



ForBio – Course 2021

Phylogenomics



Torsten Struck – t.h.struck@nhm.uio.no
NHM UiO

Phylogenomics

- Original definition by Eisen (1998):
Improving Functional Predictions for Uncharacterized Genes by Evolutionary Analysis (Gene function prediction)
- Gene family evolution
- Most commonly used nowadays:
Reconstructing phylogenetic relationships using genomes or large part of the genome (e.g., transcriptomes)
- Studying horizontal gene transfer

The era of “-omics”

- Genomics
- Transcriptomics
- Proteomics

The era of “-omics”

- Genomics
- Transcriptomics
- Proteomics

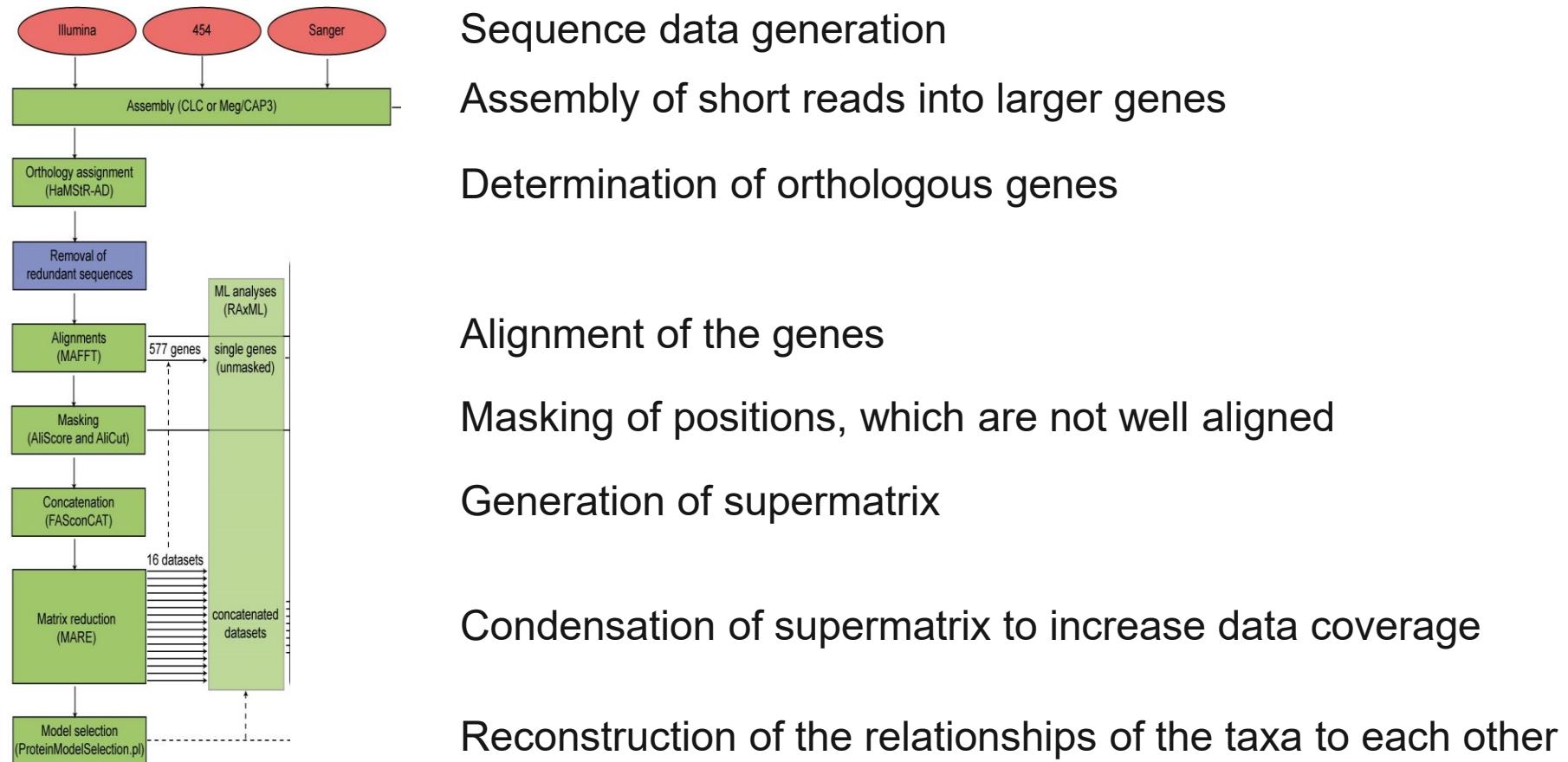


Next generation sequencing:

- 454 pyrosequencing
- Illumina (Solexa) sequencing
- SOLiD sequencing
- Ion semiconductor sequencing (Ion torrent)
- Single molecule real time sequencing (Pacific Biosciences)
- Nanopore DNA sequencing



