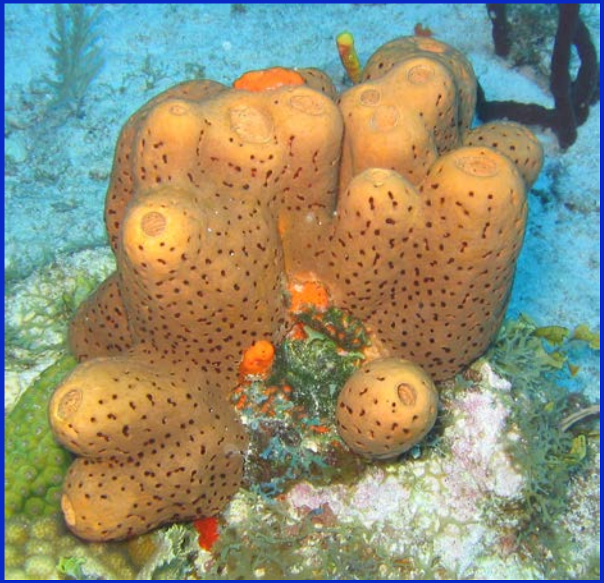
Research Computing Across the Tree of Life

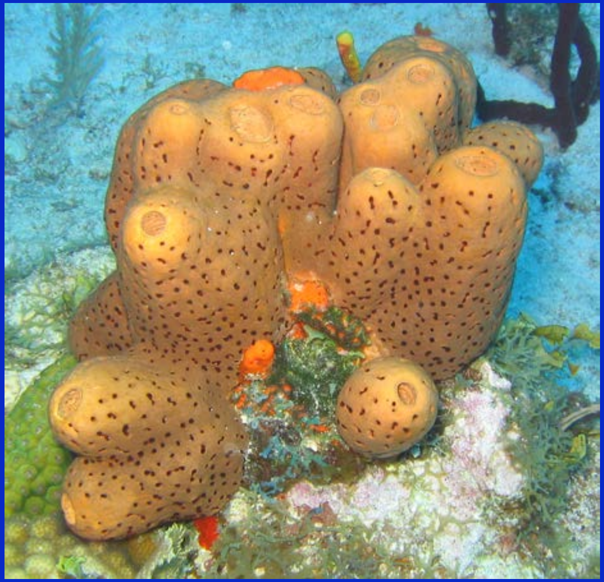
Bob Thacker

thacker@uab.edu
Department of Biology
University of Alabama at Birmingham
www.portol.org

Ecology & Evolution of Sponges



Ecology & Evolution of Sponges



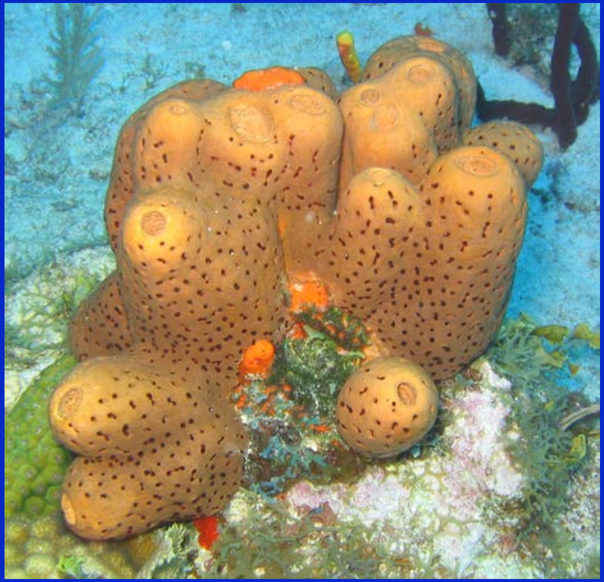
Biodiversity:

Species Discovery
& Systematics

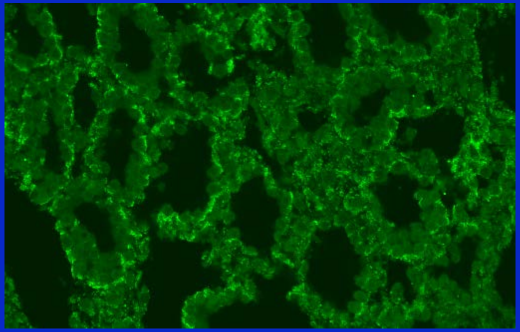
CONSERVATION INTERNATIONAL

Smithsonian

Ecology & Evolution of Sponges



Symbiotic
Interactions



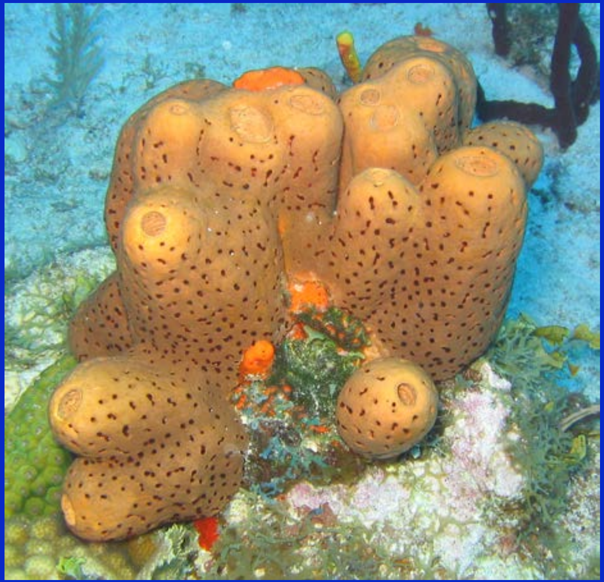
Biodiversity:
Species Discovery
& Systematics

CONSERVATION INTERNATIONAL

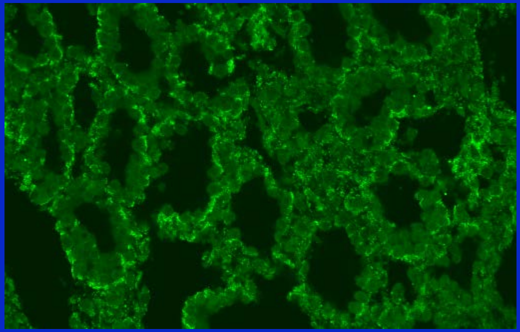
Smithsonian

NSF

Ecology & Evolution of Sponges



Symbiotic Interactions



PorToL

Porifera Tree of Life

www.portol.org

Biodiversity:

Species Discovery & Systematics

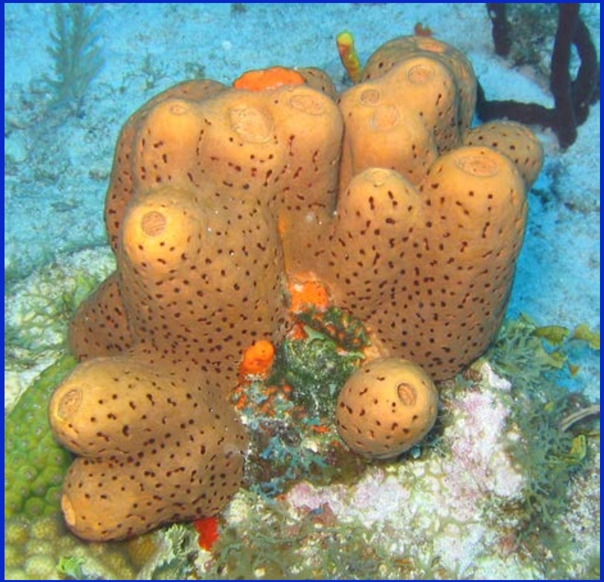
Large-Scale Phylogenetics

CONSERVATION INTERNATIONAL

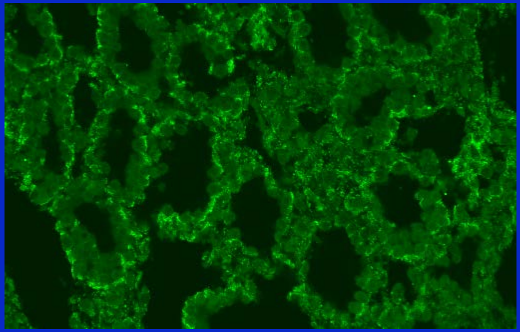
Smithsonian

NSF

Ecology & Evolution of Sponges



Symbiotic Interactions



PorToL

Porifera Tree of Life

www.portol.org

Biodiversity:

Species Discovery & Systematics

Large-Scale Phylogenetics

CONSERVATION INTERNATIONAL

Smithsonian

NSF

Arbor

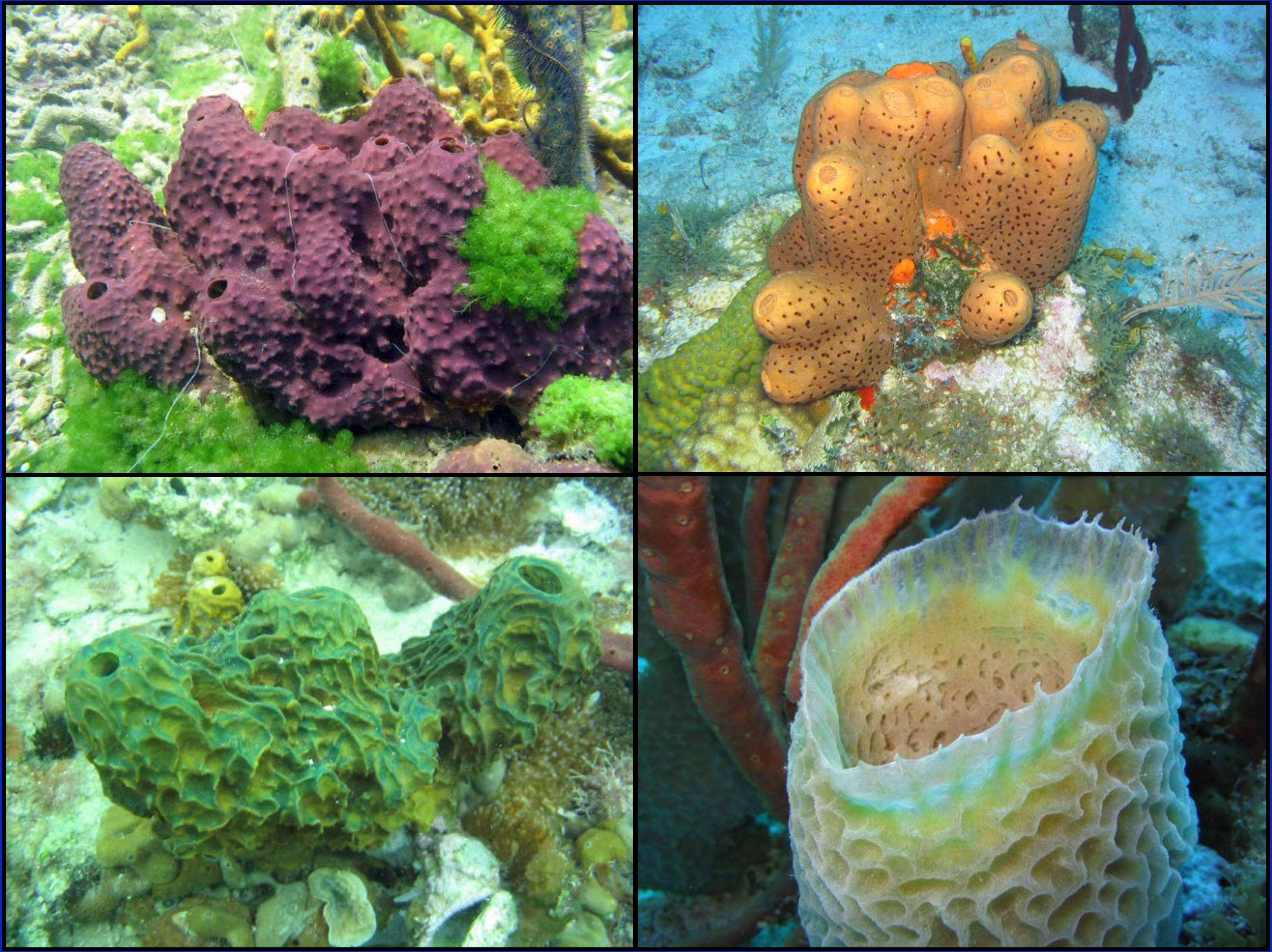
Sponge Biodiversity

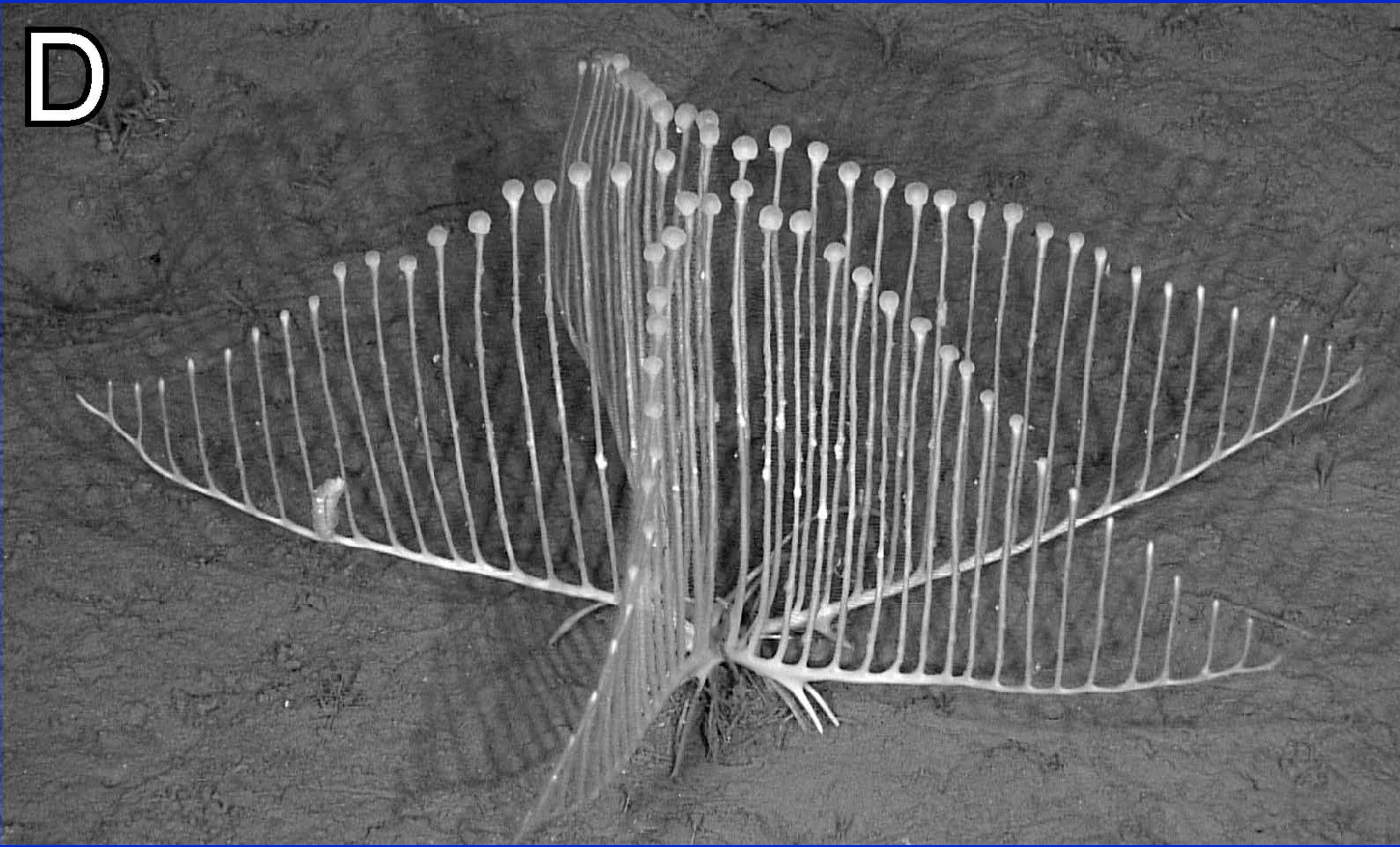
Porifera are distributed in

3 traditional classes + 1 new class = 4 classes
25 orders
127 families
682 valid genera (700+ proposed)
8,261 recognized species + estimated 8,000 to be named

ICZN sponge taxonomy is maintained as the
World Porifera Database (www.marinespecies.org/porifera)

Sponge Biodiversity

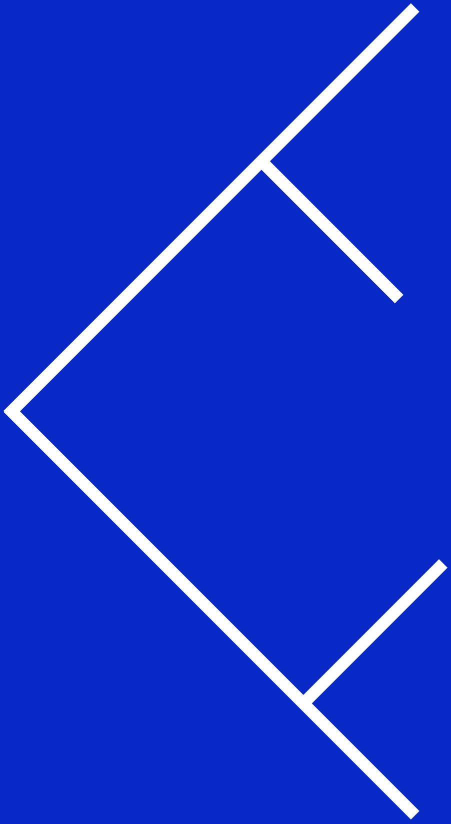




Chondrocladia lyra, a new species of carnivorous sponge from northern California (Demospongiae, Poecilosclerida, Cladorhizidae)

