Using genome-level processes to solve tricky nodes in phylogenetics
USING GENOME-LEVEL PROCESSES TO SOLVE TRICKY NODES IN PHYLOGENETICS
Using genome-level processes to solve tricky nodes in phylogenetics

LBA
From unicellular to multicellular...
From radial symmetry to bilateral
Bilaterian innovations

(a) Radial symmetry

(b) Bilateral symmetry
The New Animal Phylogeny

Bilateria

Lophotrochozoa

Ecdysozoa

Deuterostomia

Porifera
Placozoa
Cnidaria
Ctenophora

Non bilateria
Bilaterian innovations

- Symmetry
- Cephalization
- Mesoderm

(a) Radial symmetry

(b) Bilateral symmetry
Zygote
Cleavage
Eight-cell stage
Cleavage
Blastocoel
Cross section of blastula
Blastula
Gastrulation
Blastocoel
Endoderm
Ectoderm
Archenteron
Cross section of gastrula
Blastopore
Mesoderm

Diploblastic gastrula

Triploblastic gastrula

Ectoderm
Endoderm
Mesoderm

Mesoderm
Embryonic layers

- **Endoderm:** digestive tract (i.e. guts) and organs derived from it (i.e. lungs) as well as secretory organs (liver, pancreas, thyroid, thymus...).

- **Ectoderm:** epidermis and nervous system, as well as hair, cornea, tooth enamel, etc.

- **Mesoderm:** coelom, muscles, bone/cartilage, circulatory, spleen, kidney, reproductive, etc.
Triggers: Ecological

Cambrian Substrate Revolution

- Kimberella grazing on bacterial mats
- Substrate is soupier
- Bilaterian burrows

Ediacaran Substrate

Cambrian Substrate

© Royal Ontario Museum
The New Animal Phylogeny

Non bilateria

Porifera
Placozoa
Cnidaria
Ctenophora

Lophotrochozoa

Ecdysozoa

Deuterostomia

Bilateria
First bilaterian
Libbie Hyman

The Invertebrates, 1940-1967
Polycladida
Polycladida
Nemertodermatida
Xenoturbellida

The Swedish clog
Acoels, nemertodermatids, and xenoturbellids

Acoela + Nemertodermatida = Acoelomorpha

Xenoturbellida + Acoelomorpha = Xenacelomorpha
THE MOLECULAR ERA
Molecules as Documents of Evolutionary History

Emile Zuckermandl and Linus Pauling
Molecular: 18S

18S rRNA secondary structure
*Diplomona omar* (Aplousobranchia)
Evidence from 18S Ribosomal DNA That the Lophophorates Are Protostome Animals

Kenneth M. Halanych, John D. Bacheller, Anna Marie A. Aguinaldo, Stephanie M. Liva, David M. Hillis, James A. Lake

Evidence for a clade of nematodes, arthropods and other moulting animals

Anna Marie A. Aguinaldo*, James M. Turbeville†, Lawrence S. Linford*, Maria C. Rivera*, James R. Garey‡, Rudolf A. Raff§ & James A. Lake*

Aguinaldo et al, 1997

Halanych et al, 1995

Molecular: 18S
The New Animal Phylogeny

Non bilateria
- Porifera
- Placozoa
- Cnidaria
- Ctenophora

Lophotrochozoa

Ecdysozoa

Bilateria

Deuterostomia
Deuterostomia

- Cephalochordates
- Urochordates (Ascidians)
- Vertebrates

Phylum Chordata

- Hemichordates
- Echinoderms
- Ambulacraria
Long Branch Attraction

Carranza et al, 1997
Long Branch Attraction

Real phylogeny

Inferred phylogeny

A

B

D

C

D

B

C

A
Avoiding LBA

Should we be worried about long-branch attraction in real data sets? Investigations using metazoan 18S rDNA

Frank E. Anderson, David L. Swofford

* Department of Zoology and Center for Systematic Biology, Southern Illinois University, Carbondale, IL 62901, USA
b Department of Biological Science and School of Computational Science and Information Technology, Florida State University, Tallahassee, FL 32306-4120, USA

Received 4 February 2004; revised 1 June 2004

A review of long-branch attraction

Johannes Bergsten

Department of Ecology and Environmental Science, Umeå University, SE-901 87 Umeå, Sweden