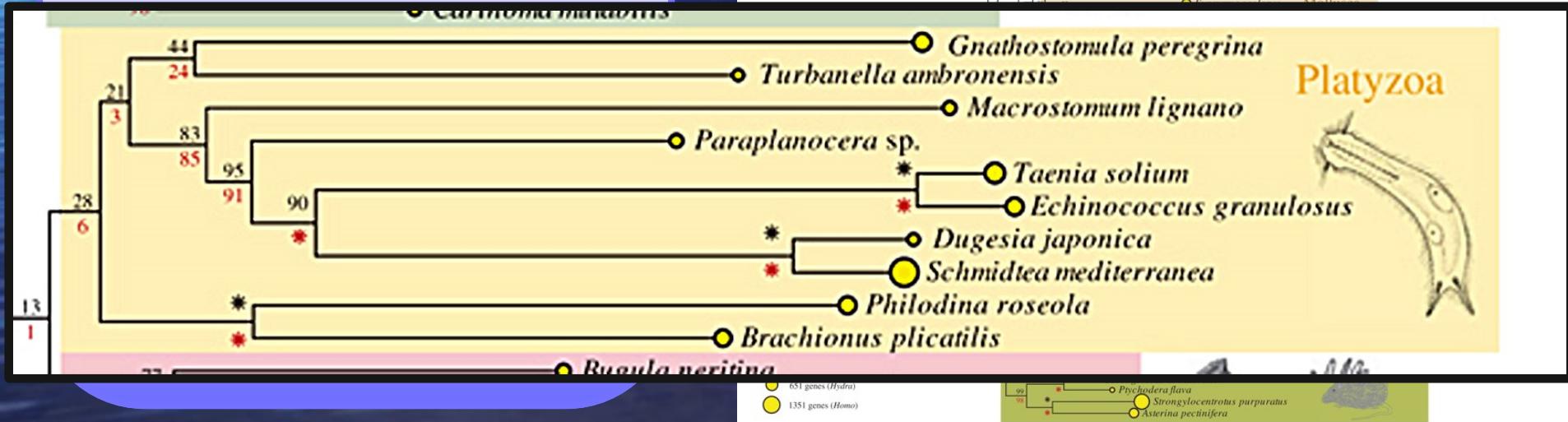
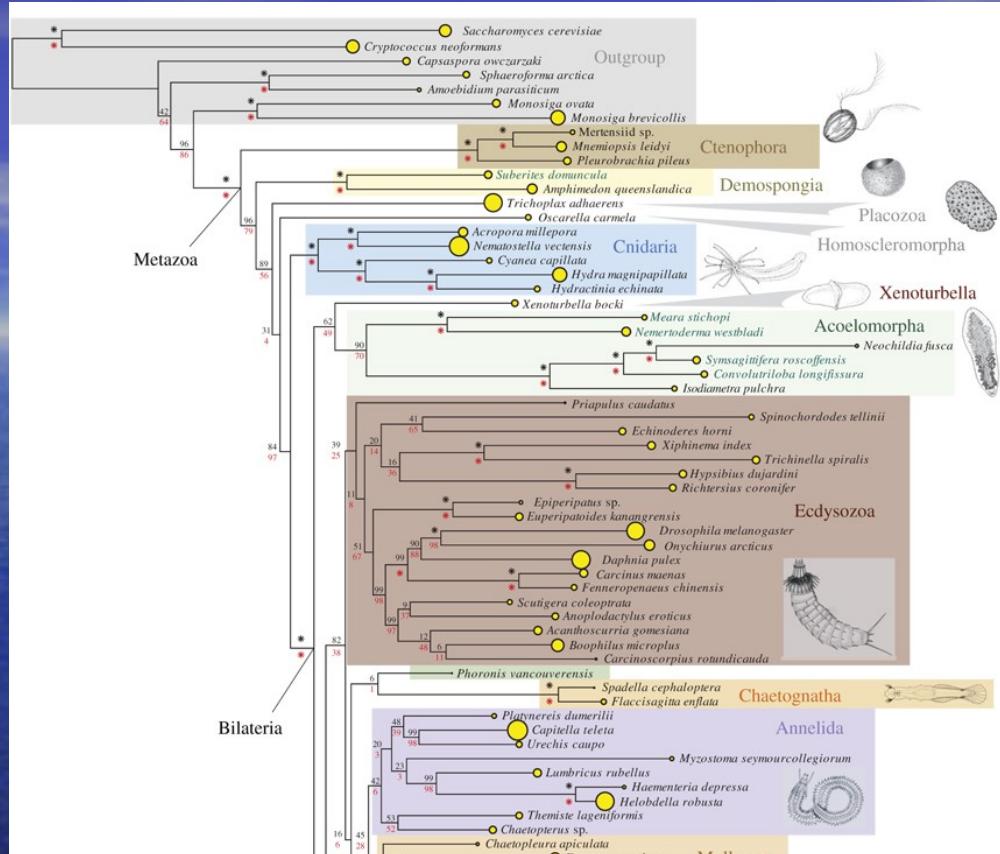
The standard procedure



Phylogenomics

- Monophyly of Platyzoa
- Except for Platyhelminthes, only one or two species

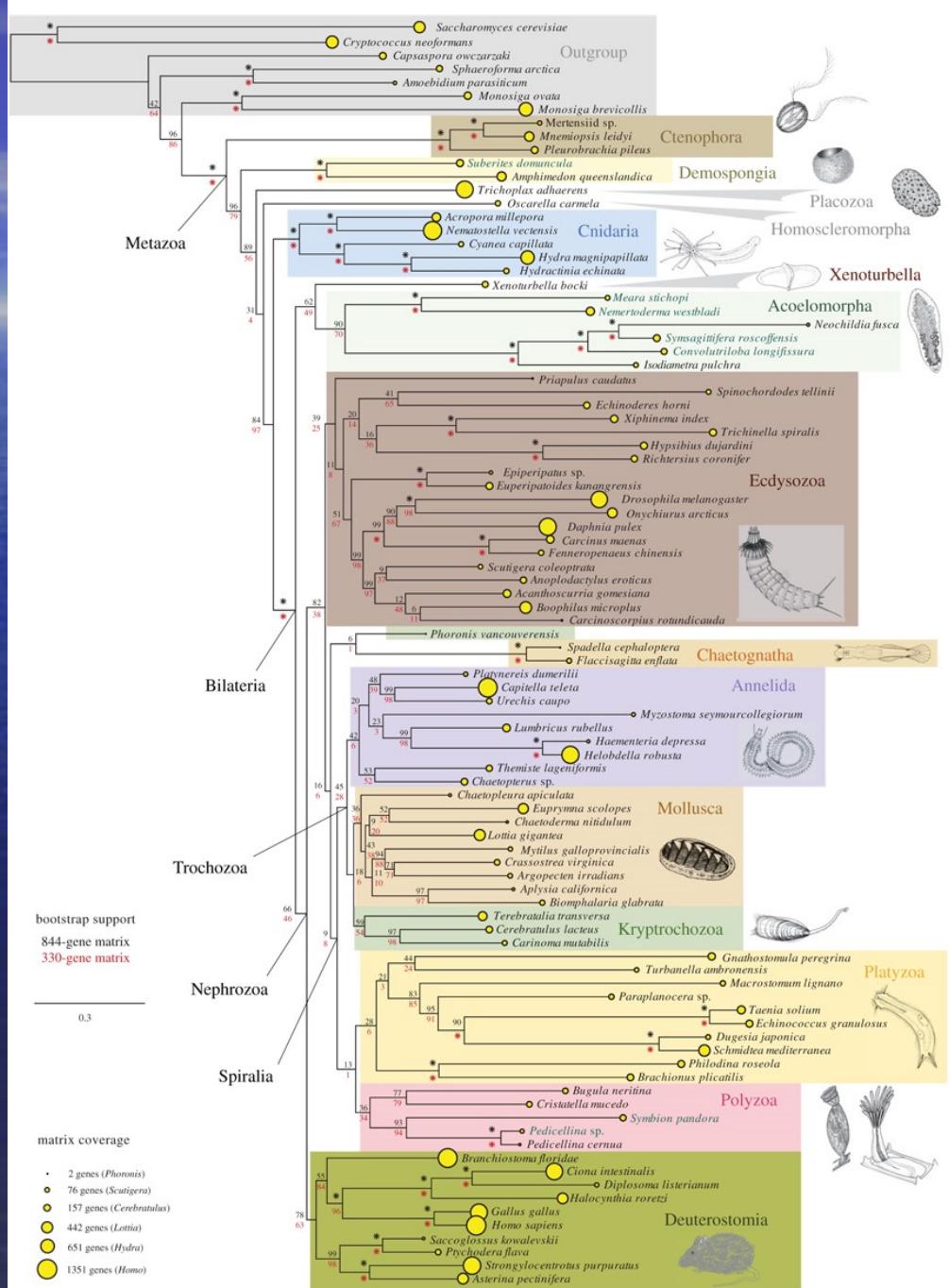
Hejnol et al. (2009). Proc. R. Soc. B: Biol. Sci.