Welton Lee, California Academy of Sciences

A phylogeny is used to visualize evolutionary relationships among groups of organisms



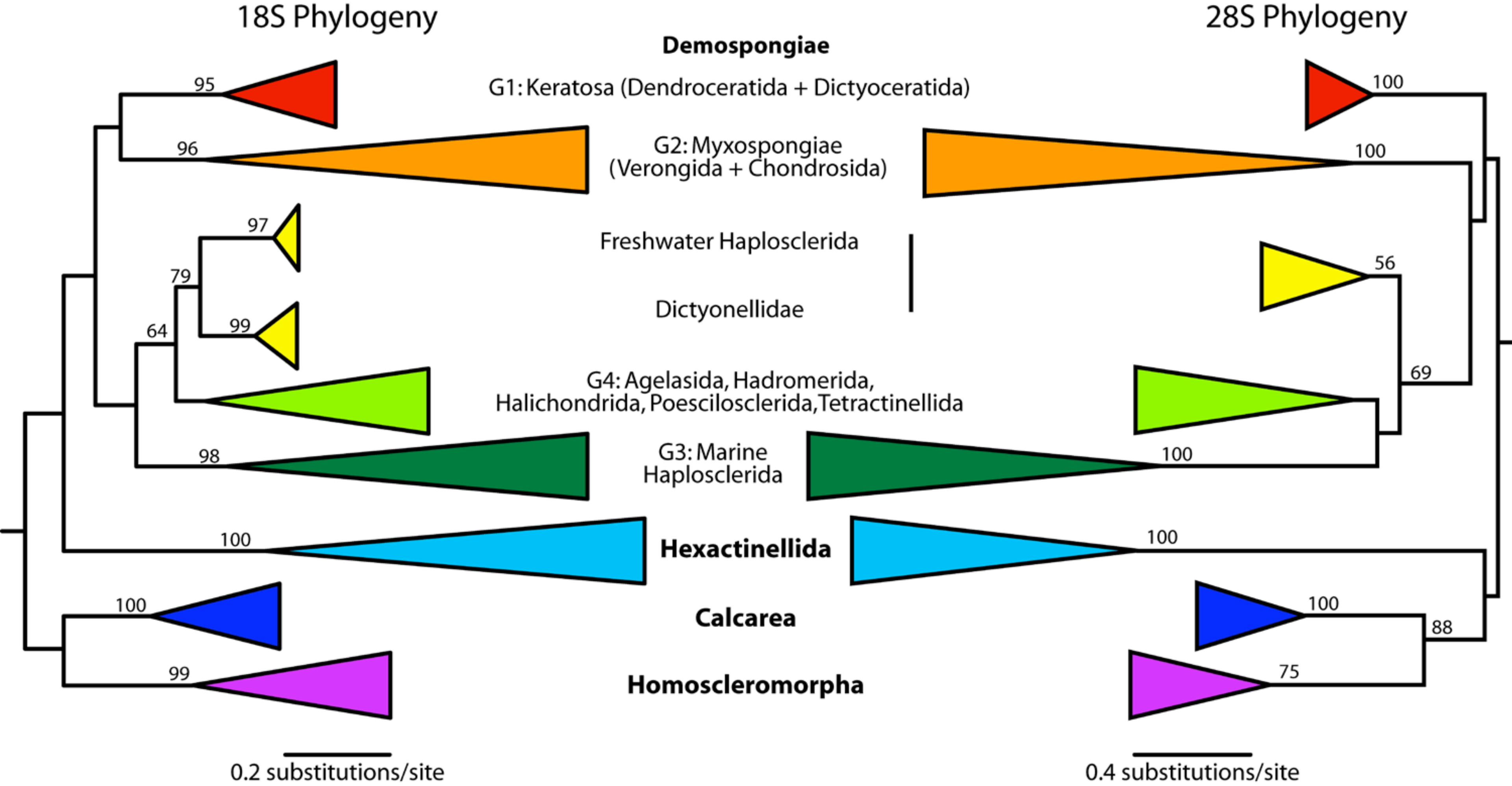
Demospongiae

Hexactinellida

Calcarea

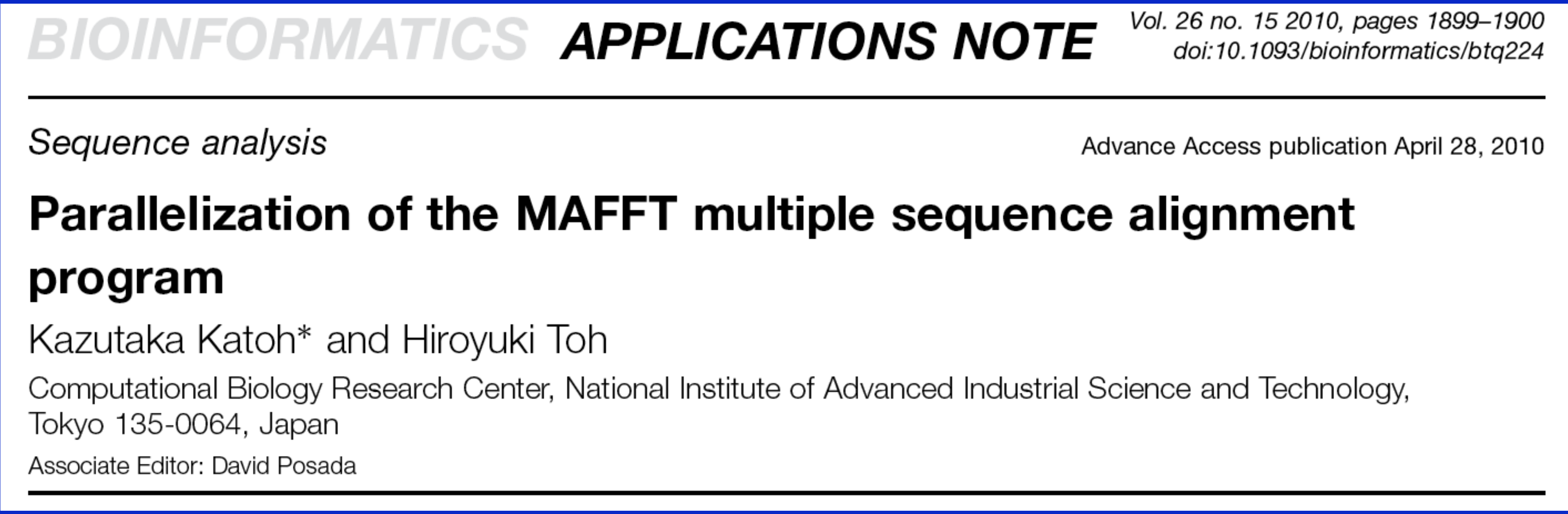
Homoscleromorpha

NSF Assembling the Tree of Life: Porifera Tree of Life (PorToL), >1300 species



Computational Needs for Phylogenetics

- Secondary Structure Alignments
  - Parallel MAFFT



BIOINFORMATICS **APPLICATIONS NOTE**
Vol. 26 no. 15 2010, pages 1899–1900
doi:10.1093/bioinformatics/btq224