Accepted 14 February 2005

Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes

Jordi Paps, Jaume Baguñà and Marta Riutort

Departament de Genetica, Universitat de Barcelona, Avenida Diagonal, 645, 08028 Barcelona, Spain
Avoiding LBA

1. Use methods less sensitive to LBA, such as ML or BI
   • Use models less sensitive to LBA
   • In the case of AA and BI, use CAT over LG

\[
Q = \begin{bmatrix}
\alpha \pi_T & \beta \pi_C & \gamma \pi_G \\
\beta \pi_A & \rho \pi_C & \sigma \pi_G \\
\gamma \pi_A & \sigma \pi_T & \tau \pi_C \\
\end{bmatrix}
\]
Avoiding LBA

1. Use methods less sensitive to LBA, such as ML or BI

2. Model modifications (Gamma-distribution, covarion, recoding)
### Gamma distribution

<table>
<thead>
<tr>
<th>Sample</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_WEAU160</td>
<td>ATCAAGAGTGAGGAGTCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_U455</td>
<td>TACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_SF1703</td>
<td>TACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_92RW020.5</td>
<td>GACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_92UG031.7</td>
<td>GACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_92UG037.8</td>
<td>AGAAGTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_TZ017</td>
<td>CACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_UG275A</td>
<td>A.C.A.G.ACA.G.G.G.AA.AA.AA.A</td>
<td>39</td>
</tr>
<tr>
<td>A_UG273A</td>
<td>A.C.A.G.ACA.G.G.G.AA.AA.AA.A</td>
<td>39</td>
</tr>
<tr>
<td>A_DJ258A</td>
<td>TACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_KENYA</td>
<td>TACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_CARGAN</td>
<td>TACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_CARSAS</td>
<td>TACAACTGGAGAATGCACTGCACTTG</td>
<td>39</td>
</tr>
<tr>
<td>A_CAR4054</td>
<td>A.CACA.G.G.G.G.AA.AA.AA.AA.A</td>
<td>39</td>
</tr>
<tr>
<td>A_CAR286A</td>
<td>A.CACA.G.G.G.G.AA.AA.AA.AA.A</td>
<td>39</td>
</tr>
<tr>
<td>A_CAR4023</td>
<td>A.CACA.G.G.G.G.AA.AA.AA.AA.A</td>
<td>39</td>
</tr>
<tr>
<td>A_CAR423A</td>
<td>A.CACA.G.G.G.G.AA.AA.AA.AA.A</td>
<td>39</td>
</tr>
<tr>
<td>A_VI191A</td>
<td>A.CACA.G.G.G.G.AA.AA.AA.AA.A</td>
<td>39</td>
</tr>
</tbody>
</table>
### Gamma distribution

![Gamma distribution diagram]

#### Example Table

<table>
<thead>
<tr>
<th>Location</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>B_WEAU160</td>
<td>ATCAAGATGAAAGGGATCAGGAAGAATTATCAGCACTTG</td>
</tr>
<tr>
<td>A_U455</td>
<td>T........ACA...G........TTG........</td>
</tr>
<tr>
<td>A_SF1703</td>
<td>T........ACA...T........C.G......AA..A</td>
</tr>
<tr>
<td>A_92RW020.5</td>
<td>G........ACA...G....GG......AA..A</td>
</tr>
<tr>
<td>A_92UG031.7</td>
<td>G.A........ACA...G....GG......A</td>
</tr>
<tr>
<td>A_92UG037.8</td>
<td>T........AGA...G........TTG...K</td>
</tr>
<tr>
<td>A_TZ017</td>
<td>G.A........ACA...G............A</td>
</tr>
<tr>
<td>A_UG275A</td>
<td>A.C.T........ACA...G....GG......A</td>
</tr>
<tr>
<td>A_UG273A</td>
<td>A.CACA.T........G......AA..K</td>
</tr>
<tr>
<td>A_DJ258A</td>
<td>T........ACA ................A.T</td>
</tr>
<tr>
<td>A_KENYA</td>
<td>T........ACA ............A........A</td>
</tr>
<tr>
<td>A_CARGAN</td>
<td>T........ACA ................A...</td>
</tr>
<tr>
<td>A_CARSAS</td>
<td>T........ACA.CACA...G....GG...CA</td>
</tr>
<tr>
<td>A_CAR4054</td>
<td>A........CACA...G...GG...CA...A</td>
</tr>
<tr>
<td>A_CAR286A</td>
<td>A........CACA...G...GG...AA...A</td>
</tr>
<tr>
<td>A_CAR4023</td>
<td>A.............A.............A...A</td>
</tr>
<tr>
<td>A_CAR423A</td>
<td>A.............A.............A...A</td>
</tr>
<tr>
<td>A_VI191A</td>
<td>........ACA...T....GG.....A......</td>
</tr>
</tbody>
</table>

