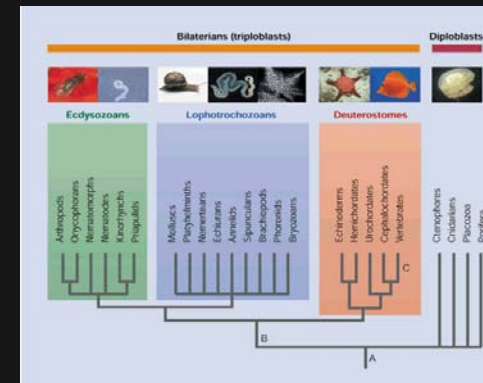




DryadUK Sept 12th 2011

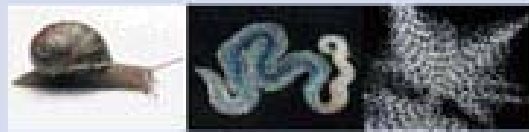


Seb Shimeld  
 Department of Zoology  
 University of Oxford



Bilaterians (triploblasts)

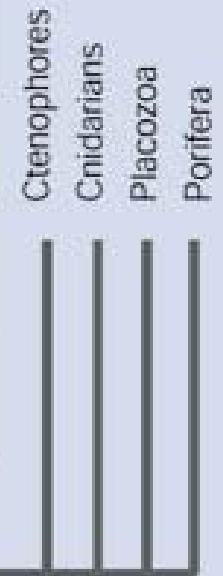
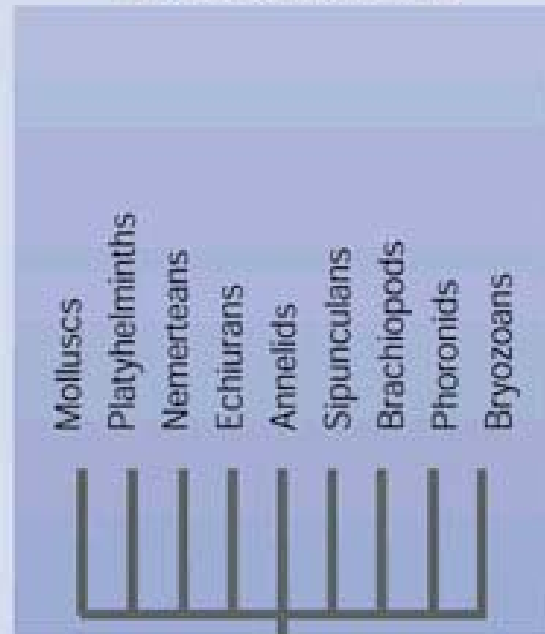
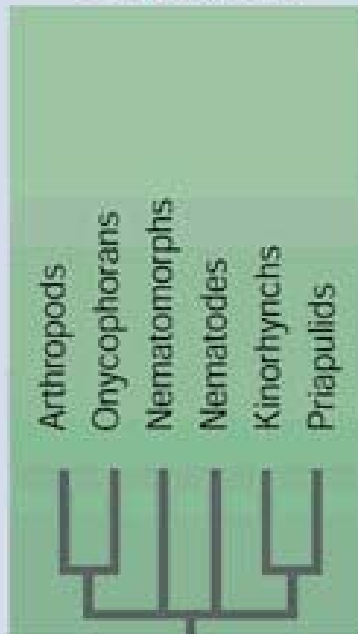
Diploblasts



Ecdysozoans

Lophotrochozoans

Deuterostomes



B

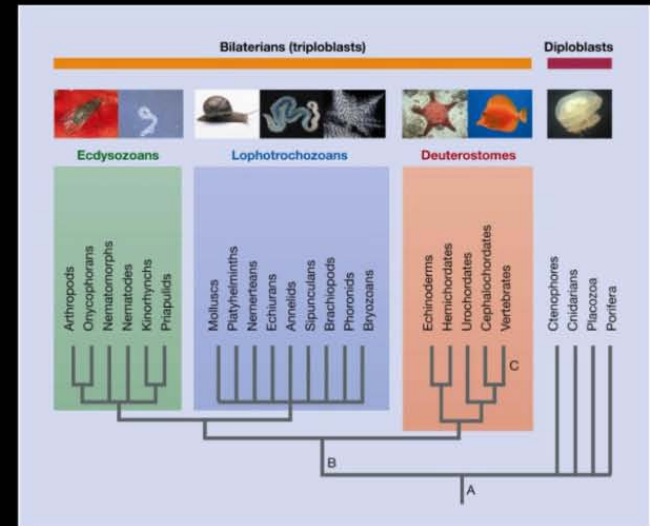
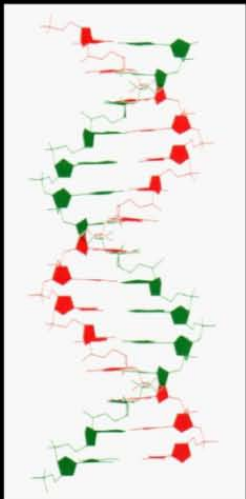
A

C

# The aim .....

A causal understanding of the path leading from genetic changes to morphological diversity

## Phenotype



## Phylogeny

# How might a typical project go today?

(data types)

- Start with a next gen sequencing run: 10-40Gbases of seq data for ~£2K (short read seq data)
- Assembly and initial analysis (sequences, spreadsheets etc.)
- Molecular phylogenetics, synteny etc. (images, tree files, output from web browsers)
- Cloning, expression, function (mostly images)
- Replicate this across a phylogeny

# What do we do with the data?

- Publication
  - Selected assembly/analysis
  - Selected images and phylogenetic trees
  - A bit more in the supplementary?
  - Sequence data in the relevant archive?
- What's left?
  - Lots of assembly and analysis
  - Lots of trees
  - 99% of image data

# Problems

- Most data never gets public; especially true of image data
- Sustainability of sequence data storage
- Integration of data

## Current fixes

- Organism-specific databases: For example Aniseed (Ascidian Network for In Situ, Expression and Embryological Data) which houses genome, image, experimental data
- Caveats: Expensive, limited, and what's in it for me?