Phylogenomics

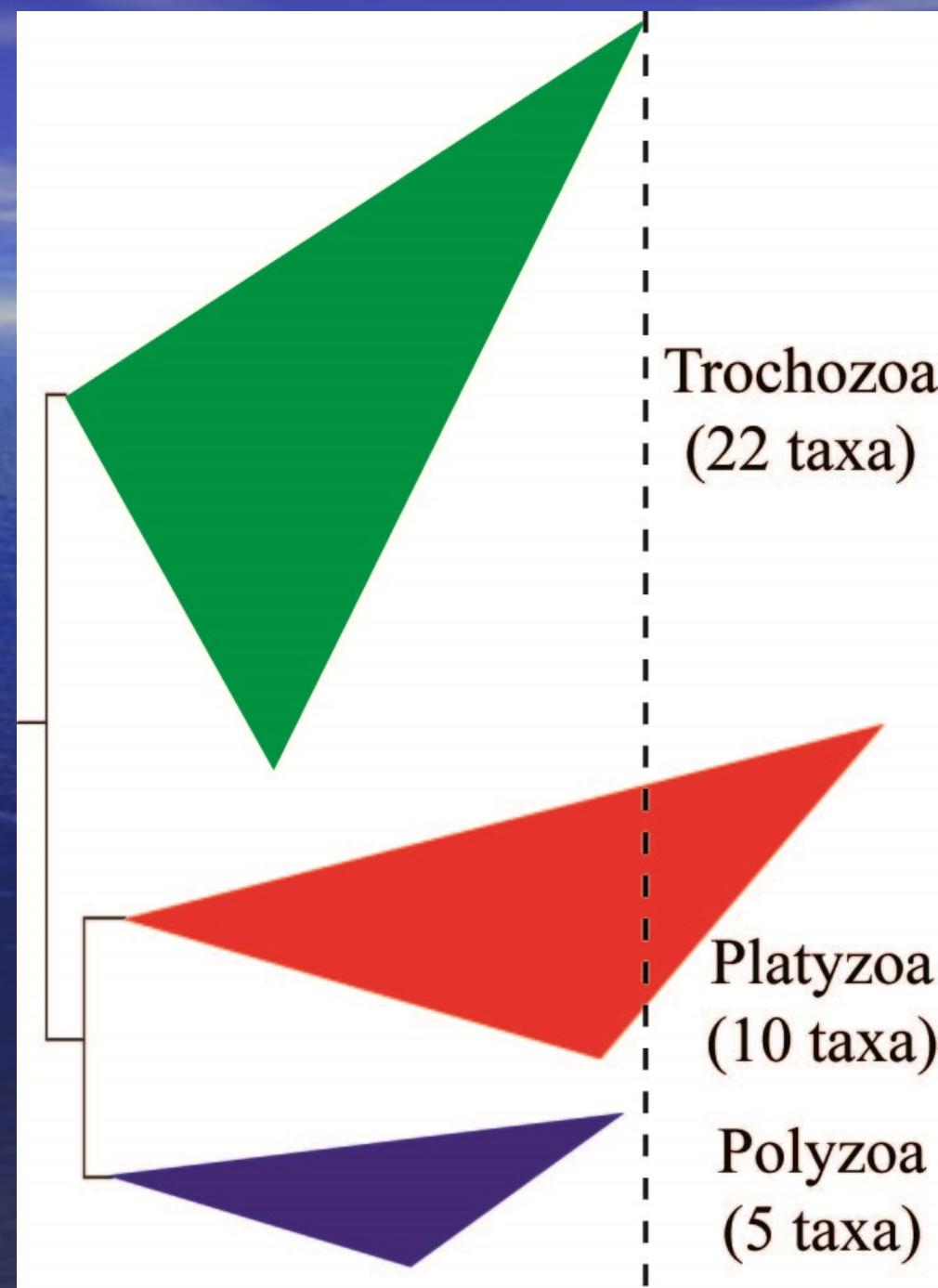
- Monophyly of Platyzoa
- Except for Platyhelminthes, only one or two species
- Small libraries
- Low coverage (74 out of 1,487 genes)

Hejnol et al. (2009). Proc. R. Soc. B: Biol. Sci.