#### Matrix Representation

\[
Q = \begin{bmatrix}
\alpha \pi_T & \beta \pi_C & \gamma \pi_G \\
\alpha \pi_A & \rho \pi_C & \sigma \pi_G \\
\beta \pi_A & \rho \pi_T & \tau \pi_G \\
\gamma \pi_A & \sigma \pi_T & \tau \pi_C
\end{bmatrix}
\]

#### Gamma Distribution

![Gamma distribution graph]
Heterotachy & Covarion

\[
Q = \begin{pmatrix}
\alpha \pi_T & \beta \pi_C & \gamma \pi_G \\
\alpha \pi_A & \rho \pi_C & \sigma \pi_G \\
\beta \pi_A & \rho \pi_T & \tau \pi_G \\
\gamma \pi_A & \sigma \pi_T & \tau \pi_C
\end{pmatrix}
\]
Recoding

Giacomelli et al, 2022
Recoding

Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding

Anthony K. Redmond & Aoife McLysaght

Six-State Amino Acid Recoding is not an Effective Strategy to Offset Composition Heterogeneity and Saturation in Phylogenetic Analyses

Alexandra M. Hernandez & Joseph F. Ryan
Avoiding LBA

1. Use methods less sensitive to **LBA**, such as **ML** or **BI**

2. **Model modifications** (Gamma-distribution, covarion, recoding)

3. Using the **shortest branched representatives** available for each phyla (Relative Rate Test)
Long Branch Attraction

Carranza et al, 1997
Relative Rate Test
Avoiding LBA

1. Use methods less sensitive to LBA, such as ML or BI

2. **Model modifications** (Gamma-distribution, covarion, recoding)

3. Using the **shortest branched representatives** available for each phyla (Relative Rate Test)

4. Use **widest taxon sampling** within the long branch group
Avoiding LBA

1. Use methods less sensitive to LBA, such as ML or BI

2. Model modifications (Gamma-distribution, covarion, recoding)

3. Using the shortest branched representatives available for each phyla (Relative Rate Test)

4. Use widest taxon sampling within the long branch group

5. Use closely related outgroups
Long Branch Attraction

Adl et al, 2012
Acoel Flatworms: Earliest Extant Bilaterian Metazoans, Not Members of Platyhelminthes

Iñaki Ruiz-Trillo,¹ Marta Riutort,¹ D. Timothy J. Littlewood,² Elisabeth A. Herniou,² Jaume Baguñà¹*

Molecular: 18S
Avoiding LBA

1. Use methods less sensitive to **LBA**, such as **ML** or **BI**

2. **Model modifications** (Gamma-distribution, covarion, recoding)

3. Using the **shortest branched representatives** available for each phyla (Relative Rate Test)

4. Use **widest taxon sampling** within the long branch group

5. Use **closely related outgroups**

6. **Compartmentalize** problematic lineages (extremely long branches)
Acoels, nemertodermatids, and xenoturbellids

Acoela + Nemertodermatida = Acoelomorpha
**Taxon sampling in 18S**

**Zrzavý et al, 1998**
MP: 18S + Morphology  
151 seqs  
28 phyla bilaterians  
Consensus sequences

**Giribet et al, 2000**
POY: 18S + Morphology  
147 spp  
28 phyla bilaterians

**Peterson & Eernisse, 2001**
MP: 18S + Morphology  
304 seqs  
28 phyla bilaterians  
"AGG" removed
Avoiding LBA

1. Use methods less sensitive to **LBA**, such as **ML** or **BI**

2. **Model modifications** (Gamma-distribution, covarion, recoding)