Sequence analysis

Advance Access publication April 28, 2010

Parallelization of the MAFFT multiple sequence alignment program

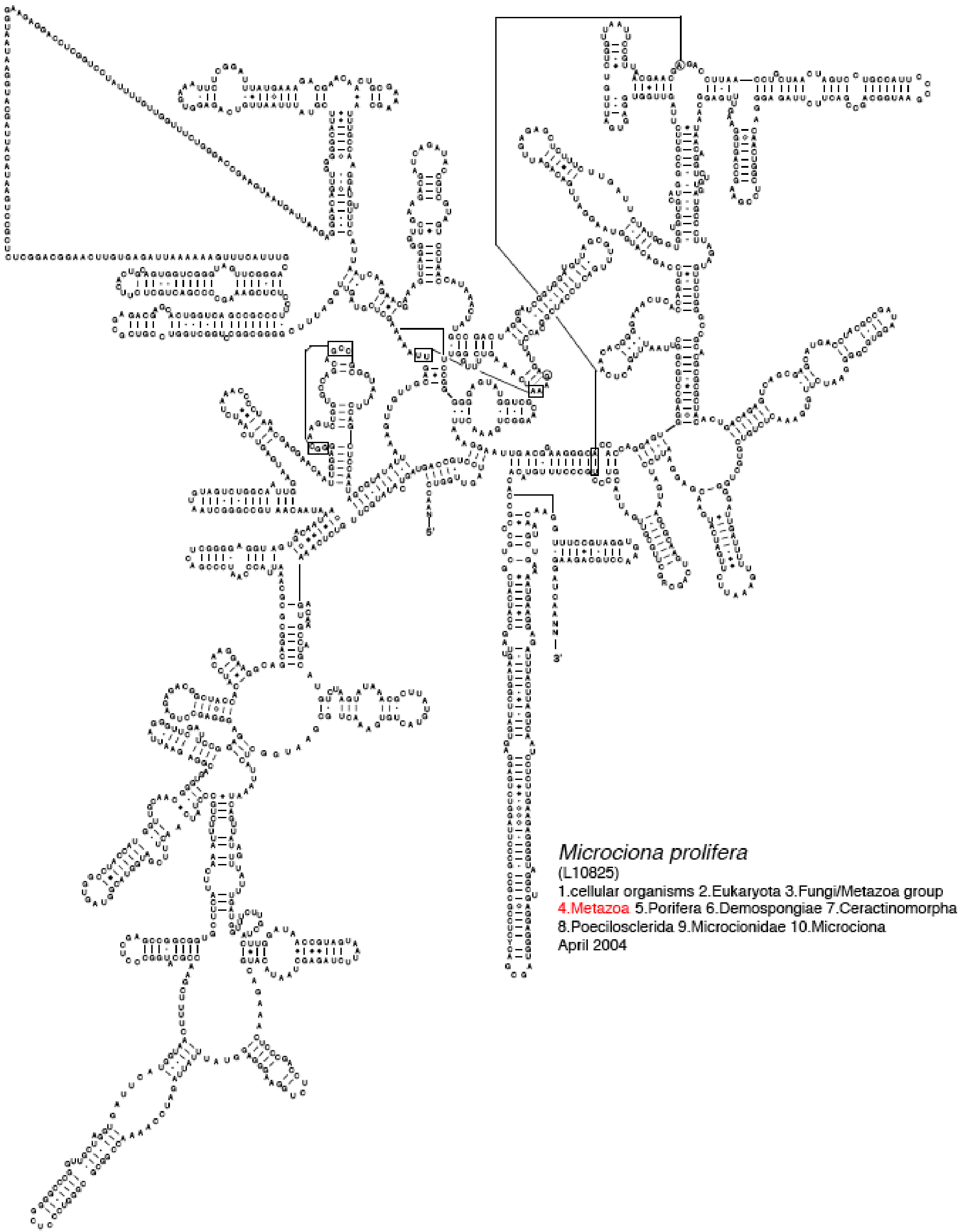
Kazutaka Katoh* and Hiroyuki Toh
Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo 135-0064, Japan

Associate Editor: David Posada

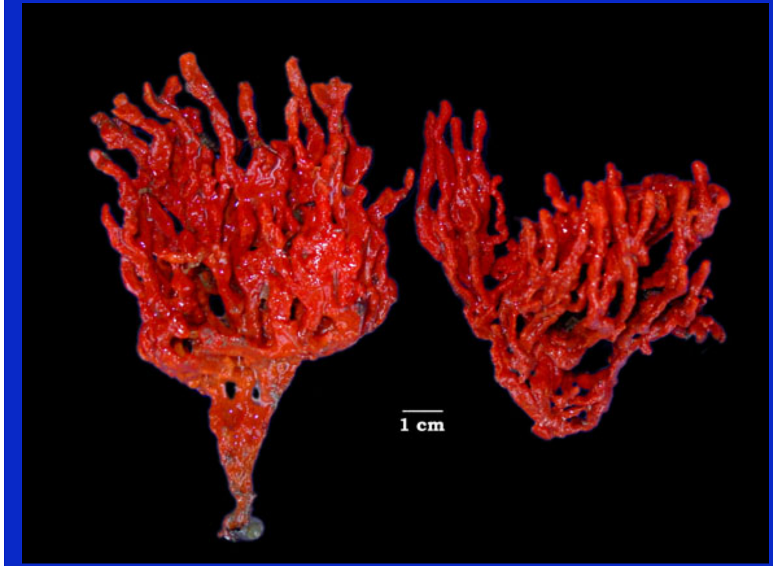
Analysis of Ribosomal Gene Sequences

Alignments based on folding of secondary structure are most accurate, but computationally complex

Secondary Structure: small subunit ribosomal RNA



Microciona prolifera
(L10825)
1.cellular organisms 2.Eukaryota 3.Fungi/Metazoa group
4.Metazoa 5.Porifera 6.Demospongiae 7.Ceractinomorpha
8.Poecilosclerida 9.Microcionidae 10.Microciona
April 2004

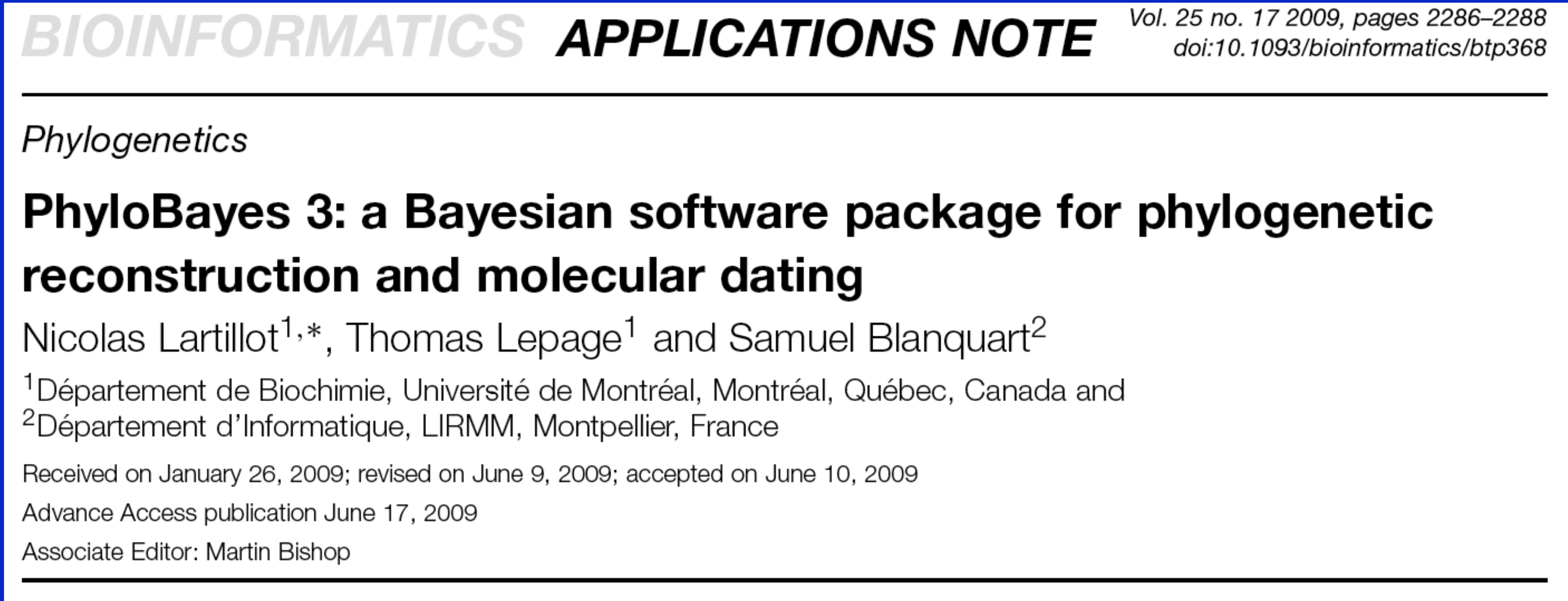


Dr. Robin Gutell
http://www.rna.ccbb.utexas.edu/
www.calacademy.org

Computational Needs for Phylogenetics

- Secondary Structure Alignments
  - Parallel MAFFT

- Bayesian Algorithms
  - Parallel PhyloBayes



BIOINFORMATICS **APPLICATIONS NOTE**
Vol. 25 no. 17 2009, pages 2286–2288
doi:10.1093/bioinformatics/btp368

Phylogenetics

PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating

Nicolas Lartillot[1,*], Thomas Lepage[1] and Samuel Blanquart[2]

