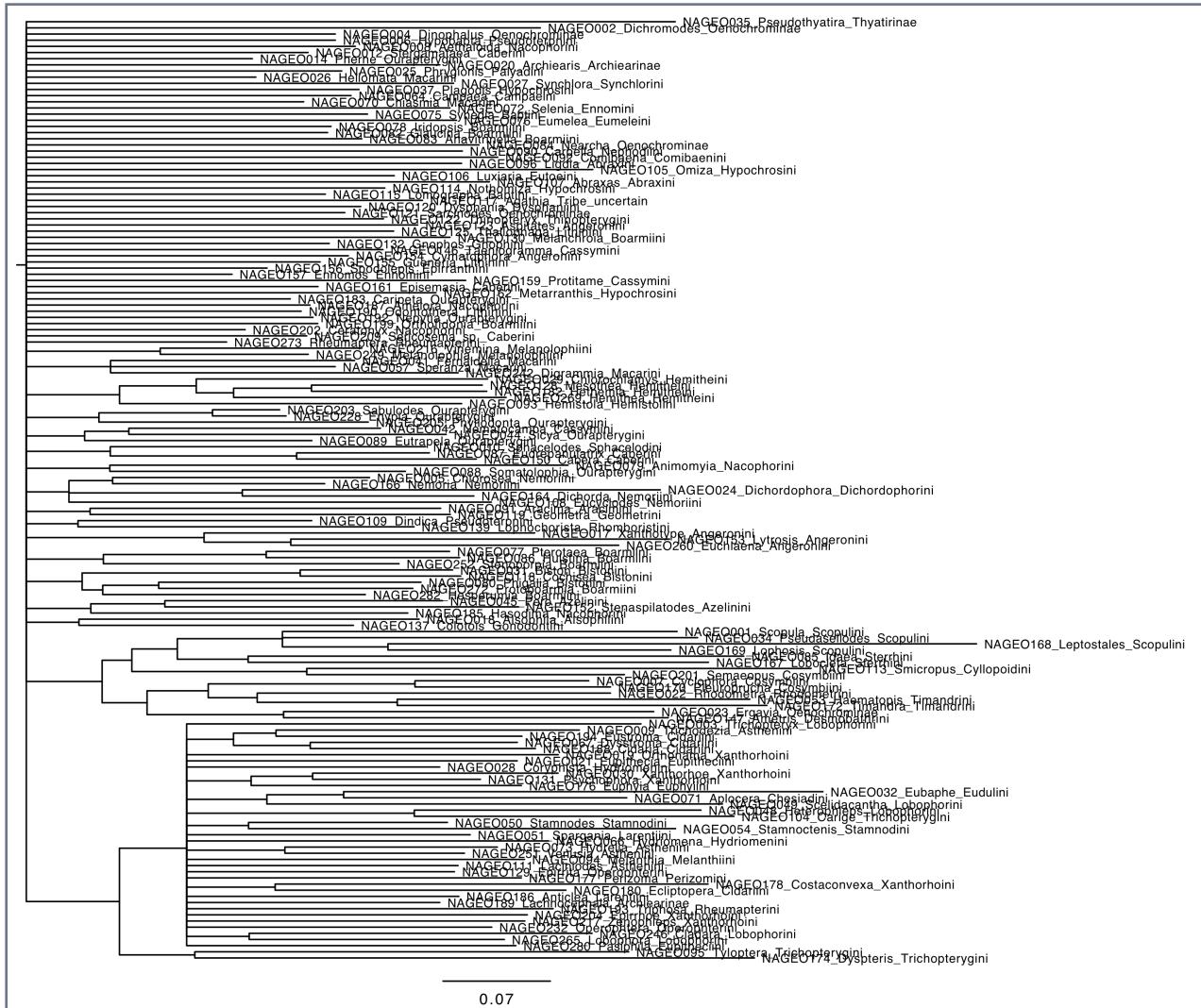
# ML tree – full dataset (RAxML)