3. Using the **shortest branched representatives** available for each phyla (Relative Rate Test)

4. Use **widest taxon sampling** within the long branch group

5. Use **closely related outgroups**

6. **Compartmentalize** problematic lineages (extremely long branches)

7. **More data!!!**
THE MOLECULAR ERA: MORE MARKERS...
New markers

Giribet, 2000
New markers

A phylogenetic analysis of myosin heavy chain type II sequences corroborates that Acoela and Nemertodermatida are basal bilaterians


Ruiz-Trillo et al, 2002
18S + Myosin

Combined large and small subunit ribosomal RNA phylogenies support a basal position of the acoelomorph flatworms

Maximilian J. Telford**, Anne E. Lockyer*, Chloë Cartwright-Finch† and D. Timothy J. Littlewood‡

Telford et al, 2003
18S + 28S

Bilaterian Phylogeny Based on Analyses of a Region of the Sodium–Potassium ATPase α-Subunit Gene

Frank E. Anderson,†,‡ Alonso J. Córdoba,‡ Mikael Thollesson†,∥

Anderson et al, 2004
ATPase alpha

Lophotrochozoa internal phylogeny: new insights from an up-to-date analysis of nuclear ribosomal genes

Jordi Paps, Jaume Baguñà and Marta Riutort*

Paps et al, 2009
18S + 28S
Long Branch Attraction

Real phylogeny

Inferred phylogeny

A
B
D
C

D
B

C
A
Poor taxon sampling

28S, Mallat & Whinchell 2002
Orthology concerns

- Lineage Sorting
- Horizontal Gene Transfer
- Gene Duplication and Loss
- Hybridization
- Recombination
13 Genes

- 27 Phyla
- 90 OTU
- 8,800 bp
- ML & Mr Bayes

Xenoturbella is a deuterostome that eats molluscs

Sarah J. Boulrat¹, Claus Nielsen², Anne E. Lockyer³, D. Timothy J. Littlewood³ & Maximilian J. Telford¹

Paps et al 2009b
Deuterostomia

Cephalochordates

Urochordates (Ascidians)

Vertebrates

Phylum Chordata

Hemichordates

Ambulacraria

Echinoderms

Xenoturbella
Lophotrochozoa

Porifera
Placozoa
Cnidaria
Ctenophora

Acoelomorpha

Lophotrochozoa

Ecdysozoa

Bilateria

Nephrozoa

Xenoturbella

Deuterostomia
The last march of the PCR...

<table>
<thead>
<tr>
<th>Publication</th>
<th>Phylogenetic position</th>
<th>Dataset</th>
<th>Statistical support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carranza et al. (1997)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA</td>
<td>67 % NJ BS&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ruiz-Trillo et al. (1999)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA</td>
<td>100 % ML BS&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ruiz-Trillo et al. (2002)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, myosin heavy chain type II</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Jondelius et al. (2002)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA</td>
<td>92%/82 % ML BS&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Telford et al. (2003)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, 28S rDNA</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ruiz-Trillo et al. (2004)</td>
<td>First-splitting bilaterians</td>
<td>mtDNA</td>
<td>97 % QP&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Glenner et al. (2004)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA; morphology</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Wallberg et al. (2007)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, 28S rDNA</td>
<td>1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Philippe et al. (2007)</td>
<td>Deuterostomes</td>
<td>68 protein-coding genes</td>
<td>0.34 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Dunn et al. 2008</td>
<td>Lophotrochozoa</td>
<td>150 protein-coding genes</td>
<td></td>
</tr>
<tr>
<td>Paps et al. (2009a)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, 28S rDNA</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Paps et al. (2009b)</td>
<td>First-splitting bilaterians</td>
<td>13 nuclear markers</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Heijnol et al. (2009)</td>
<td>First-splitting bilaterians</td>
<td>1500 protein-coding genes</td>
<td>66 % ML BS&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mwinyi et al. (2010)</td>
<td>First-splitting bilaterians</td>
<td>mtDNA</td>
<td>0.91 BI BPP&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mallatt et al. (2010)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, 28S rDNA</td>
<td>&gt;90 % ML BS&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Philippe et al. (2011a)</td>
<td>Deuterostomes</td>
<td>197 protein-coding genes and mtDNA</td>
<td>63 % BI BS&lt;sup&gt;g&lt;/sup&gt; and 0.99 BI BPP</td>
</tr>
<tr>
<td>Ryan et al. (2013)</td>
<td>First-splitting bilaterians</td>
<td>242 protein-coding genes</td>
<td>38–52 % ML BS&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>Srivastava et al. (2014)</td>
<td>First-splitting bilaterians</td>
<td>Different sets, ranging from 69 to 442</td>
<td>96 % ML BS and 1.0 BI-WAG BPP&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
</tbody>
</table>


