Today

● This morning
  ● What is a phylogenetic tree?
  ● How to “read” simple phylogenetic trees
  ● How can you make a phylogeny?
  ● How can you root a phylogeny?

● Gene Duplications & orthology
So far: one gene, one individual, one species

- Biology is more complex
- Not all internal nodes are divergences between species
General idea of today

- Having all the genomes …
- What is the relation of my gene to homologous genes in the same and other organisms (what are the relations within the gene family)
Not just divergence between species; what is up here?

Kinase-5: essential for centrosome separation in mitosis
Gene duplication is the duplication of a region of DNA that contains a gene (important force in evolution of gene families / genome evolution).
Two genes per species: how to differentiate between one ancient or two recent duplications?

- Two genes in Human chromosomes (Human A & Human B) & two genes in mouse chromosomes (Mouse A & Mouse B)
Duplications, Speciations

1. Mouse A
2. Mouse B
3. Human A
4. Human B

Speciation  Gene Duplication
One more genome evolution process / gene family evolution process: what is up here? (besides duplication & speciation)

Gene loss
Interpreting the tree: duplications vs speciations, gene tree inside species tree.
Interpreting the tree: duplications vs speciations, gene tree inside species tree loss.
Blast with a newly sequenced globin from frog

What kind of globin is it?

Anja Roesner,* Christine Fuchs,† Thomas Hankeln,† and Thorsten Burmester*

*Institute of Zoology, Johannes Gutenberg University, Mainz, Germany; and †Institute of Molecular Genetics, Johannes-Gutenberg-University, Mainz, Germany

Hemoglobin, myoglobin, neuroglobin, and cytoglobin are four types of vertebrate globins with distinct tissue distributions and functions. Here, we report the identification of a fifth and novel globin gene from fish and amphibians, which has apparently been lost in the evolution of higher vertebrates (Amniota). Because its function is presently unknown, we tentatively call it globin X (GbX). Globin X sequences were obtained from three fish species, the zebrafish...
Fitch 1970
Two genes in two species are orthologous if they derive from one gene in their last common ancestor; **This means that the “node” in the tree where they diverge is a speciation node.**

"the corresponding gene"

Genes can diverge by
- Speciation, or
- Duplicaion

NOT orthologous relative to the human-mouse speciation
Retraction in MBE because of duplication / speciation, orthology / paralogy

- [http://retractionwatch.wordpress.com](http://retractionwatch.wordpress.com)

- Quite simply, our study targetted specifically a comparative genomics analysis of salmonid species. However, because the whole family went through a relatively recent whole genome duplication event, sorting out paralogs from true orthologs remains a serious challenge. We realised that some targets retained in our analyses were likely to be paralogs, which could cause some biases in our results. We now need to revisit that very carefully using alternative analytical approaches /

- They wanted to compare the sequences of genes across different species, and thought that’s what they had done. However, in these species, apparently a lot of genes are duplicated, such that there are two relatively similar copies (call them gene 1a and gene 1b). If you’re looking at a copy of a gene in a species, it’s difficult to tell whether it’s gene 1a or gene 1b. So presumably what happened is that they thought they were looking at gene 1a in both species, but realized they were looking at gene 1a in one species, but gene 1b in another species. This is an easy mistake to make, and could definitely lead to major problems. Though again, without the text of the original article, it’s tough to be more precise than that.
InParalogs (/ Co-orthologs)  
(Sonnhammer & Koonin TiG)

If the node where two genes diverged was a speciation than two genes are orthologs; if it was a duplication they are paralogs

2 genes in mouse are both orthologs to two genes in human!!
“Phylogenetic timing of duplicates”
How to time duplications and detect losses given that single gene trees are often based on too little data to reliably infer their evolutionary history?
How to time duplications and detect losses given that single gene trees are often based on too little data to reliably infer their evolutionary history?

Strict reconciliation?
How to time duplications and detect losses given that single gene trees are often based on too little data to reliably infer their evolutionary history?

Species tree guided gene tree

... Species tree guided tree building

Is there significant signal in aspects of the data (the alignment) to support a gene tree that implies less genome evolution events? e.g. TreeBEST as used in ENSEMBL COMPARA

Prevents spurious duplications and losses
Example of difference between species tree guided reconstruction vs strict tree reconciliation for duplication.
Example of difference between species tree guided reconstruction vs strict tree reconciliation for duplication.
RAL evolution?

Animal RAS  Fungal RAS  Animal RAL

Animal invention and wrong tree ("consensus" in the RAS field) OR old duplication and loss
RalGEF subcluster of RasGEF tree  Ral subcluster of Ras tree

B. dendrobatidis  R. oryzae  P. blakesleeanus
RAL evolution?

Animal RAS  Fungal RAS  Animal RAL  Early branching fungi RAL

Old duplication and loss. No more OR.
Rooting 2: via gene duplications: a not unimportant example

Phylogenetic Tree of Life

Bacteria
- Spirochetes
- Proteobacteria
- Cyanobacteria
- Planctomyces
- Bacteroides
- Cytophaga
- Thermotoga
- Aquifex

Archaea
- Green filamentous bacteria
- Methanosarcina
- Methanobacterium
- Methanococcus
- T. celer
- Thermoproteus
- Pyrodictium

Eucarya
- Entamoeba
- Slime molds
- Animals
- Fungi
- Plants
- Ciliates
- Flagellates
- Trichomonads
- Microsporidia
- Diplomonads

“three kingdoms”
How to root the tree of life?
1: Find paralogs that duplicated before the LUCA

6 found so far
How to root the tree of life? 2: Make a tree of paralogs that duplicated before the LUCA

Gribaldo 1998 J Mol Evol
The root of the tree of life

One gene duplication
Zero losses

Six losses
Three gene duplication

A poor alternative, tree reconciliation if e.g. euk vs bac & arch