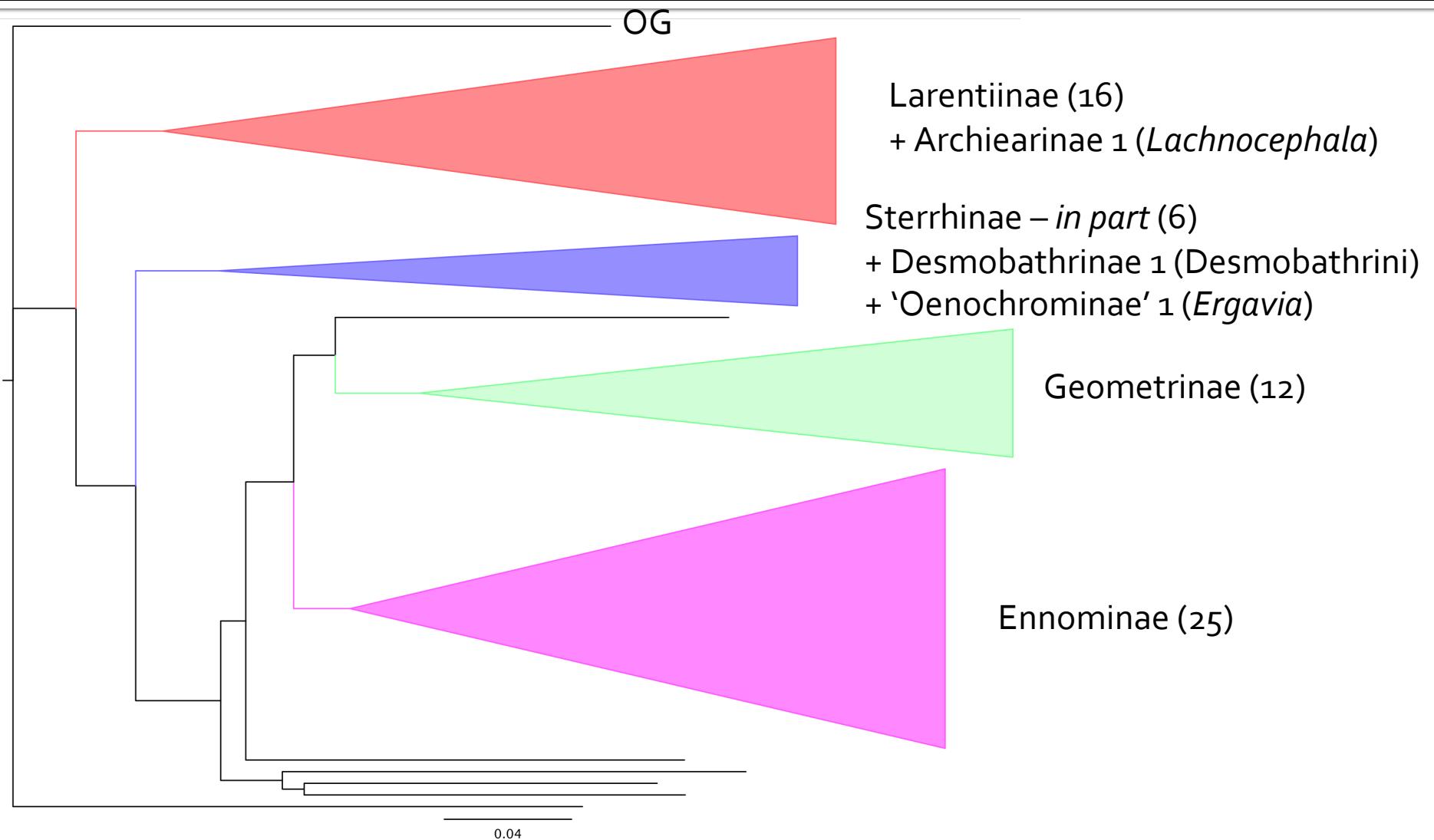
# ML tree – full dataset (RAxML)



