

BIG4: Biosystematics, informatics and genomics of the big 4 insect groups- training tomorrow's researchers and entrepreneurs

Kick-Off Meeting
14-18 September 2015
Copenhagen, Denmark



This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 642241

Genomic data and evolutionary history

Niklas Wahlberg (Lund)

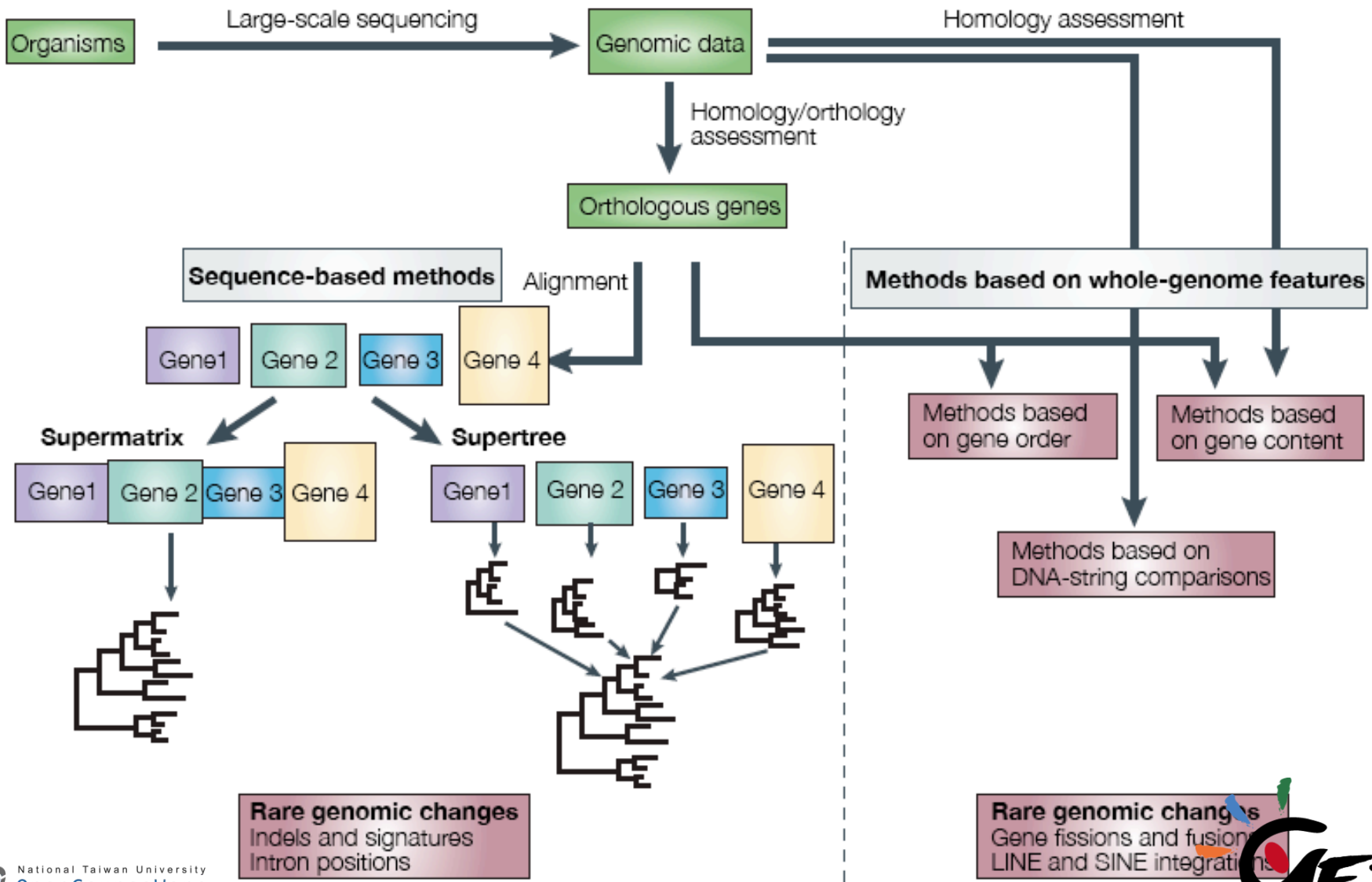
What kind of data?