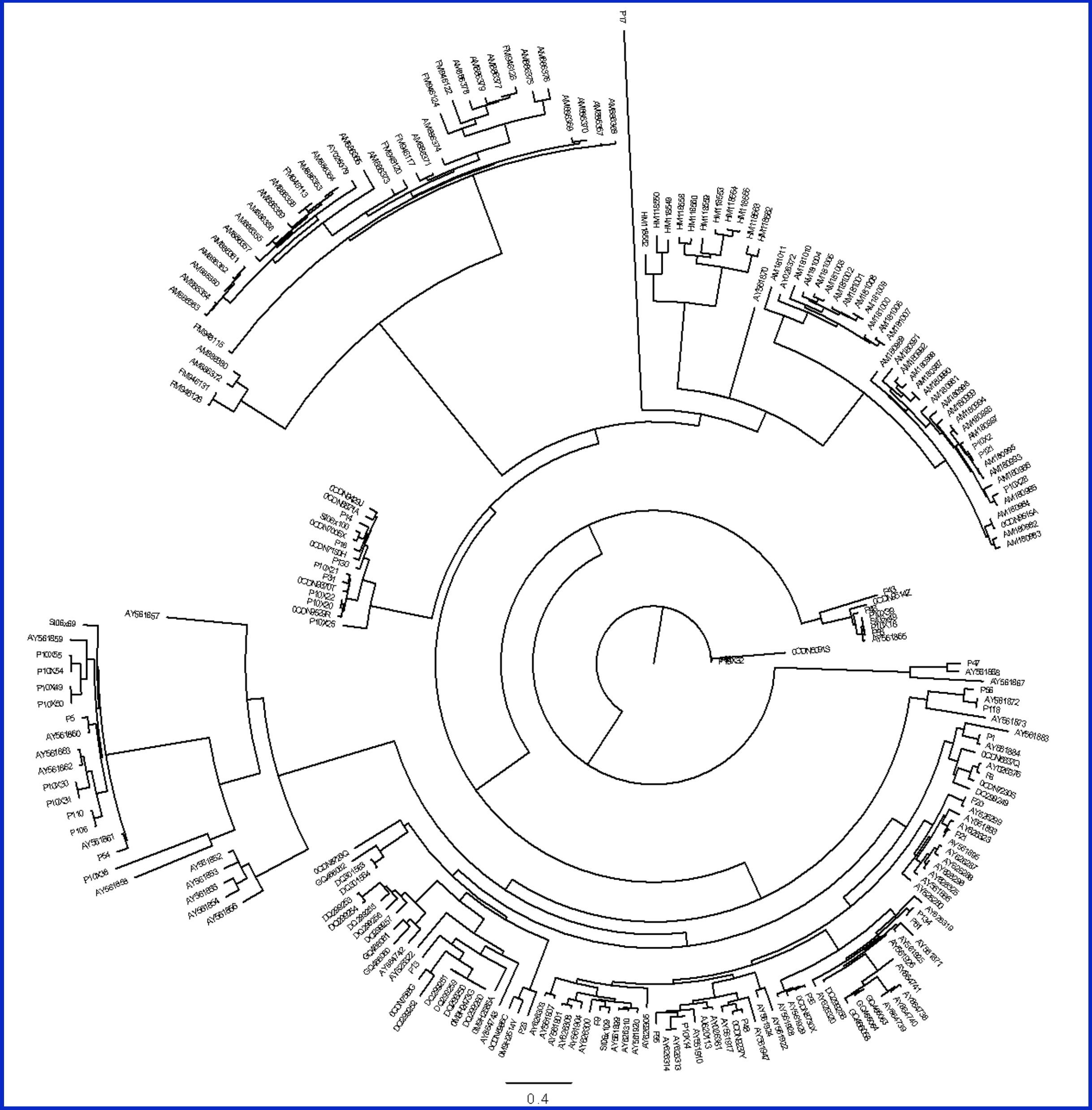
[1]Département de Biochimie, Université de Montréal, Montréal, Québec, Canada and
[2]Département d'Informatique, LIRMM, Montpellier, France

Received on January 26, 2009; revised on June 9, 2009; accepted on June 10, 2009
Advance Access publication June 17, 2009
Associate Editor: Martin Bishop

PhyloBayes

New parallel version implemented by Puri Bangalore

CAT-GTR model of sequence evolution, with over 50 different models specified across 3500 bp alignment of 250 taxa



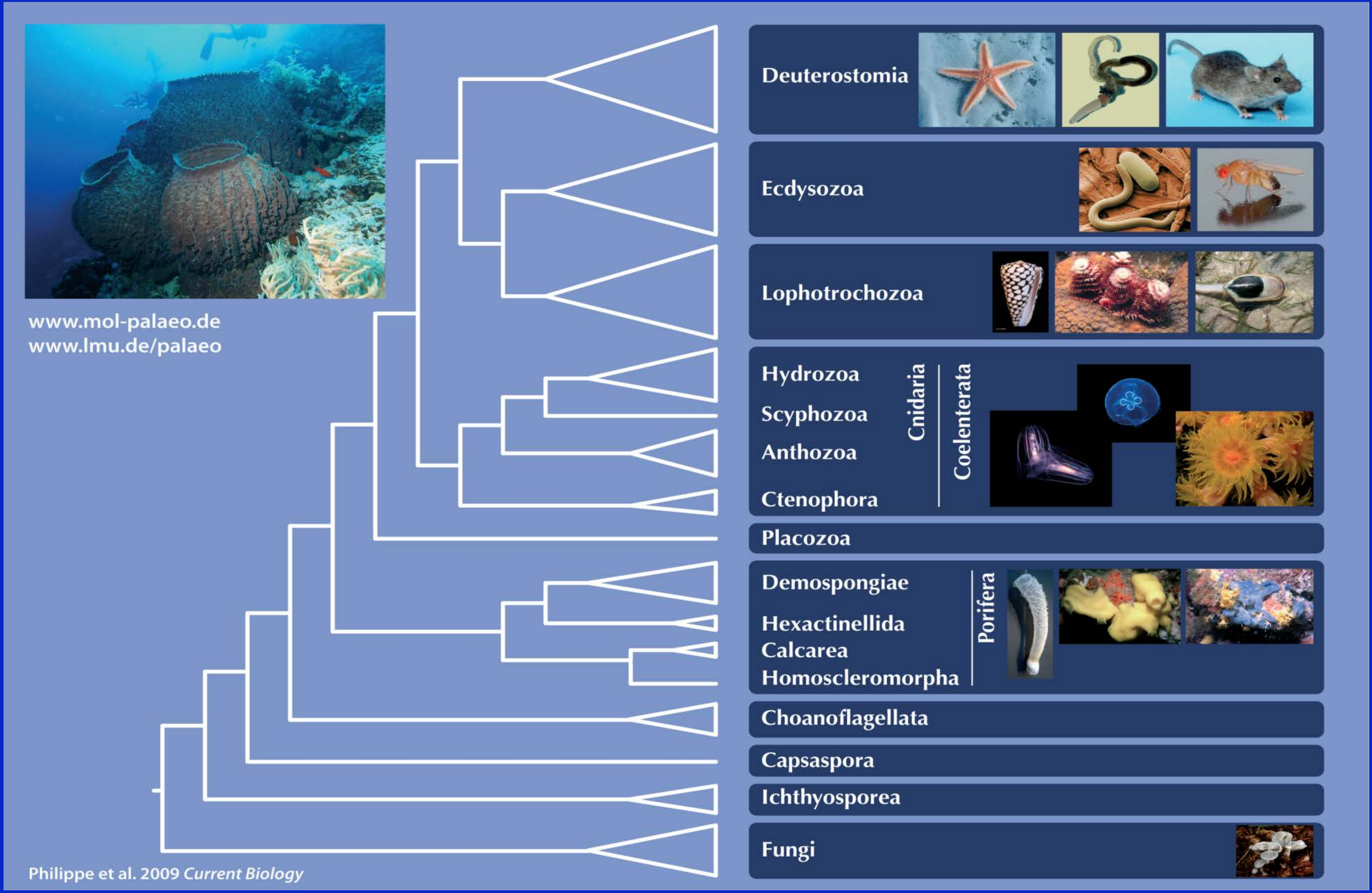
Computational Needs for Phylogenetics

- Secondary Structure Alignments
  - Parallel MAFFT

- Bayesian Phylogenies
  - Parallel PhyloBayes

- EST Libraries
  - Custom Software to Select Homologous Genes
    - 55 species
    - 128 nuclear-encoded proteins
    - 30,257 aligned amino acid positions

Sponges are an ancient lineage of metazoans.