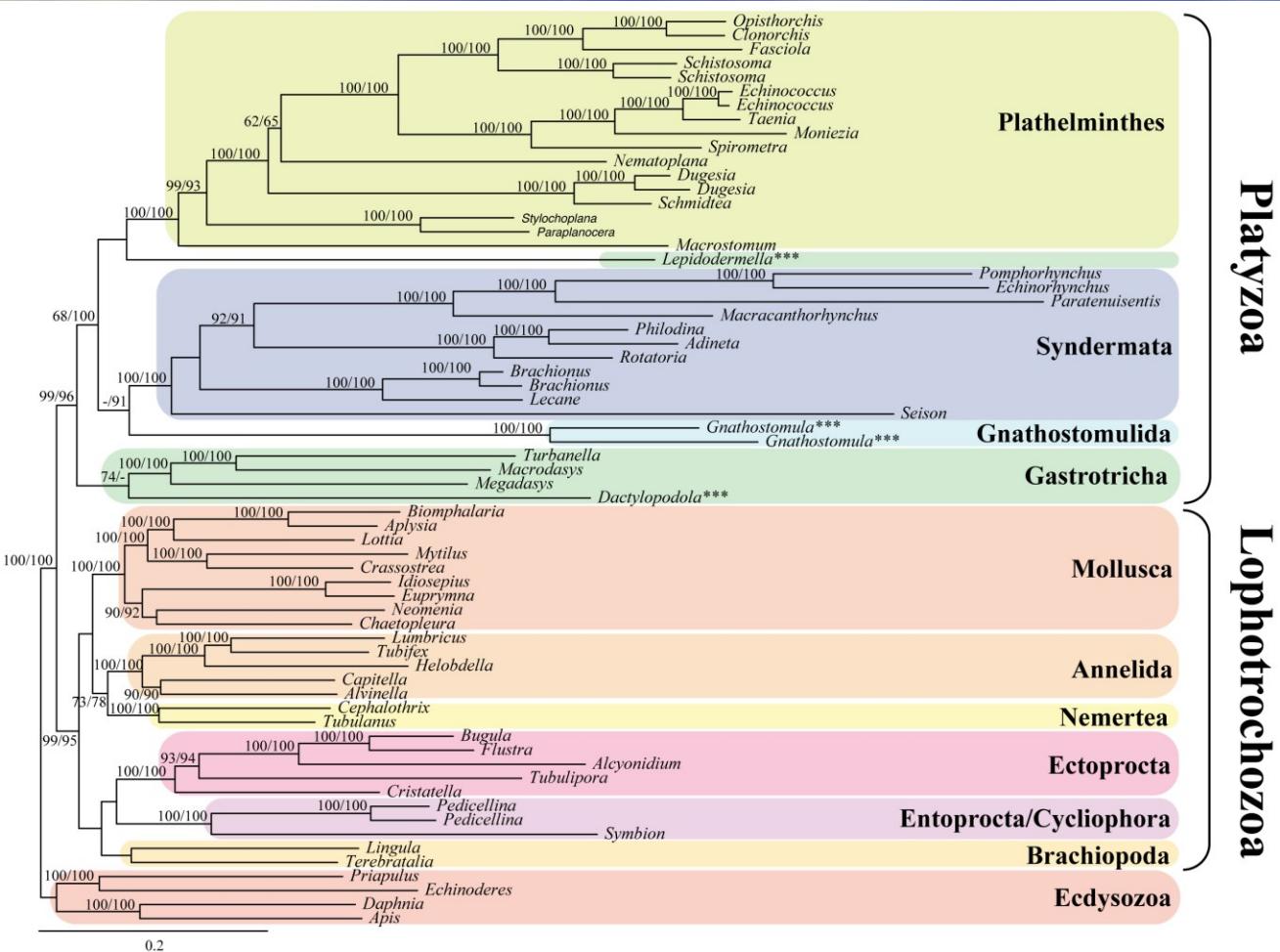


Phylogenomics

- Monophyly of Platyzoa
- Except for Platyhelminthes, only one or two species
- Small libraries
- Low coverage (74 out of 1,487 genes)
- Long branches



