Probing early eukaryotic evolution using phylogenetic methods

The Universal SSU rRNA Tree
Wheelis et al. 1992 PNAS 89: 2930

The SSU Ribosomal RNA Tree for Eukaryotes

The Archezoa Hypothesis - primitively amitochondriate eukaryotes
(Tom Cavalier-Smith 1983)

The Archezoa Hypothesis
T. Cavalier-Smith (1983)

- The Archezoa hypothesis would fall if:
  - Find mitochondrial genes on archeozoan genomes
  - Find that archezoans branch among aerobic species with mitochondria
  - Find mitochondrion-derived organelles in archezoans
Mitochondrial HSP 70 is encoded by the host nucleus but is of endosymbiotic origin

Targeting sequences

<table>
<thead>
<tr>
<th>ATPase DOMAIN</th>
<th>Peptide Binding DOMAIN</th>
<th>Retention sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
</tr>
</tbody>
</table>

Cytoplasmic HSP70

mitochondrial HSP70

other proteobacteria

cyanobacteria

Endosymbiotic origin

Mitochondrial genes in Archezoa

<table>
<thead>
<tr>
<th>Archezoa</th>
<th>Proteins of mitochondrial origin*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giardia / Spironucleus</td>
<td>Heat shock 70, Chaperonin 60</td>
</tr>
<tr>
<td>Trichomonas</td>
<td>Heat shock 70, Chaperonin 60</td>
</tr>
<tr>
<td>Microsporidia</td>
<td>Heat shock 70</td>
</tr>
<tr>
<td>Entamoeba</td>
<td>Heat shock 70, chaperonin 60</td>
</tr>
</tbody>
</table>

*defined as forming a monophyletic group with mitochondrial homologues in a non-controversial species phylogeny

Long branches may cause problems for phylogenetic analysis

- Felsenstein (1978) made a simple model phylogeny including four taxa and a mixture of short and long branches

TRUE TREE

\[ \begin{align*}
  A & \quad B \quad C \\
  q & \quad p & \quad p > q
\end{align*} \]

WRONG TREE

\[ \begin{align*}
  A & \quad B \\
  p & \quad q
\end{align*} \]

- Methods which assume all sites change at the same rate (e.g. PROTML) may be particularly sensitive to this problem

Chaperonin 60 Protein Maximum Likelihood Tree

(PROTML, Roger et al. 1998, PNAS 95: 229)

Long branches

Chaperonin 60 Protein Maximum Likelihood Tree

(PROTML, Roger et al. 1998, PNAS 95: 229)
A simple experiment:

- Does the Cpn60 tree topology change:
  - If we remove long-branch outgroups
  - If we remove sites where every species has the same amino acid

The Archezoa Hypothesis
T. Cavalier-Smith (1983)

- The Archezoa hypothesis would fall if:
  - Find mitochondrial genes on archezoan genomes
  - Find that archezoans branch among aerobic species with mitochondria
  - Find mitochondrion-derived organelles in archezoans

Microsporidia are primitive and early branching eukaryotes or relatives of fungi?

- Microsporidia branch deep in some trees but not in others
  - Deep: SSU rRNA, EF-2
  - With fungi: tubulin and HSP70
- Microsporidia contain a mitochondrial HSP70
  - They once contained the mitochondrial endosymbiont and thus are not Archezoa sensu Cavalier-Smith
    - Hirt et al., 1997
  - Or they obtained this gene (and the one for tubulin) from fungi via HGT
The tree shown is the maximum likelihood tree obtained using a program called PROTML.

Maximum likelihood and phylogenetic inference
- The maximum likelihood tree is the one with the highest probability of giving rise to the data under a particular model.
- It is not the probability that it is the "true" tree.

Assumptions underpinning the PROTML model
- Assumes all sites in the molecule can change
- Assumes that all sequences are evolving in the same way
- EF-2 violates both of these assumptions:

A combination of factors (outgroup GC content and site rate heterogeneity) influence the EF-2 DNA tree

Eukaryotic phylogeny

LogDet/Paralinear distances for EF-2 DNA variable sites
codon positions 1+2
A combination of factors (outgroup GC content & site rate heterogeneity) influence the EF-2 DNA tree

Are the former Archezoa really ancient offshoots?
- Probably not:
  - Microsporidia are related to fungi (Tubulin, RNA polymerase, LSU rRNA, HSP70, TATA binding protein, EF-2, EF-1 alpha, SSU rRNA)
  - Evidence for Giardia and Trichomonas branching deeper than other eukaryotes is based on trees made using unrealistic assumptions (often PROTML)
  - There is plenty of room for new hypotheses

If former archezoa contain genes from the mitochondrial endosymbiont what happened to the organelle?

Microsporidia are obligate intracellular parasites

Life cycle of microsporidia

mtHsp70 has diverse roles in mitochondria

mtHsp70 also plays a role in assembly of Fe-S clusters
- an essential function for yeast mitochondria (Lill & Kispal, 2000)
Microsporidian HSP70 have no obvious N-terminal targeting signal

<table>
<thead>
<tr>
<th>Organism</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elachista cast</td>
<td>M-----------------------------TR</td>
</tr>
<tr>
<td>Haemoproteus antarcticus</td>
<td>M-------------------------------TR</td>
</tr>
<tr>
<td>Trachipleistophora hominum</td>
<td>M-------------------------------TR</td>
</tr>
<tr>
<td>Aeropyrum oxalaticum</td>
<td>M-----------------------------TR</td>
</tr>
<tr>
<td>Salmo salar nebulosus</td>
<td>M-----------------------------TR</td>
</tr>
<tr>
<td>Schizosaccharomyces cellulare</td>
<td>-----------------------------TR</td>
</tr>
<tr>
<td><em>T. pallidum</em></td>
<td>M-----------------------------TR</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>M-----------------------------TR</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>M-----------------------------TR</td>
</tr>
</tbody>
</table>

Localisation of a mthsp70 in Trachipleistophora

Confocal microscopy using an antibody to Th-mtHsp70

Infected cells

Spores

Rabbit

Western blot using an antibody to Th-mtHsp70

Trees are a good way of exploring the history of genes but care is needed!

- Making trees is not easy:
  - Among-site rate heterogeneity, “fast clock” species, shared nucleotide or amino acid composition biases
  - Different data sets may be affected by individual phenomena to different degrees
  - Biases need not be large if phylogenetic signal is weak

Phylogenetic analysis requires careful thought

- Phylogenetic analysis is frequently treated as a black box into which data are fed (often gathered at considerable cost) and out of which “The Tree” springs
- (Hillis, Moritz & Mable 1996, Molecular Systematics)