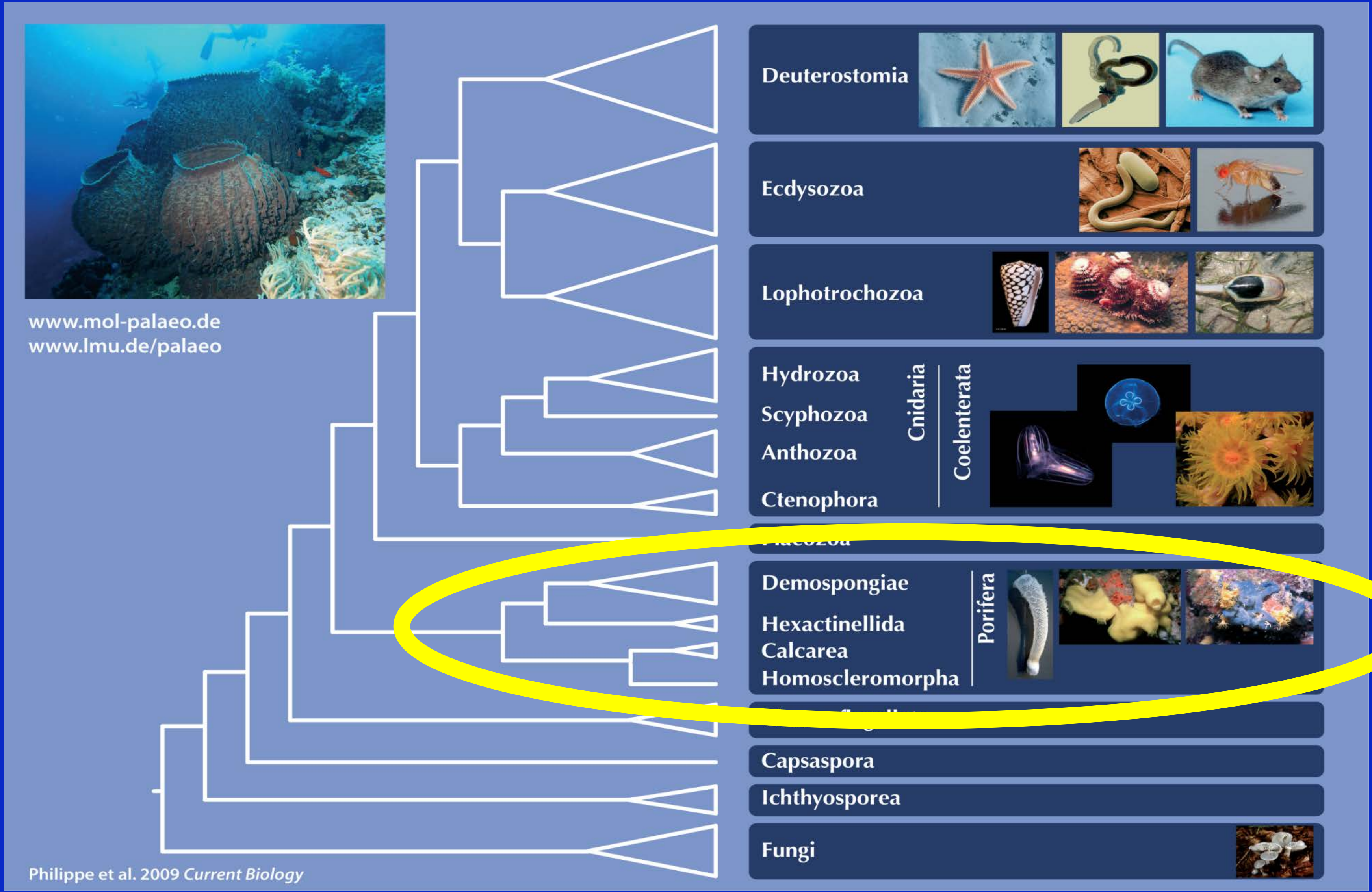


www.mol-palaeo.de
www.lmu.de/palaeo

Philippe et al. 2009 *Current Biology*

PorToL team members contributed to a recent phylogenomic study of metazoan relationships that places sponges in a monophyletic clade that is basal to other animals (Philippe et al. 2009). Bayesian tree obtained from the analysis of 128 nuclear-encoded proteins (30,257 aligned amino acid positions) for 55 terminal taxa with the CAT model of evolution.

Sponges are an ancient lineage of metazoans.



www.mol-palaeo.de
www.lmu.de/palaeo

Philippe et al. 2009 *Current Biology*

PorToL team members contributed to a recent phylogenomic study of metazoan relationships that places sponges in a monophyletic clade that is basal to other animals (Phillippe et al. 2009). Bayesian tree obtained from the analysis of 128 nuclear-encoded proteins (30,257 aligned amino acid positions) for 55 terminal taxa with the CAT model of evolution.

Computational Needs for Phylogenetics

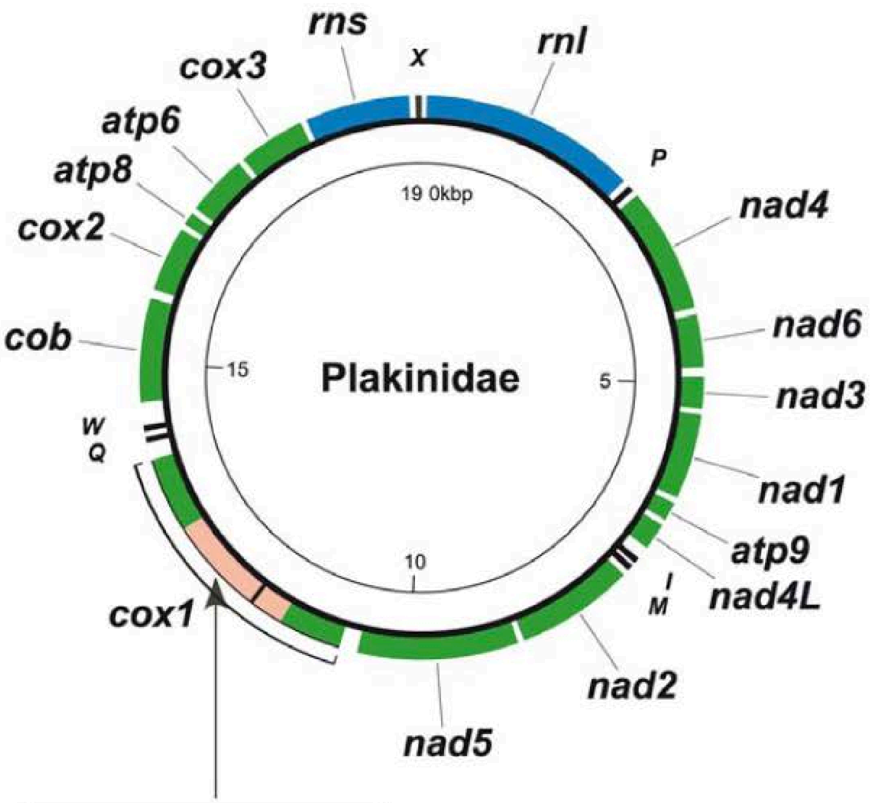
- Secondary Structure Alignments
  - Parallel MAFFT

- Bayesian Phylogenies
  - Parallel PhyloBayes

- EST Libraries
  - Custom Perl Scripts to Select Homologous Genes

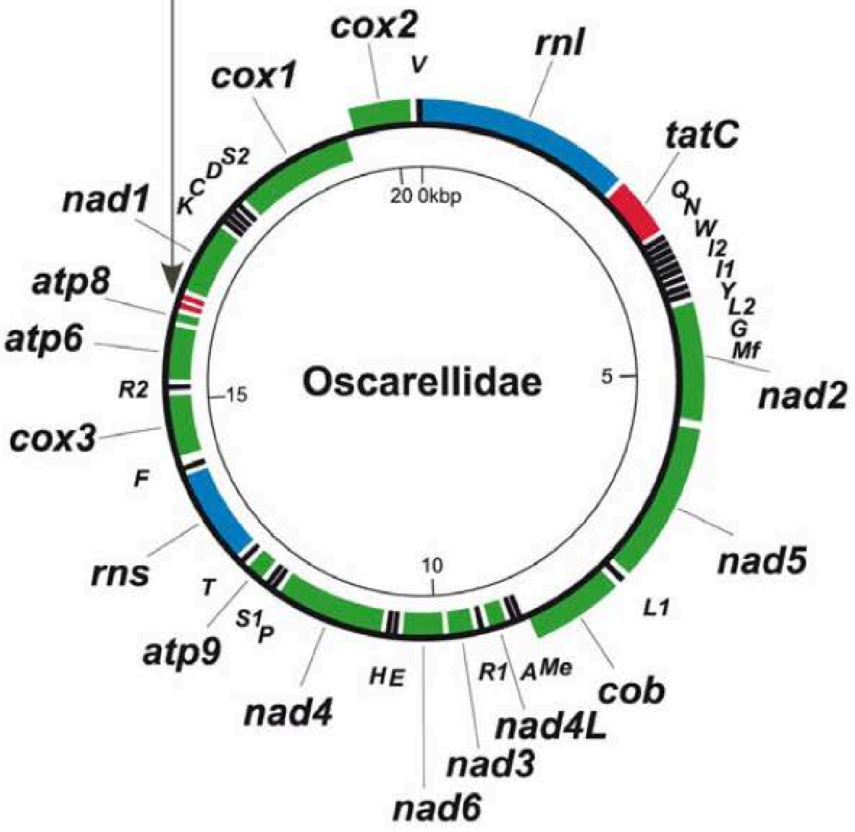
- Illumina Sequencing of Mitochondrial Genomes
  - Galaxy