Brute-force phylogenomics

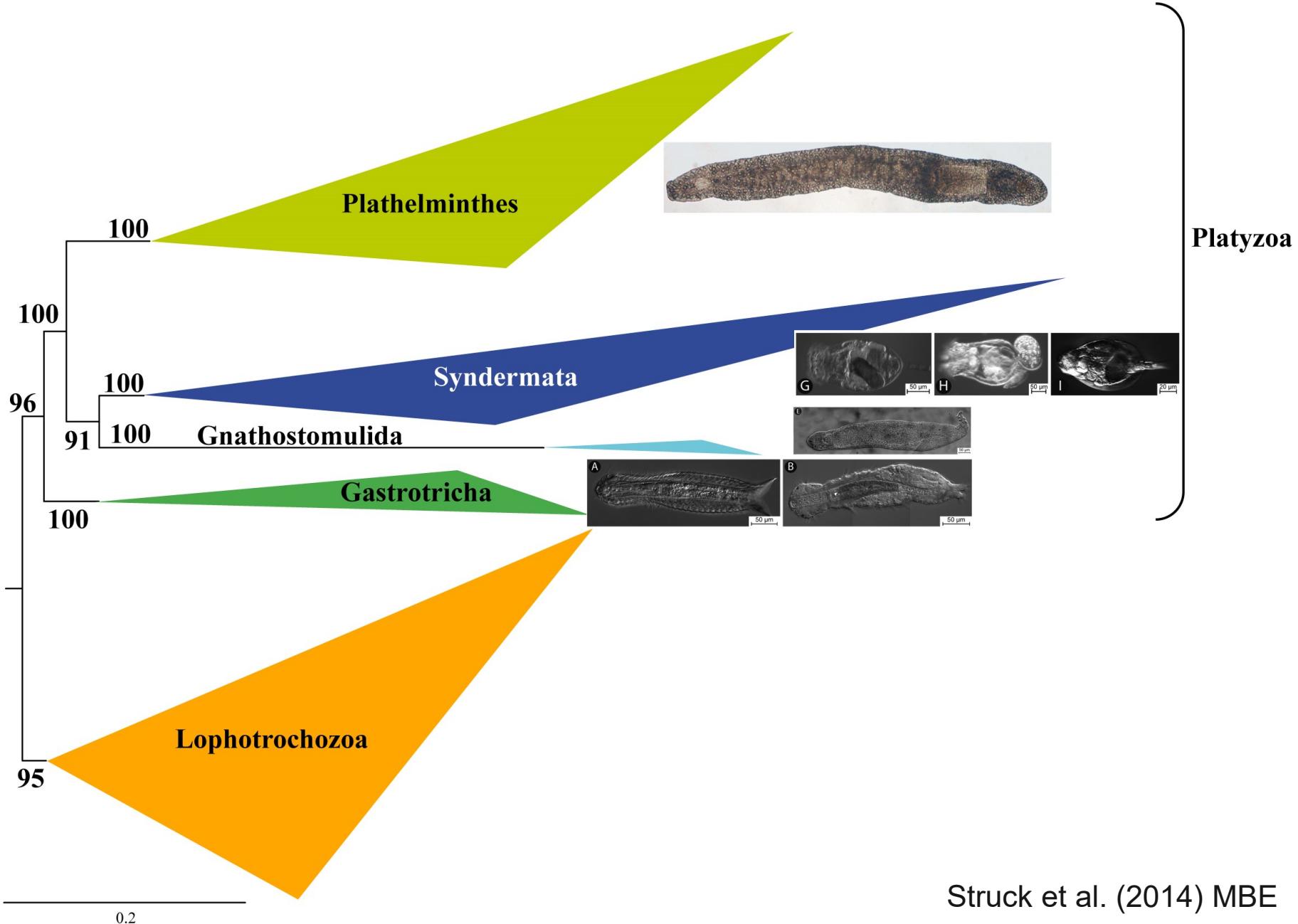


559 ortholog genes
82162 aa
35.7% coverage/taxon

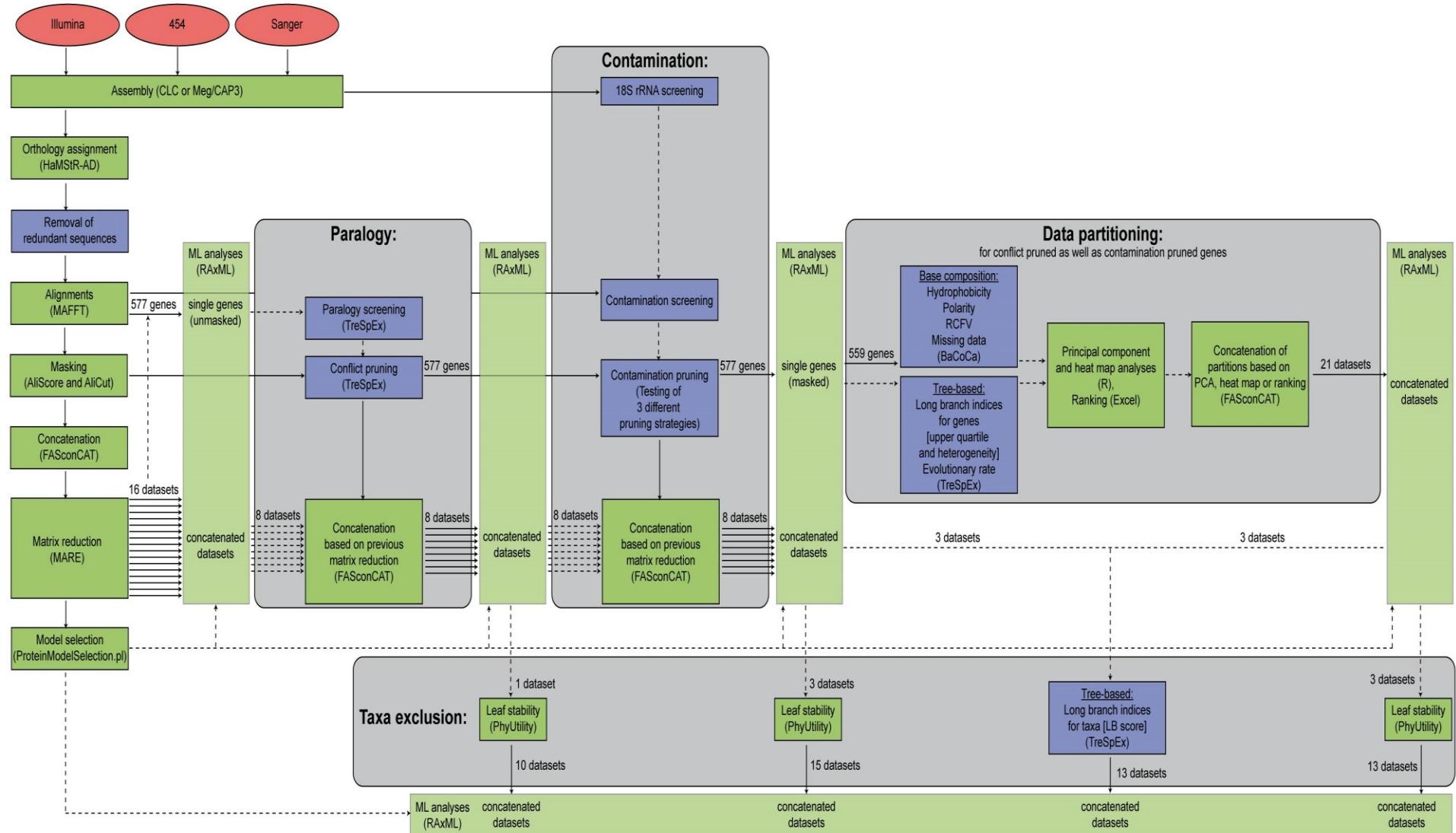
RAxML
Only bootstrap values $\geq 70\%$
Sensitivity analyses
*** non-stable taxa

Latter BS values without *Lepidodermella* and *Dactylopodola* due to long-branch issues