The start of the phylotranscriptomics age

Table 1: A list of some of the molecular phylogenies dealing with the position of acoelomorphs, their results, the dataset and the statistical support

<table>
<thead>
<tr>
<th>Publication</th>
<th>Phylogenetic position</th>
<th>Dataset</th>
<th>Statistical support</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carranza et al. (1997)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA</td>
<td>67 % NJ BS&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ruiz-Trillo et al. (1999)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA</td>
<td>100 % ML BS&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ruiz-Trillo et al. (2002)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, myosin heavy chain type II</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Jondelius et al. (2002)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA</td>
<td>92 %/82 % ML BS&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Telford et al. (2003)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, 28S rDNA</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Ruiz-Trillo et al. (2004)</td>
<td>First-splitting bilaterians</td>
<td>mtDNA</td>
<td>97 % QP&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Glenner et al. (2004)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA; morphology</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Wallberg et al. (2007)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, 28S rDNA</td>
<td>1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Philippe et al. (2007)</td>
<td>Deuterostomes</td>
<td>68 protein-coding genes</td>
<td>0.34 BI BPP&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Dunn et al. 2008</strong></td>
<td>Lophotrochozoa</td>
<td>150 protein-coding genes</td>
<td>–</td>
</tr>
<tr>
<td>Paps et al. (2009a)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, 28S rDNA</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td>Paps et al. (2009b)</td>
<td>First-splitting bilaterians</td>
<td>13 nuclear markers</td>
<td>1.0/1.0 BI BPP&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Hejnol et al. (2009)</strong></td>
<td>First-splitting bilaterians</td>
<td>1500 protein-coding genes</td>
<td>66 % ML BS&lt;sup&gt;f&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mwinyi et al. (2010)</td>
<td>First-splitting bilaterians</td>
<td>mtDNA</td>
<td>0.91 BI BPP&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Mallatt et al. (2010)</td>
<td>First-splitting bilaterians</td>
<td>18S rDNA, 28S rDNA</td>
<td>&gt; 90 % ML BS&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>Philippe et al. (2011a)</td>
<td>Deuterostomes</td>
<td>197 protein-coding genes and mtDNA</td>
<td>63 % BI BS&lt;sup&gt;g&lt;/sup&gt; and 0.99 BI BPP</td>
</tr>
<tr>
<td>Ryan et al. (2013)</td>
<td>First-splitting bilaterians</td>
<td>242 protein-coding genes</td>
<td>38–52 % ML BS&lt;sup&gt;h&lt;/sup&gt;</td>
</tr>
<tr>
<td>Srivastava et al. (2014)</td>
<td>First-splitting bilaterians</td>
<td>Different sets, ranging from 69 to 442</td>
<td>96 % ML BS and 1.0 BI-WAG BPP&lt;sup&gt;i&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Ruiz-Trillo and Paps 2015
PHYLOTOKTRANSCRIPTOMICS
Phylogenomics

Philippe 2005
146 genes
35,000 chars
10 phyla
Broad phylogenomic sampling improves resolution of the animal tree of life

Casey W. Dunn, Andreas Hejnol, David Q. Matus, Kevin Pang, William E. Browne, Stephen A. Smith, Elaine Seaver, Greg W. Rouse, Matthias Obst, Gregory D. Edgecombe, Martin V. Sørensen, Steven H. D. Haddock, Andreas Schmidt-Rhaesa, Akiko Okusu, Reinhardt Meiberg Kristensen, Ward C. Wheeler, Mark Q. Martin, and Gonzalo Giribet