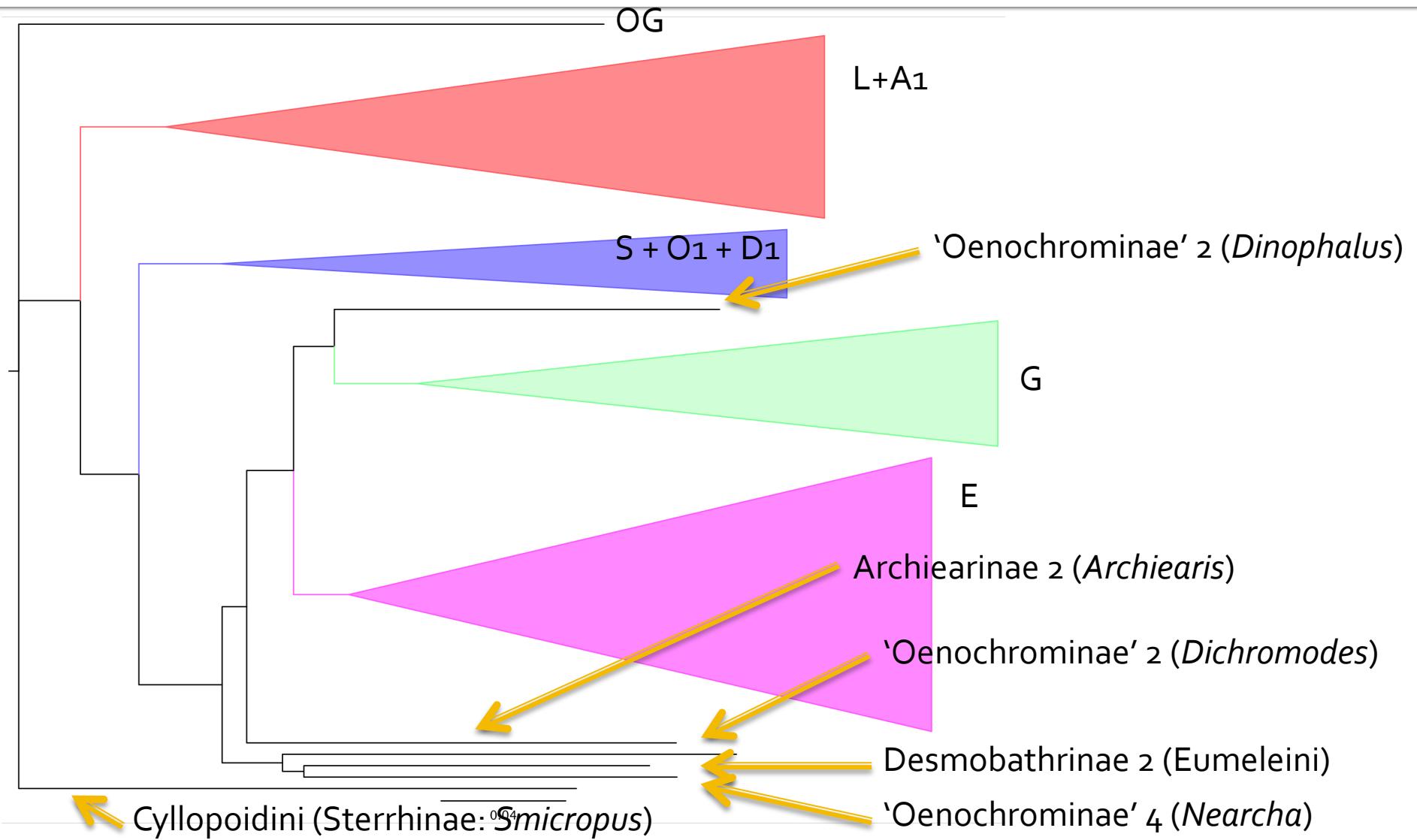
# BI tree - full dataset