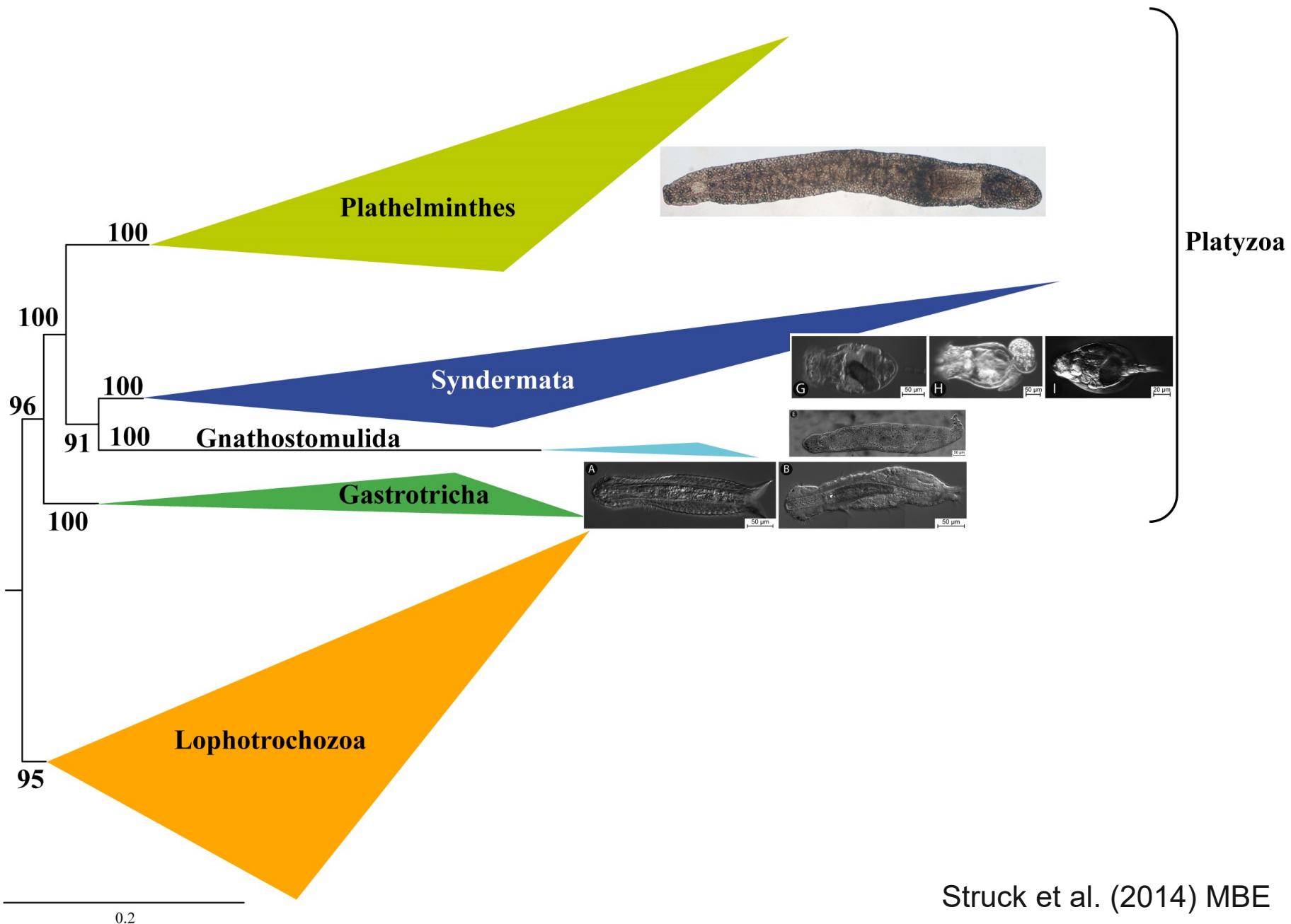
Brute-force phylogenomics



Going beyond the standard



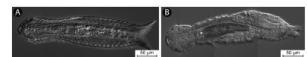
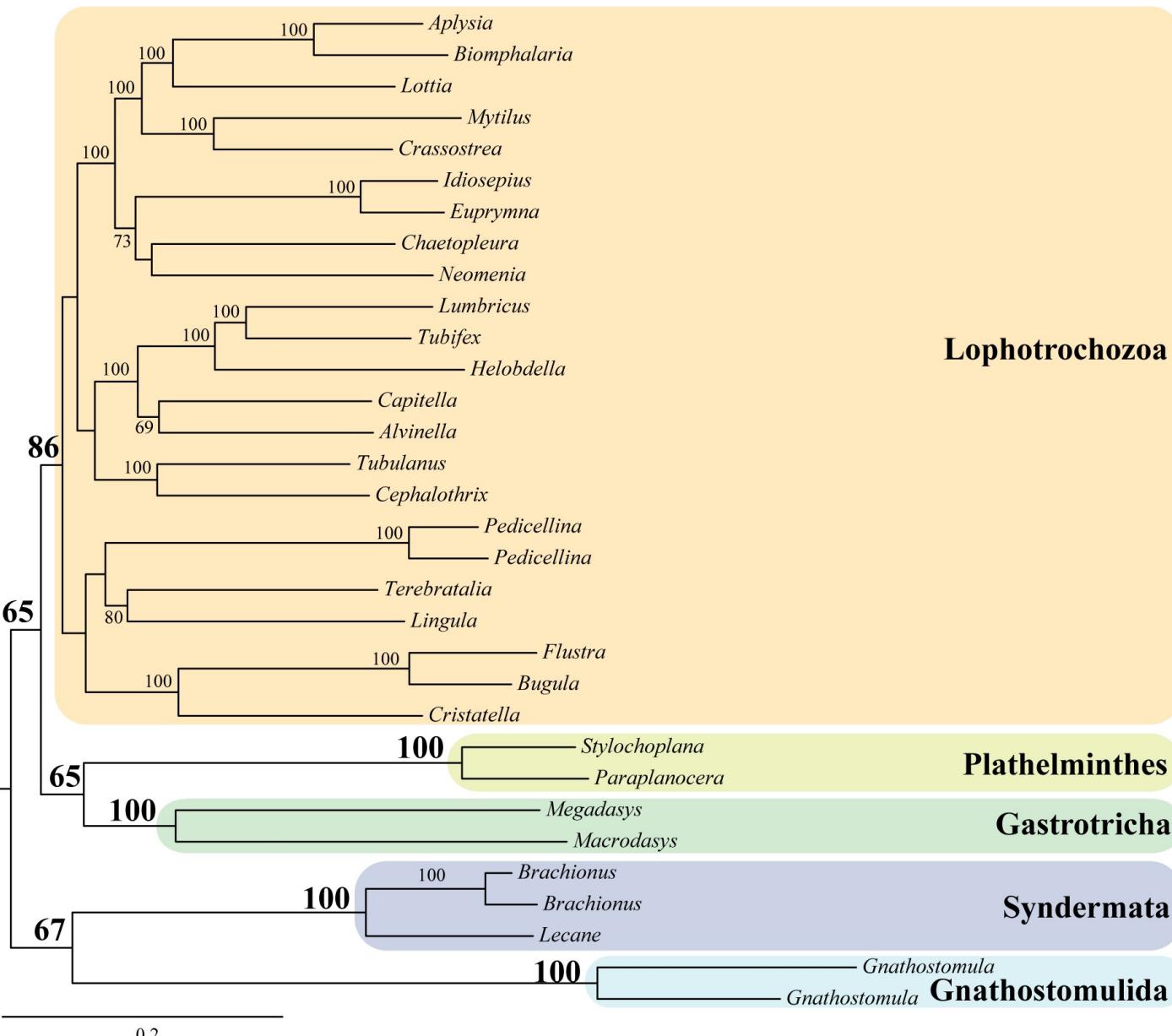
Brute-force phylogenomics