Dunn et al. (2008)
150 genes or 21,000 chars
26 phyla

Phylogenomics 2008

Gonzalo Giribet
Casey W. Dunn
Andreas Hejnol
Phylogenomics: Acoelomorpha

Hejnol et al. (2009)
1,500 genes or 270,000 chars
27 phyla
Acoels, nemertodermatids, and xenoturbellids

Acoela + Nemertodermatida = Acoelomorpha

Xenoturbellida + Acoelomorpha = Xenacelomorpha
Hejnol et al, 2009
1,500 genes
270,000 chars

Philippe et al, 2011
197 markers
38,000 chars
CAT model
Deuterostomia

Cephalochordates

Urochordates (Ascidians)

Vertebrates

Phylum Chordata

Hemichordates

Ambulacraria

Echinoderms

Xenacoelomorpha
Deuterostome synapomorphy: Gill slits
Miniature deuterostomes still have coelom

An Anatomical Description of a Miniaturized Acorn Worm (Hemichordata, Enteropneusta) with Asexual Reproduction by Paratomy

Katrine Worsaae¹*, Wolfgang Sterrer², Sabrina Kaul-Strehlow³, Anders Hay-Schmidt⁴, Gonzalo Giribet⁵
Philippe et al, 2011
197 markers
38,000 chars
30% missing data
CAT model
Xenacoelomorpha is the sister group to Nephrozoa

Johanna Taylor Cannon¹, Bruno Cossermelli Vellutini², Julian Smith III³, Fredrik Ronquist¹, Ulf Jondelius¹ & Andreas Hejnol²

LETTER
doi:10.1038/nature16520

New deep-sea species of *Xenoturbella* and the position of Xenacoelomorpha

Greg W. Rouse¹, Nerida G. Wilson¹,²,³, Jose I. Carvajal¹ & Robert C. Vrijenhoek⁴

LETTER
doi:10.1038/nature16545
Mystery of deep-sea 'purple sock' solved

By Rebecca Morelle
Science Correspondent, BBC News

3 February 2016 | Science & Environment

Research Highlight: Deep-Sea Churro Finds its Evolutionary Home
**Xenacoelomorpha 2016**

- Cannon et al, 2016
- 217 markers
- 45,000 chars
- CAT-GTR + Gamma
Laumer et al, 2019
422 markers
53,000 chars
60% completeness
167 taxa
Almost all phyla
CAT-GTR + Gamma model
Xenambulacrararia 2019
Paraphyletic deuterostomes

Philippe et al, 2019
1,173 markers
350,000 chars
CAT-GTR model
Xenambulacraria
2019

Philippe et al, 2019
1,173 markers
350,000 chars
CAT-GTR model
Filtering artifactual signal increases support for Xenacoelomorpha and Ambulacraria sister relationship in the animal tree of life

Highlights
- High rate of violation of incontrovertible splits in animal phylogenomic datasets
- Filtering gene families accordingly increases phylogenetic information
- Increased, but not conclusive, support for the existence of Xenambulacraria

Authors
Peter O. Mulhair, Charley G.P. McCarthy, Karen Siu-Ting, Christopher J. Creevey, Mary J. O’Connell

Correspondence
mbzmjo@nottingham.ac.uk
Filtering artifactual signal increases support for Xenacoelomorpha and Ambulacraria sister relationship in the animal tree of life

Highlights
- High rate of violation of incontrovertible splits in animal phylogenomic datasets
- Filtering gene families accordingly increases phylogenetic information
- Increased, but not conclusive, support for the existence of Xenambulacraria

Authors
Peter O. Mulhair, Charley G.P. McCarthy, Karen Siu-Ting, Christopher J. Creevey, Mary J. O’Connell

Correspondence
mbzmjo@nottingham.ac.uk

The slowly evolving genome of the xenacoelomorph worm
Xenoturbella bocki

*Corresponding authors: p.schiffer@uni-koeln.de, m.telford@nottingham.ac.uk

Lack of support for Deuterostomia prompts reinterpretation of the first Bilateria

Topology-dependent asymmetry in systematic errors affects phylogenetic placement of Ctenophora and Xenacoelomorpha
Paschalia Kapli and Maximilian J. Telford