- Genomes are a massive source of information
- But not all information in genomes is useful for phylogenetics
- The most important concept is homology/orthology

The rise of phylogenomics

- The field of systematics is still trying to figure out how best to utilize genomic level data
- What parts of the genome should be used?
- How can we get at those parts in the most efficient way?

Summary



Rare genomic changes
 Indels and signatures
 Intron positions

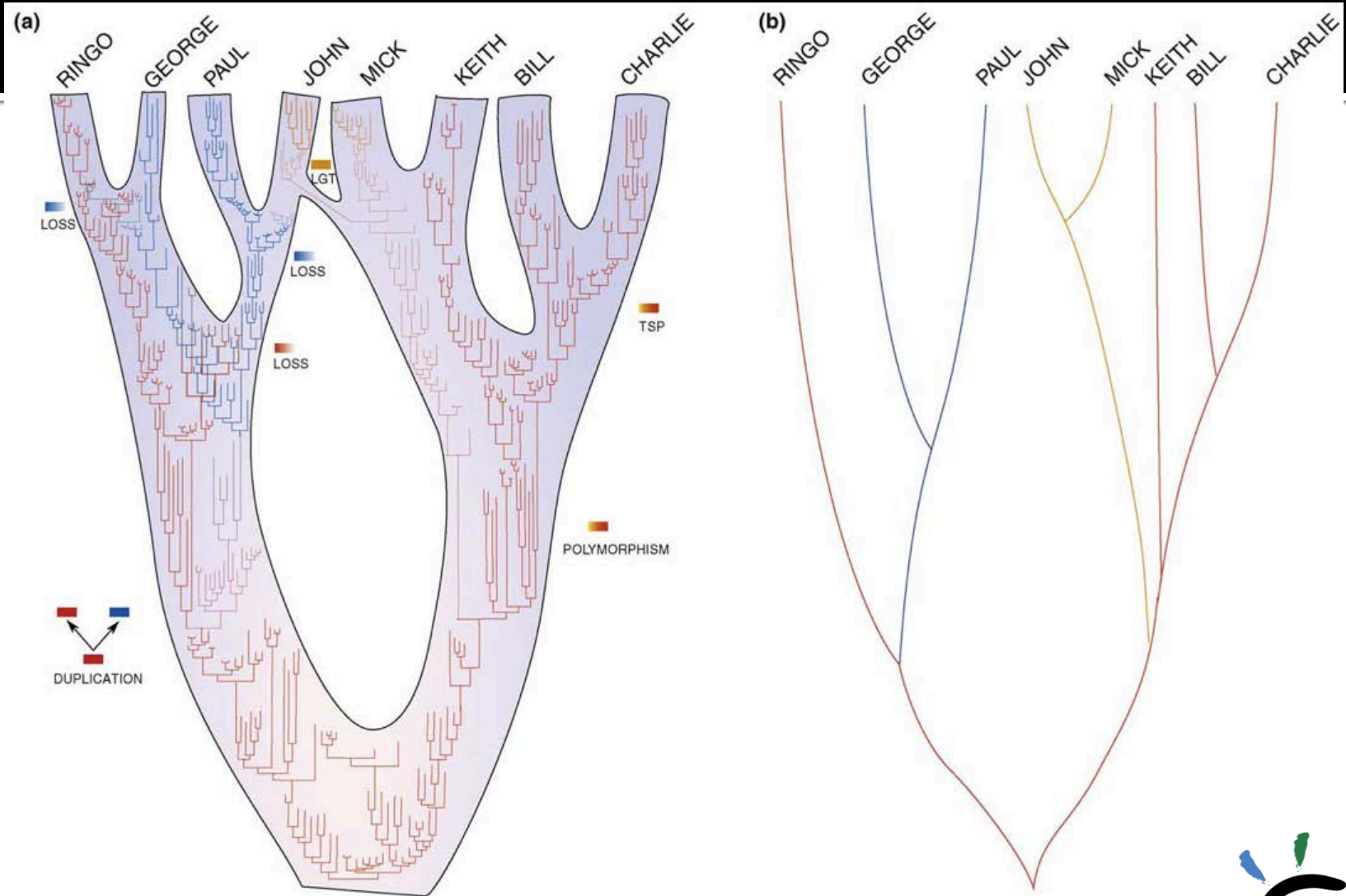
Rare genomic changes
 Gene fissions and fusion
 LINE and SINE integrations