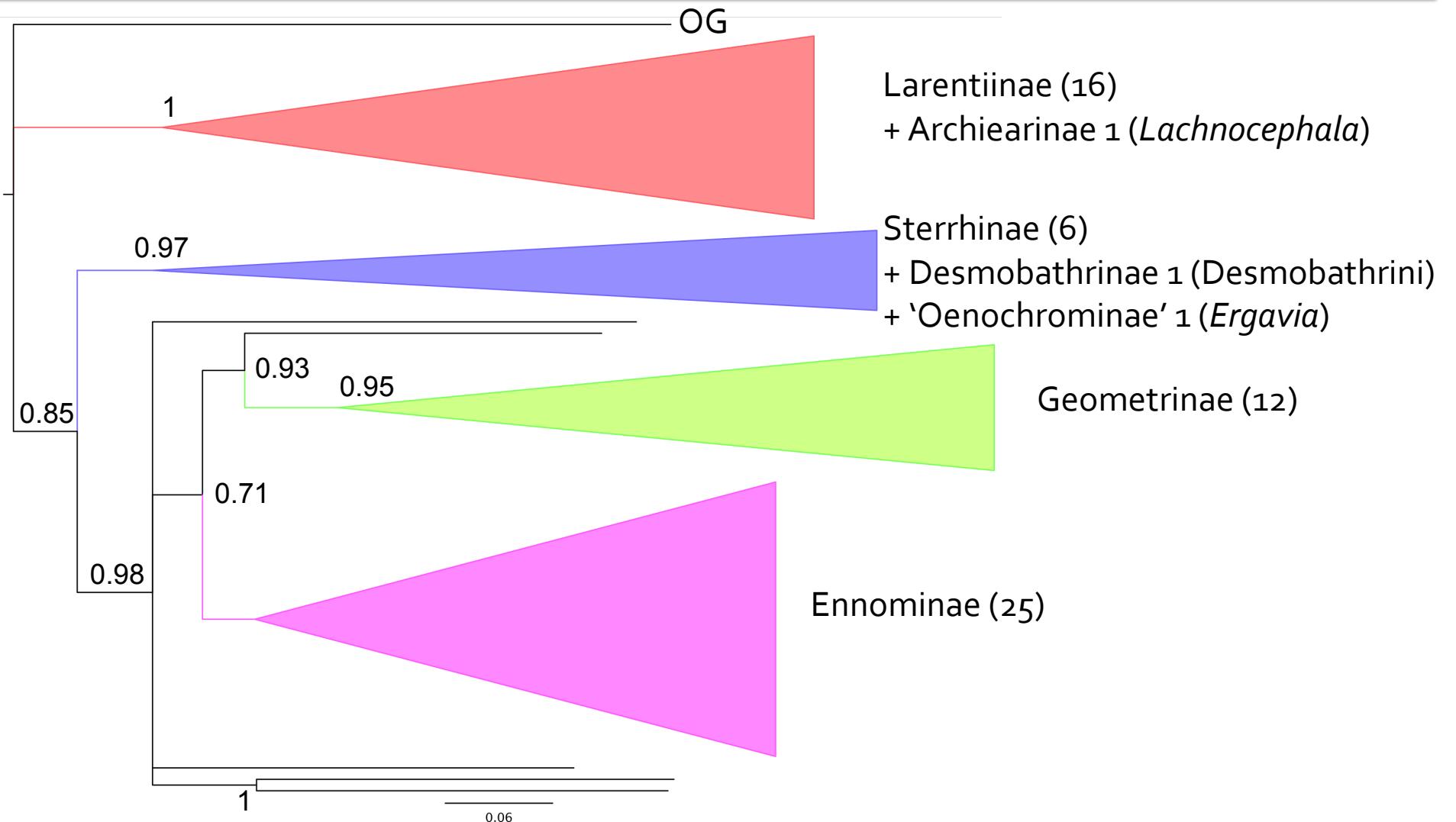
# ML tree – reduced dataset (Garli)