SCIENCE ADVANCES | RESEARCH ARTICLE
EVOLUTIONARY BIOLOGY
Philippe et al, 2019
1,173 markers
350,000 chars
CAT-GTR model
## Patchy matrices

<table>
<thead>
<tr>
<th>A</th>
<th><img src="image" alt="Main matrix" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th><img src="image" alt="Images 1 to 10" /></th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>C</th>
<th><img src="image" alt="Images C" /></th>
</tr>
</thead>
</table>

---

Hartman and Vision, 2008
COMPARATIVE GENOMICS
2010 Postdoc
Genomic content as phylogenetic marker

Rare Genomic Changes (RGCs)

(a) Intron indels
- Intron insertion
  - Flies
  - Butterflies
  - Strepsiptera

(b) Retroposon integrations
- SINE insertions
  - Whales
  - Hippos
  - Other Artiodactyla

(c) Signatures sequences
- Lox5 peptide signature
  - Dicynid mesozoa
    - Flatworms
    - Other Lophotrochozoa
  - Ecdysozoa

(d) mtDNA code variants
- Codon reassignment ATA = Ile
  - Echinoderms
  - Hemichordates
  - Amphioxus

(e) mtDNA gene order
- mtDNA gene order change
  - Insects
  - Crustaceans
  - Myriapods

(f) Gene duplications
- Gene duplication
  - Jawed vertebrates
  - Lampreys
  - Amphioxus

Rokas & Holland 2000
TRE
2010 Postdoc

2015 Lecturer

2019 Senior Lecturer
Widespread patterns of gene loss in the evolution of the animal kingdom

Cristina Guijarro-Clarke¹, Peter W. H. Holland⁰² and Jordi Paps⁰¹³
The Origin of Land Plants Is Rooted in Two Bursts of Genomic Novelty

Graphical Abstract

New genes involved in multicellularity
New genes involved in terrestrialisation

Chlorophyta
Charophyta
Bryophytes
Vascular plants
Land plants
Streptophyta

Authors
Alexander M.C. Bowles, Ulrike Bechtold, Jordi Paps

Correspondence
ubech@essex.ac.uk (U.B.), jordi.paps@bristol.ac.uk (J.P.)

In Brief
Bowles et al. show that two consecutive bursts of genomic novelty predate the origin of land plants. Identifying these events provides insights into the evolution of flora that has defined modern ecosystems.
Water-related innovations in land plants evolved by different patterns of gene cooption and novelty

Alexander M. C. Bowles$^{1,2}$, Jordi Paps$^{1,3}$ and Ulrike Bechtold$^{1,4}$

$^1$School of Life Sciences, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK; $^2$School of Geographical Sciences, University of Bristol, University Road, Bristol, BS8 1RL, UK; $^3$School of Biological Sciences, University of Bristol, 24 Tyndall Avenue, Bristol, BS8 1TQ, UK; $^4$Present address: Department of Biosciences, Durham University, South Road, Durham, DH1 3LE, UK.
Using genome-level processes to solve tricky nodes in phylogenetics
THERE IS NO PLANET B
Two major issues

- Orthology issues derived from “phylotranscriptomics”
- ESTs/RNAseq only represent a fraction of the genes in a genome, limited to a tissue and/or developmental stage.
Two major issues

- Orthology issues derived from “phylotranscriptomics”
  - ESTs/RNAseq only represent a fraction of the genes in a genome, limited to a tissue and/or developmental stage.
  - Genes are often incomplete and fragmented
  - More prone to contamination than genomes
Homology and genes

Orthologous genes: a product of speciation

Paralogous genes: duplication within a genome

- Orthologs: frog α, chick α, mouse α
- Paralogs: mouse β, chick β, frog β

Gene duplication:

early globin gene
Orthologous genes: a product of speciation

Paralogous genes: duplication within a genome

Hidden paralogy

Homologs

Orthologs

Paralogs

Orthologs

Frog α

Chick α

Mouse β

α-chain gene

β-chain gene

gene duplication

Early globin gene
Orthology issues

- **Lineage Sorting**
- **Horizontal Gene Transfer**
- **Gene Duplication and Loss**
- **Hybridization**
- **Recombination**
Two major issues