Homoscleromorpha: Plakinidae vs. Oscarellidae

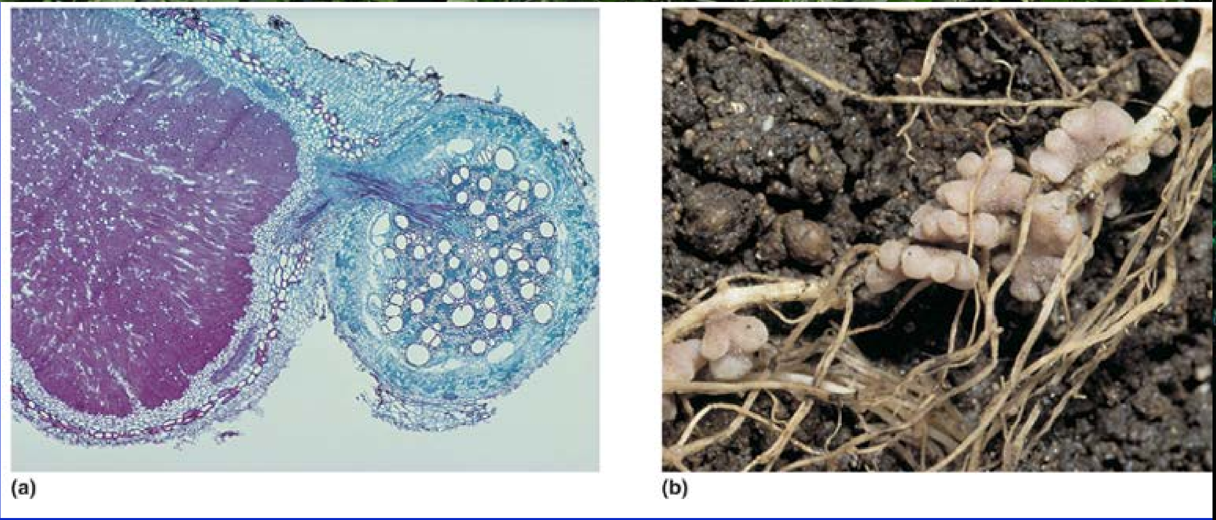


Introns:	
Plakinastrella sp.:	++
Plakina crypta:	++
Plakina trilopha:	+
Otherwise:	-

Oscarella microlobata:	*+trnY+trnT'*
Pseudocorticium jarrei:	*+trnV+trnX*
Otherwise:	*+trnV'+trnT'*



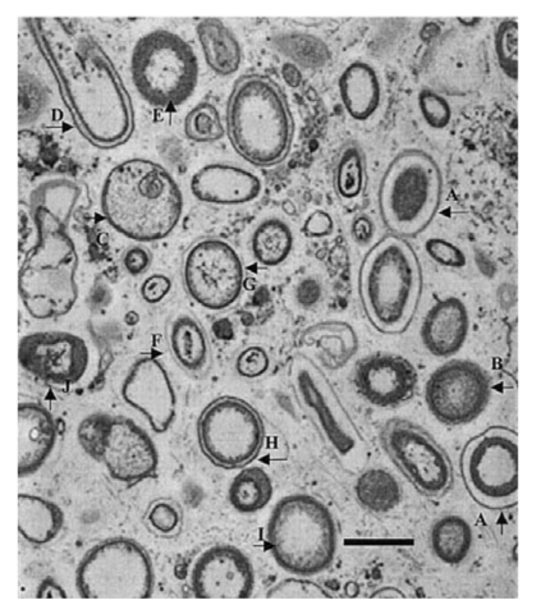
Symbiotic Interactions



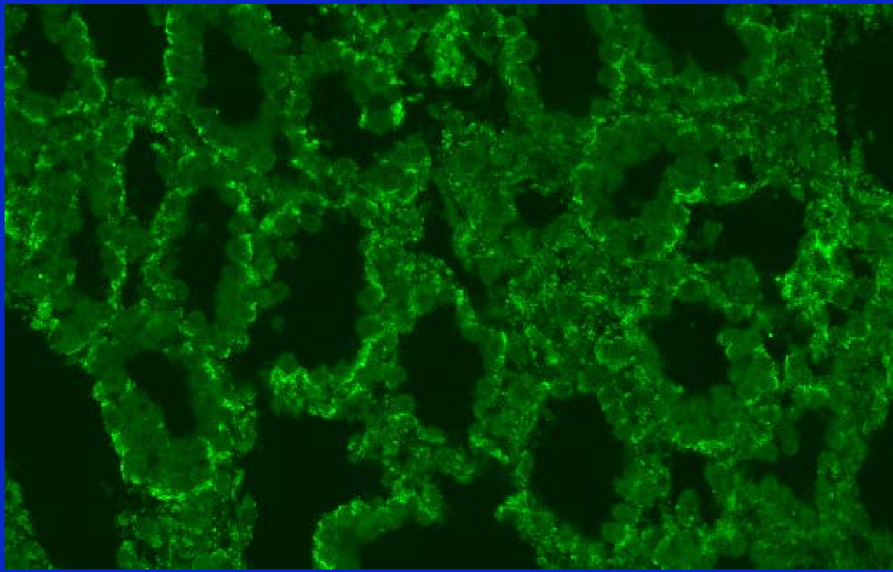
Sponge Microbiology

Sponges can host abundant and diverse communities of symbiotic microbes.

- How common are these associations?
- How diverse are these associations?
- How do microbial symbionts interact with the host sponge?



Webster et al. (2001) TEM of mesohyl of *Rhophaloeides odorabile*, with exclusively bacterial cells in this image.



Eubacterial FISH probe in the mesohyl of *Hyrtios violaceus* (Thacker et al. in prep)

