Assembling the tree of Eukaryotic Diversity
AIMS

The aims of this study are to use multigene phylogenetic analyses to:

– address specific hypotheses about eukaryotic evolution
– elucidate a robust scaffold for the eukaryotic tree of life
Previous molecular hypotheses

(A) SSU-rDNA (Sogin 1998)
Previous molecular hypotheses

(A) SSU-rDNA (Sogin 1998)

(B) Big Bang (Philippe et al 2000)
Previous molecular hypotheses

(A) SSU-rDNA (Sogin 1998)

(B) Big Bang (Philippe et al. 2000)

(C) Multigene (Baldauf et al. 2000)
Previous molecular hypotheses

(A) SSU-rDNA (Sogin 1998)

(B) Big Bang (Philippe et al. 2000)

(C) Multigene (Baldauf et al. 2000)

(D) Gene fusion (Stechmann & CS, 2003)
Previous molecular hypotheses

(A) SSU-rDNA (Sogin 1998)
(B) Big Bang (Philippe et al 2000)
(C) Multigene (Baldauf et al 2000)
(D) Gene fusion (Stechmann & CS, 2003)
Previous molecular hypotheses

LIMITED TAXON & GENE SAMPLING!
Eukaryotic relationships

• No evidence of “base” and “crown”

• Some well supported clades are emerging from comparisons of multiple genes

• Deep nodes remains elusive
Proposed “super-kingdoms”

Opisthokonts

Plantae

Amoebozoa

“Chromalveolates”

“Cercozoa”

“Excavata”

PLANTAE

Glaucocystophytes

Red Algae

Green algae & Land Plants
AMOEBOZOA

Lobose Amoebae

Slime molds

Pelobionts & Entamoeba

AMOEBOZOA
“CHROMALVEOLATES”

Ciliates

Dinoflagellates

Apicomplexa

Alveolates
“CHROMALVEOLATES”

Cryptomonads

Haptophytes

Stramenopiles

‘Chromists’
“CERCOZOA”

Cercomonads
Chlorarchniophyte

Foraminifera

Radiolaria
“EXCAVATA”

**Euglenozoa, Heterolobosea & Jakobids**

**Trimastix, Oxymonads & Malawimonas**

**Parabasalia**

**Diplomonads, Retortamonads & Carpediemonas**
Our Approach

- **TAXON SAMPLING:**
  - ~10kb from $\geq 200$, predominantly free-living microbial eukaryotes
  - well-circumscribed clades and taxa of as yet unknown affinities.

- **DATA ANALYSIS:**
  - combine existing approaches with newly developed methods for partitioning multigene data.
<table>
<thead>
<tr>
<th>Table 1. Protist taxa to be sampled and currently available data.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GROUP</strong></td>
</tr>
<tr>
<td>ALVEOLATA</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Note:** Available in existing collections (predominantly ATCC) or already in hand; S, A, B, E, R, 9, 7: genes targeted for sequencing (names at bottom right).
<table>
<thead>
<tr>
<th>Total number of genes available</th>
<th>140</th>
<th>23</th>
<th>23</th>
<th>22</th>
<th>20</th>
<th>8</th>
<th>8</th>
<th>11</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSU rRNA</td>
<td>cob</td>
<td>Actin</td>
<td>α-Tubulin</td>
<td>β-Tubulin</td>
<td>EF1-α</td>
<td>RPB1</td>
<td>HSP90</td>
<td>HSP70</td>
<td></td>
</tr>
</tbody>
</table>
“CERCOZOA”

Cercomonads
Chlorarchnhiophyte

Foraminifera

Radiolaria
## Taxa (cont)

<table>
<thead>
<tr>
<th>CERCOMONADS</th>
<th>Cercomonas longicauda</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cercomonas sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Heteromita globosa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Massisteria marina</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Spongomonas minima</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TESTATE AMOEBAE</th>
<th>Ovulinata parva</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
<th>+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amphitremids</td>
<td>Rhogostoma sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Euglyphids</td>
<td>Corythionella sp.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Euglypha rotunda</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trinema lineare</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Database

• Molecular data: U. Iowa
  – GenBank formatted entries, easy overview of completeness, readily generate multigene alignments

• Linked to www.mbl.edu/microscope
Outreach

- “Protist Diversity” workshop
- Training: grad, undergrad, postdoc
- www.mbl.edu/microscope
- Inform protist collections
RESPONSIBILITIES OF PIs

Laura Katz (LK), Smith College & University of Massachusetts
- management of this collaborative project, include coordinating DNA acquisition, assessment, storage and transfer, overseeing regular communications and organizing annual meetings
- collection of 1/3 of molecular data

Debashish Bhattacharya (DB), University of Iowa
- collection of 1/3 of molecular data
- development of appropriate databases for data organization and analyses.

Donald Burgess (DEB), American Type Culture Collection
- supervising protist culture and quality assurance.
- provide DNA samples for characterization by sequencing, PCR etc.

John Huelsenbeck (JH), University of California – San Diego
- provide guidance on phylogenetic analyses using existing approaches
- develop novel partitioning methods for multigene data

John Logsdon (JL), University of Iowa
- collection of 1/3 of molecular data
- support in database development

David J. Patterson (DP), Marine Biological Laboratories
- develop micro*scope (www.mbl.edu/microscope) into a resource to will link data to other relevant internet sites.
- development of a microscopy workshop, which will take place at MBL in years 2 and 4.
UNIQUE CHALLENGES

Combination of:

- Considerable rate heterogeneity among lineages
- Potentially large and complex genomes
- Possibility of lateral gene transfers and/or symbioses
- Modeling nucleotide, codon & amino acid evolution over long time scales
Anticipated Outcomes

Resulting phylogeny of eukaryotes is essential for:

1) interpreting origin and diversification of eukaryotic cells

2) understanding multiple origins of multicellular eukaryotes

3) unifying the universal tree of life that includes both prokaryotes and eukaryotes