• Orthology issues derived from “phylotranscriptomics”
  • ESTs/RNAseq only represent a fraction of the genes in a genome, limited to a tissue and/or developmental stage.
  • Genes are often incomplete and fragmented
  • More prone to contamination than genomes

• Systematic errors (e.g., LBA) cause by evolutionary model misspecification applied to an alignment
The approaches

- Orthology issues -> use only complete genomes
- Systematic errors -> embrace genome-level processes!
Scenario 3

Xenambulacraria & paraphyletic deuterostomes

Scenario 4

Xenoturbella sister to Ambulacraria
Sampling

• 100+ genomes, no transcriptomes
• 91 BUSCO filtered -> OrthoFinder
• Different outgroups (Holozoa, Metazoa, Cnidaria Bilateria)

16 non-animal holozoan species
24 non-Bilateria
  6 Porifera
  2 Ctenophora
  2 Placozoa
  **14 Cnidaria**
12 Lophotrochozoa
13 Ecdysozoa
**7 Ambulacraria**
13 Chordata
**7 Xenacoelomorpha**
  1 Xenoturbellida
  4 Acoela
  2 Nemertodermatida
Canary in the coal mine

Bilateria

Porifera
Ctenophora
Placozoa
Cnidaria

Xenacoelomorpha
Lophotrochozoa
Ecdysozoa
Deuterostomia
Gene content for 549,544 gene families

Modes of gene family evolution (duplications, transfers, and losses; DTL)

Multispecies coalescence model (MSC)
Gene content

- **Character matrix** records presence/absence of traits
- Then this data is used to infer the tree

(a) Character table

<table>
<thead>
<tr>
<th>CHARACTERS</th>
<th>TAXA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lancelet (outgroup)</td>
<td>0</td>
</tr>
<tr>
<td>Lamprey</td>
<td>1</td>
</tr>
<tr>
<td>Bass</td>
<td>1</td>
</tr>
<tr>
<td>Frog</td>
<td>1</td>
</tr>
<tr>
<td>Turtle</td>
<td>1</td>
</tr>
<tr>
<td>Leopard</td>
<td>1</td>
</tr>
</tbody>
</table>

Vertebral column (backbone) | 0 | 1 | 1 | 1 | 1 | 1 |
Hinged jaws                | 0 | 0 | 1 | 1 | 1 | 1 |
Four walking legs           | 0 | 0 | 0 | 1 | 1 | 1 |
Amnion                     | 0 | 0 | 0 | 0 | 1 | 1 |
Hair                        | 0 | 0 | 0 | 0 | 0 | 1 |

(b) Phylogenetic tree

- **Lancelet** (outgroup)
- Lamprey
- Bass
- Frog
- Turtle
- Leopard

Character matrix records presence/absence of traits. Then this data is used to infer the tree.
Gene content: 549,544 gene families

Orthogroups

Presence / Absence Matrix

OG1
OG2
OG3
OG4

Markov Chain (CTCM) & Dollo models

Correction for unobserved gene losses

RevBayes Phylogeny
Duplications, Transfers, Losses (DTLs)

- **ALE**: cannot infer tree, but can compare topologies statically
- **SpeciesRax**: infers tree from DTL in all the gene trees of all the genes in all genomes
ASTRAL: infers all gene trees of all the genes in all genomes using the MSC
Then it summarises all the gene trees into one.
Phylogenomics: three approaches

1. Maximise taxon sampling.
3. Minimise paralogy.
Phylogenomics: three approaches

- **ML**: IQTree, partitions, modelfinder (LG + C60), 1000 bootstraps

- **BI**: Phylobayes, CAT-GTR, all chains converged (0.1 maxdiff)
Philippe et al, 2019
1,173 markers
350,000 chars
Philippe et al, 2019
1,173 markers
350,000 chars

This study
14,000 markers
5,300,000 chars
x15 larger
1. **Maximise taxon sampling.**
   - 14k genes and 5.3 M positions

2. **Minimise noise: ALE25T**
   - 6.4k genes and 3M positions

3. **Minimise paralogy**
   - Single-copy filtered (3k genes and 1M, and smaller)