Earth Microbiome Project

300 TAYLOR ET AL. MICROBIOL. MOL. BIOL. REV.

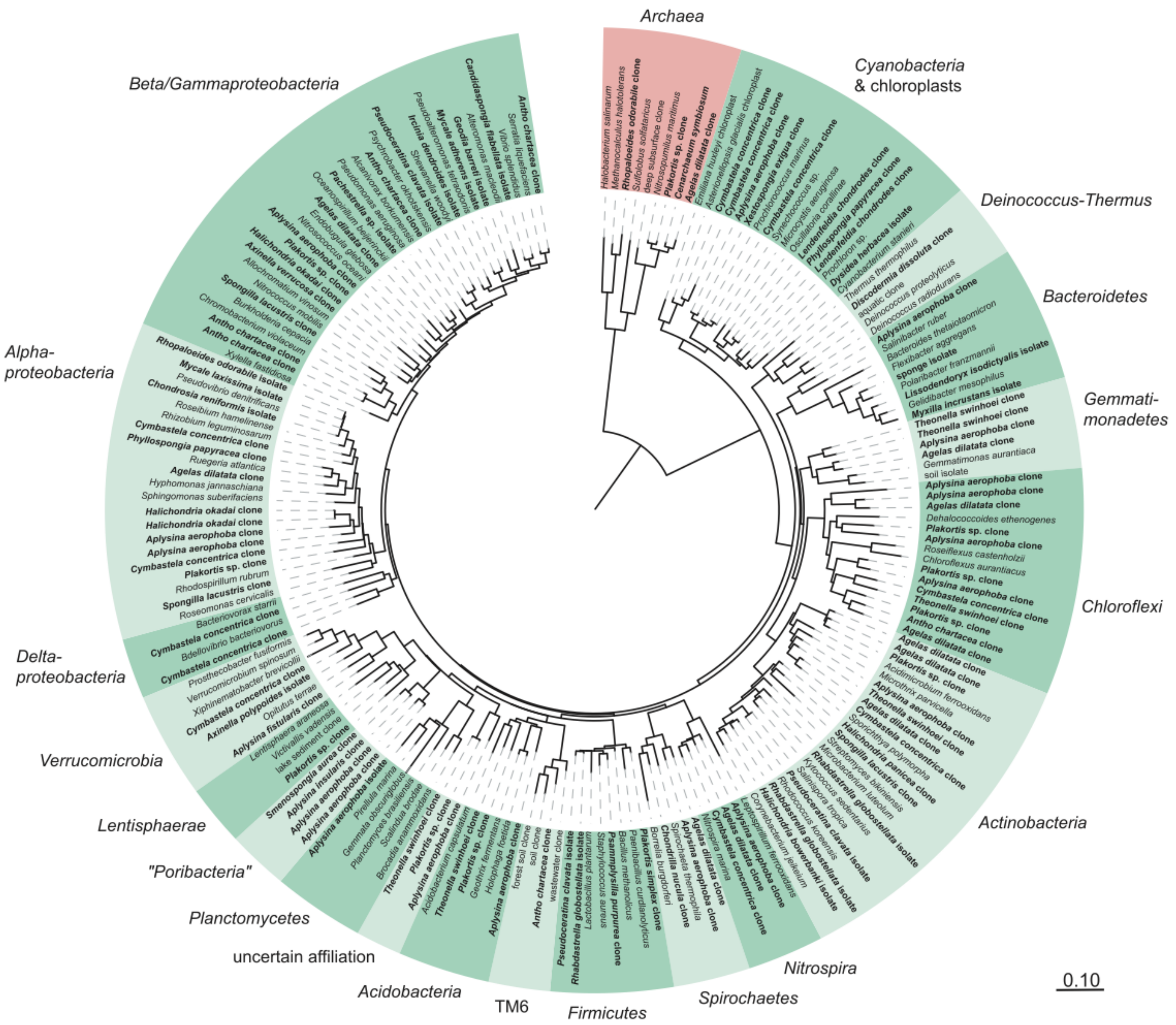
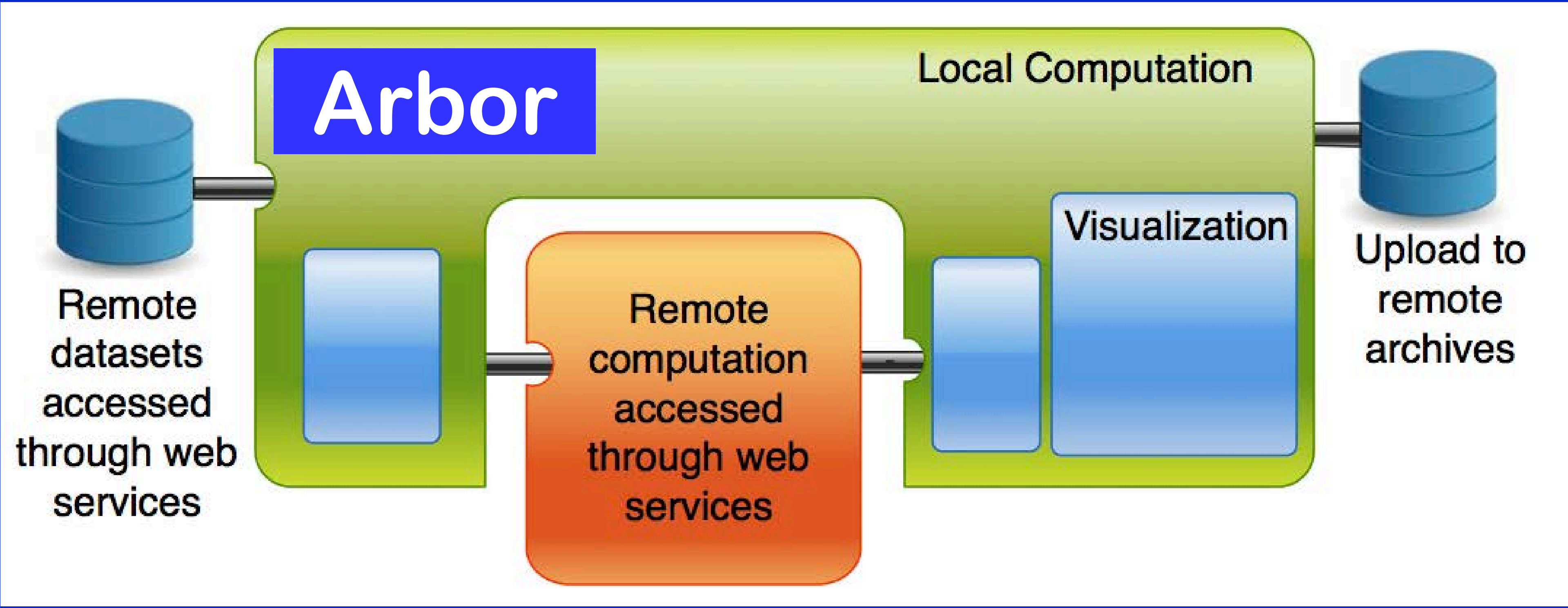


FIG. 4. 16S rRNA-based phylogeny showing representatives of all bacterial and archaeal phyla from which sponge-derived sequences have been obtained. Sponge-derived sequences are shown in bold, with additional reference sequences also included. The displayed tree is based on maximum likelihood analysis. Bar, 10% sequence divergence.

Representatives of 16 bacterial phyla and both major archaeal lineages are reported as sponge symbionts.

Comparative Analysis Workflows for the Tree of Life



- Shared workflows archived at Dryad
- Facilitate access to TeraGrid / XSEDE

NSF-Sponsored DRYAD
http://datadryad.org

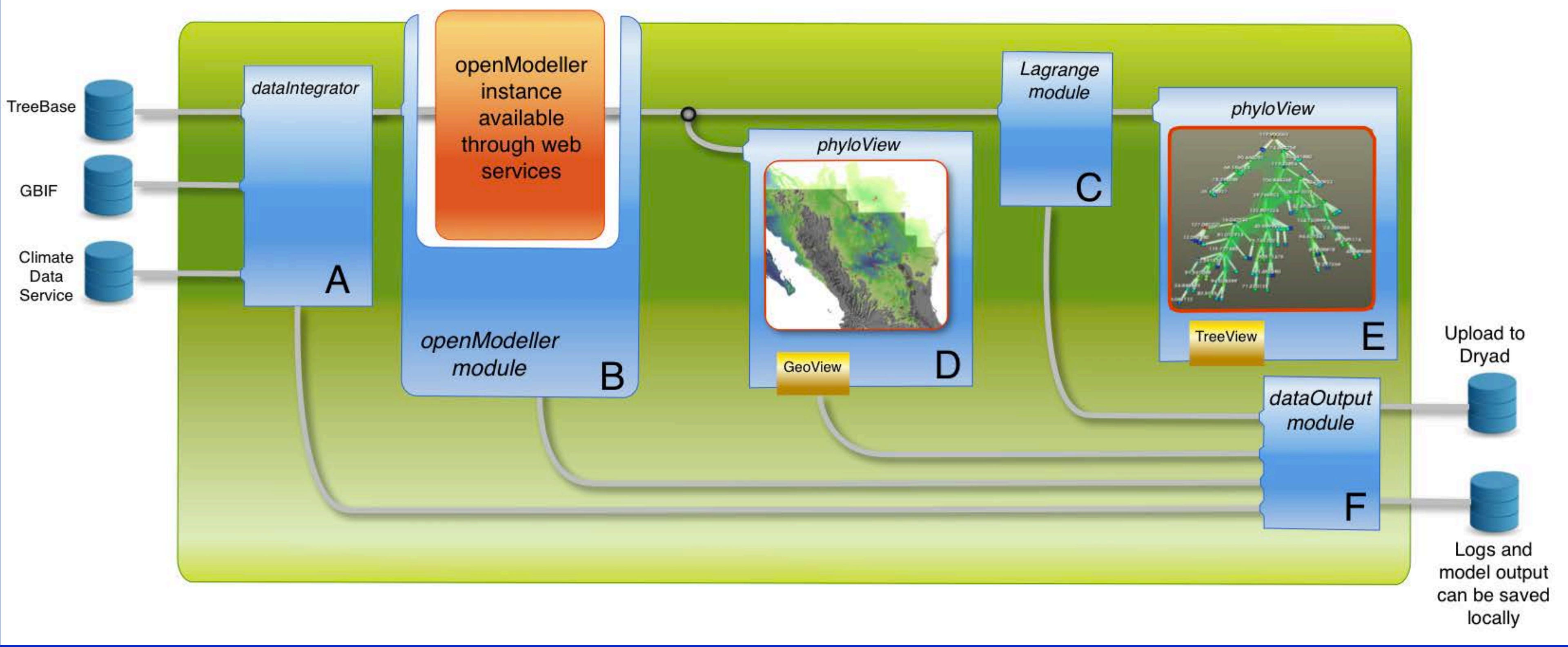
- Respository for phylogenetic data sets
- Currently expanding to accommodate any bioscience data
- Arbor workflows will be stored in DRYAD
- Reveal exactly how authors analyze data
- Retrieve methods and substitute your own data

Arbor: Case Studies

I. The Evolutionary Process of Spatial Diversification

II. The Evolution of Symbiotic Communities

III. The Evolution of Complex Interactions



How do species interactions shape long-term patterns of evolution?

Arbor Team

Curt Liesl, Kitware
Bob Thacker, UAB
Chelsea Specht, UC Berkeley
Luke Harmon, U Idaho
Jorge Soberon, U Kansas

PorToL Team

University of Alabama at Birmingham
Bob Thacker
Puri Bangalore
Smithsonian Institution / NMNH
Allen Collins
Niamh Redmond
Klaus Ruetzler
University of Richmond
Malcolm Hill
April Hill
Nova Southeastern University
Joe Lopez
Iowa State University
Dennis Lavrov

Harbor Branch Oceanographic Institution
Shirley Pomponi
Dartmouth College
Kevin Peterson
Museo Marino de Margarita, Venezuela
Cristina Diaz
Ludwig-Maximilians-Universitat Munchen,
Gert Woerheide
Queensland Museum, Australia
John Hooper
Zoological Museum Amsterdam
Rob van Soest
And many other international collaborators!