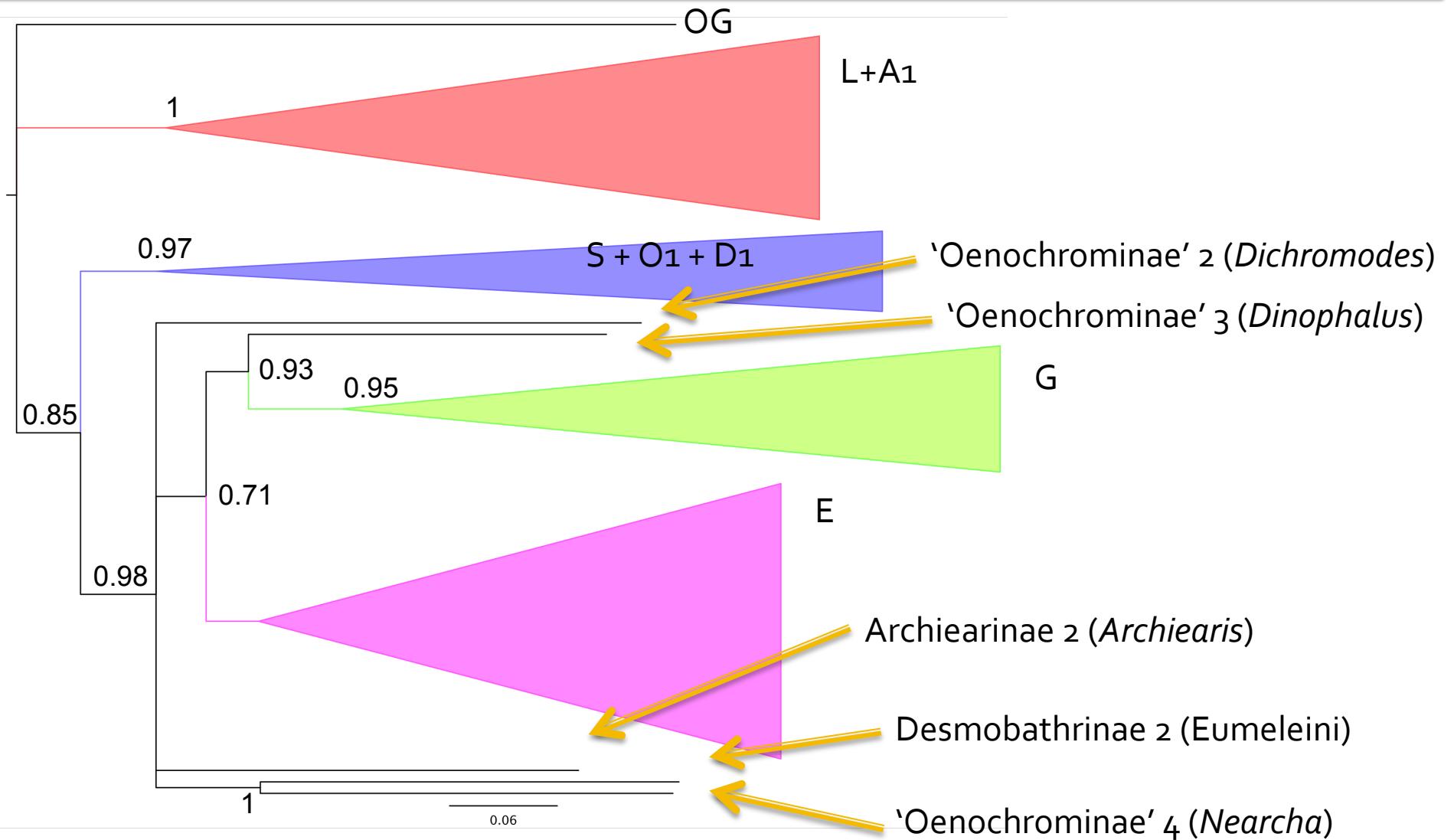
# ML tree – reduced dataset (Garli)