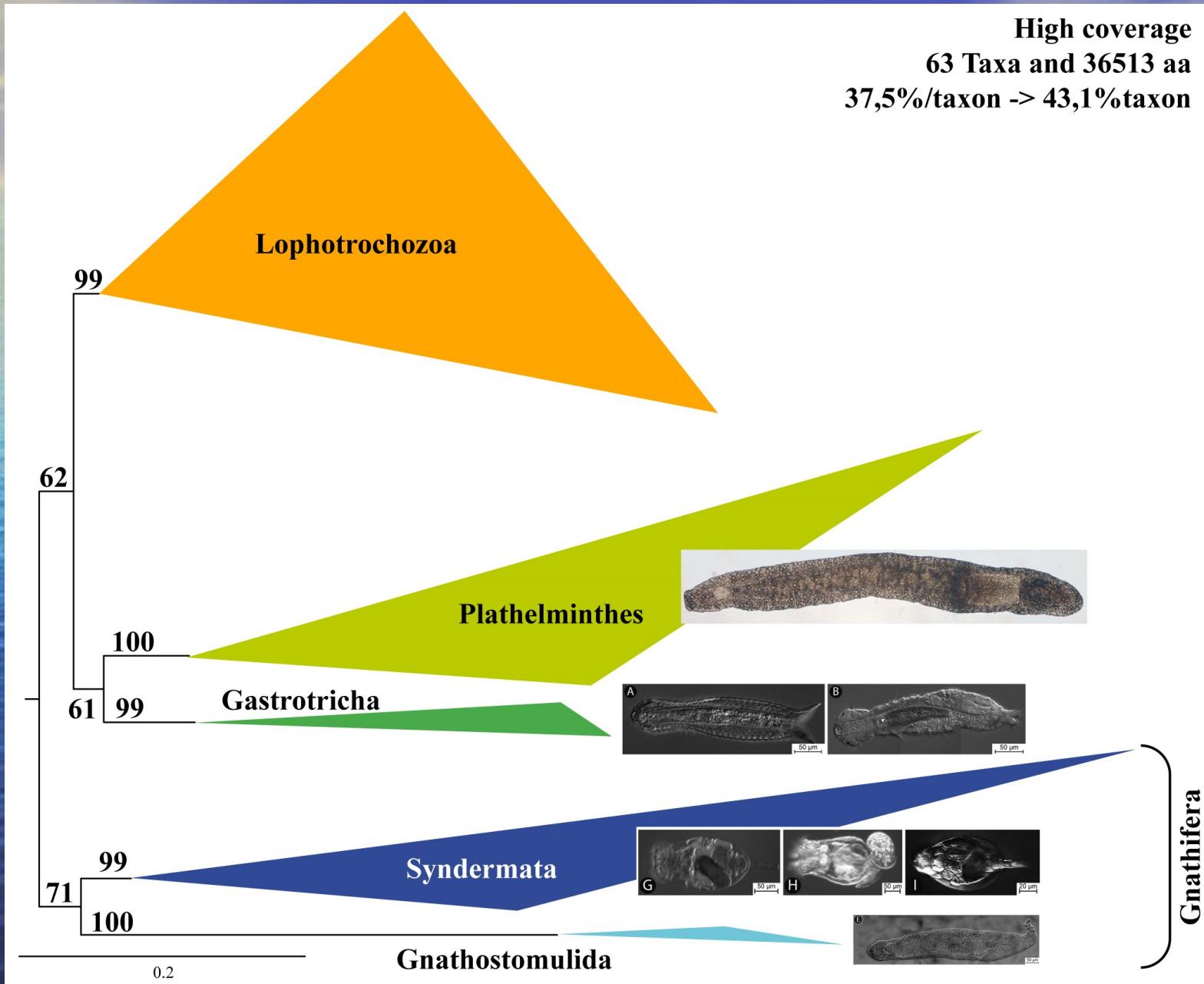
Exclusion of long-branched taxa

36 taxa instead of 63 taxa

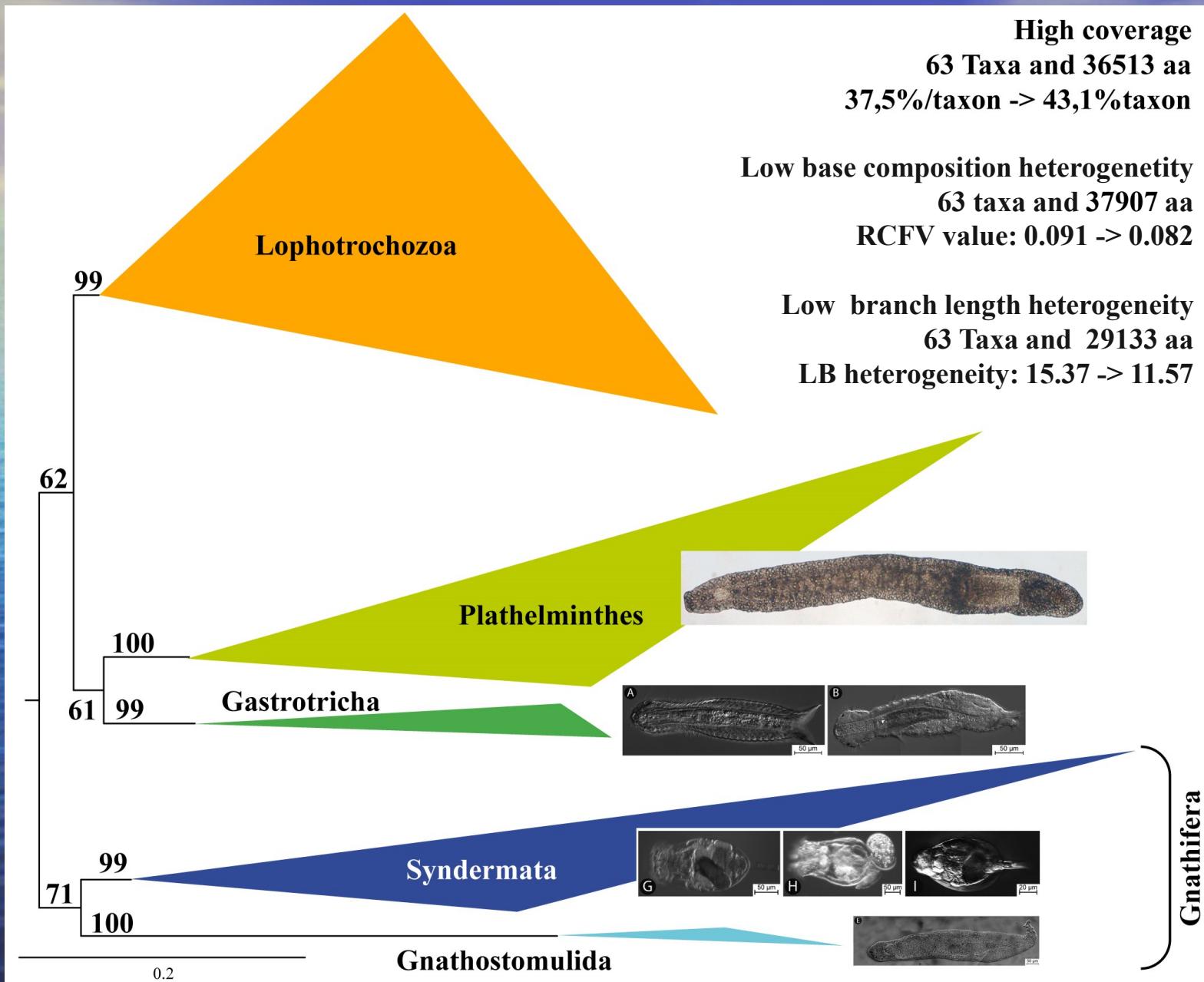
82162 aa



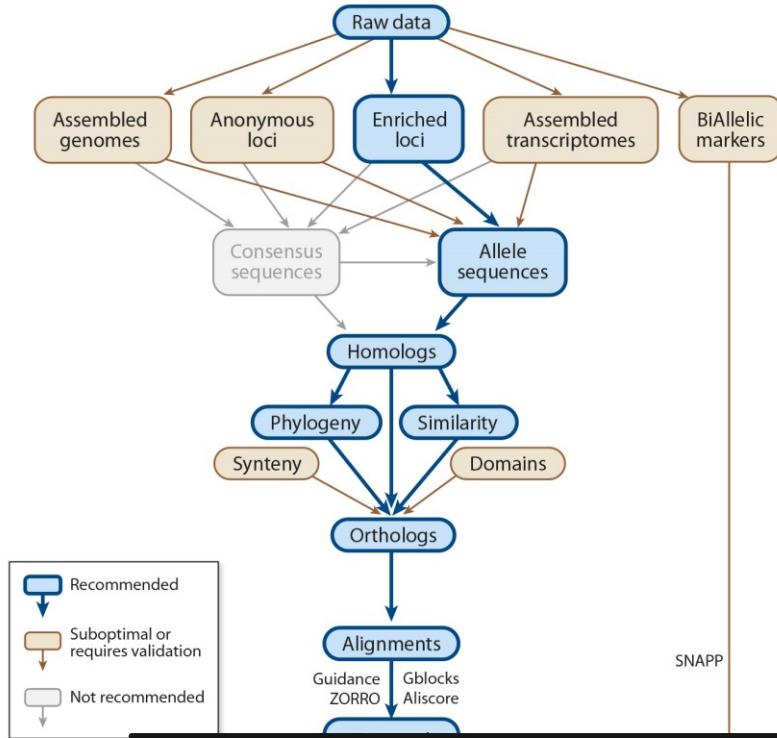
Exclusion of biased data



Exclusion of biased data

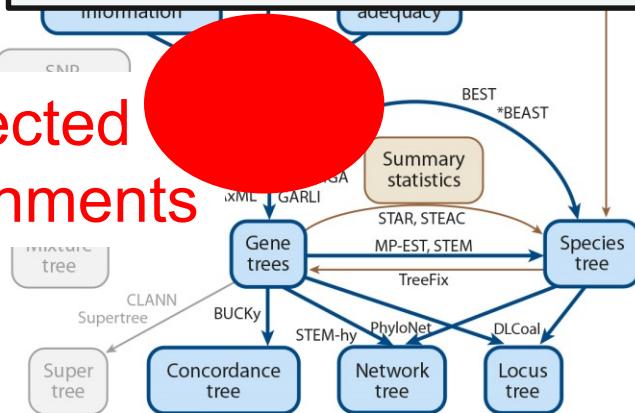


Data quality



Given the rapidly increasing quantity of available data, systematists would benefit from an integrated system for data subsampling that would allow selection of a sufficient number of loci using objective criteria for minimizing phylogenetic error (indicated by a *star* in **Figure 4**).

Selected Alignments



Your own computer & SAGA

<https://documentation.sigma2.no/jobs/submitting.html>

Login into SAGA

ssh \$USERNAME@saga.sigma2.no

Password: *your password*

Project number: NN9458K