# Bayesian analysis – reduced dataset



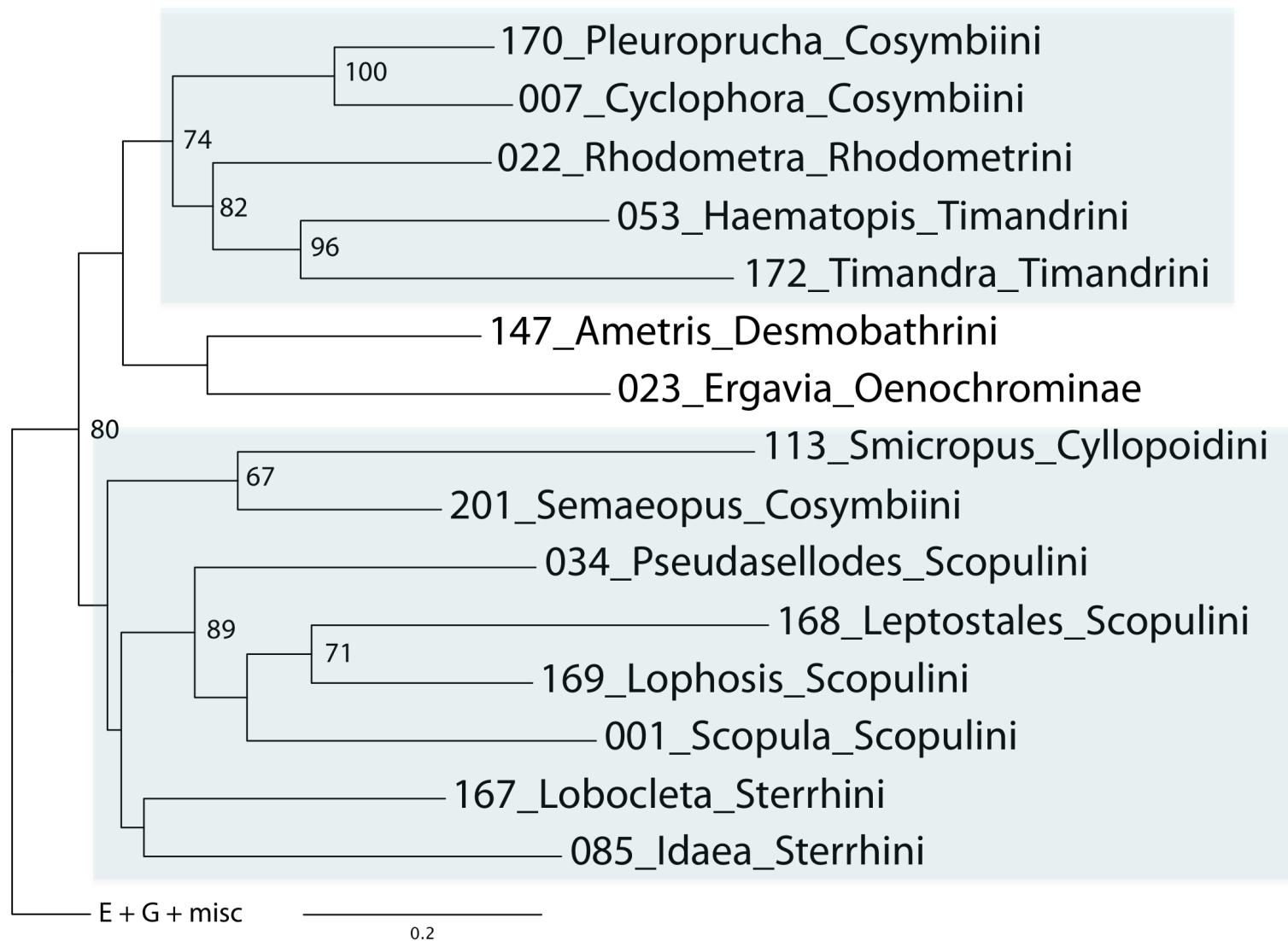
# Bayesian analysis – reduced dataset



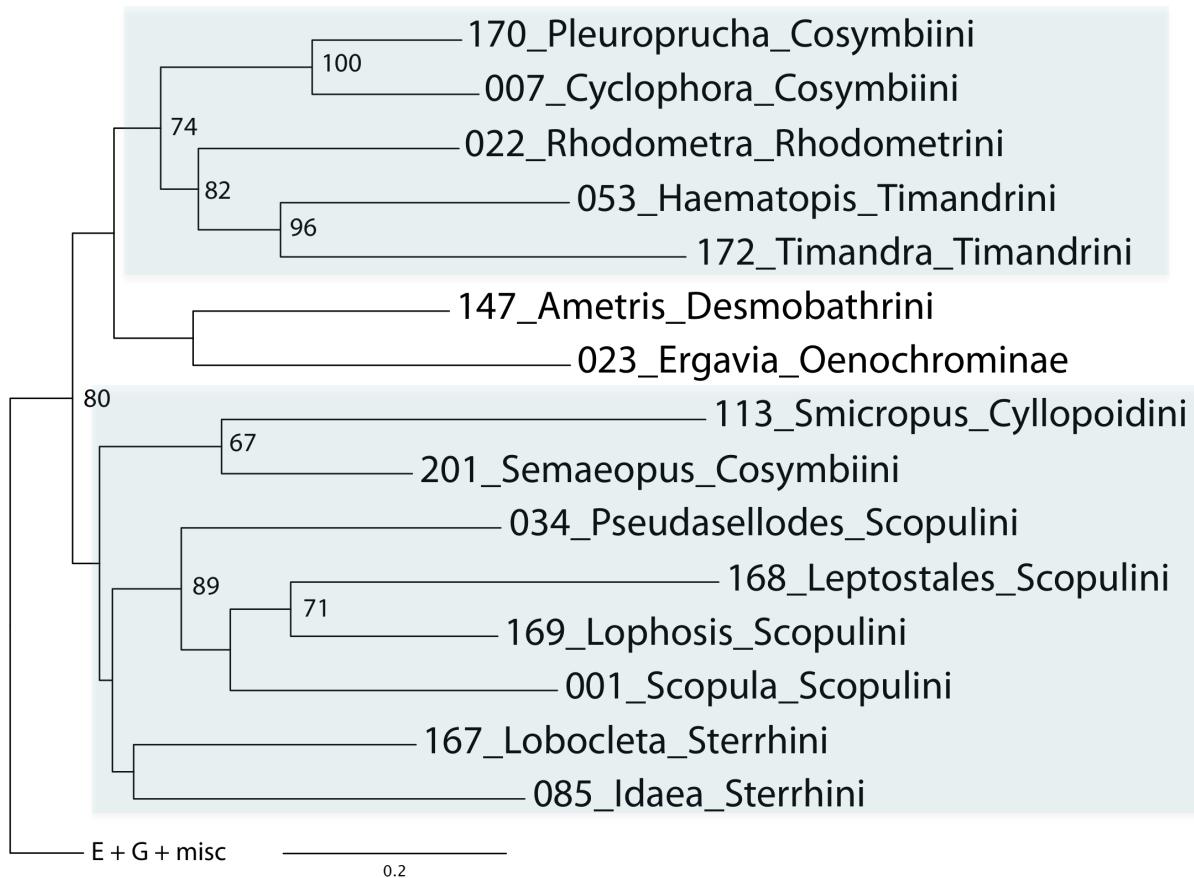
# Recovery of selected relationships

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

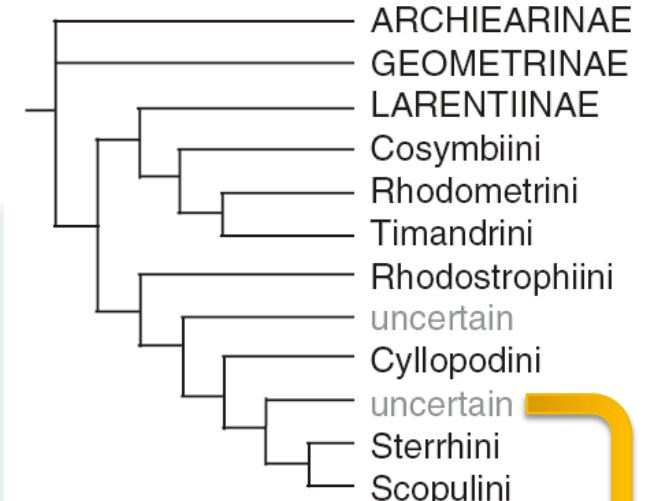
# Close-up of Sterrhinae (ML-FD)



# Close-up of Sterrhinae (ML-FD)



Sihvonen & Kaila 2004:



**Fig. 3.** A simplified cladogram of Fig. 2 transformed to show tentative tribal relationships within Sterrhinae. Taxa of uncertain tribal association are shown in grey.

Includes *Semaeopus*

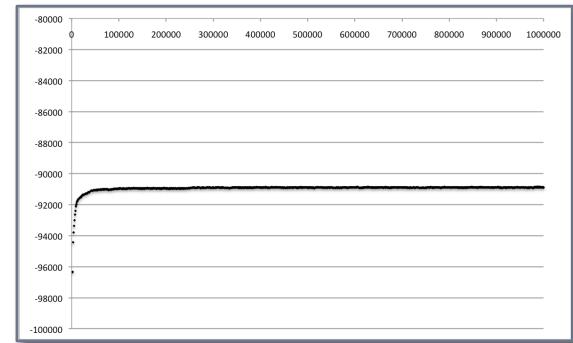
# Other interesting relationships...

- Alsophilini + Colotoini + Campaeini as in Yamamoto & 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 effect of missing characters in the matrix



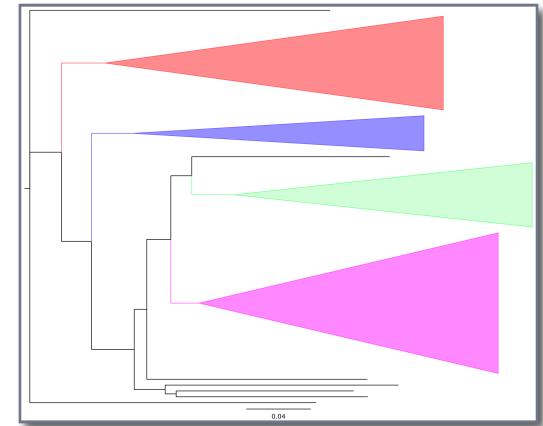
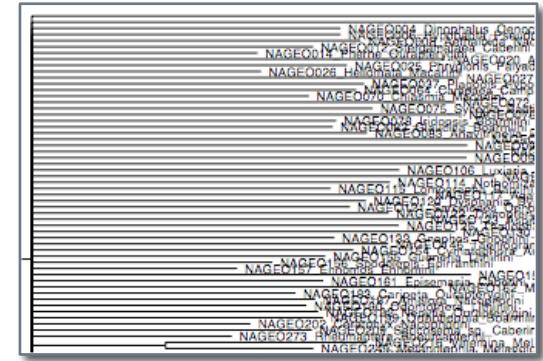
*Urania* (Uraniidae)



*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:

- Sujeewan 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