Phylogenomics

- Whole genome analyses will allow us to understand:
 - Intron-exon boundary dynamics
 - Gene duplication-deletion dynamics
 - Gene transfer dynamics
- For deeper level phylogeny, it appears that single-copy protein-coding genes are the most useful

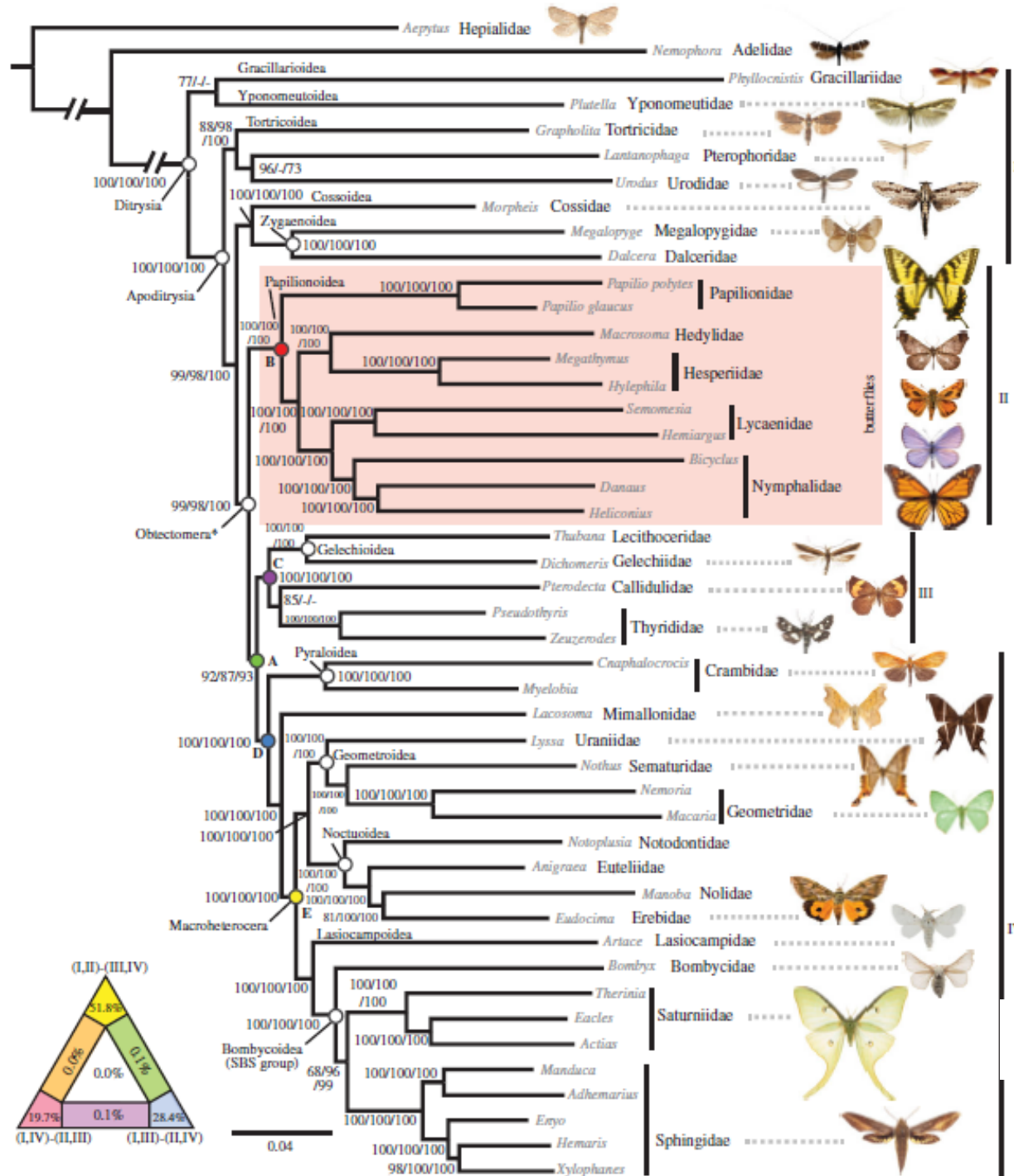
Gene dynamics over evolutionary time



Finding data

- Genome-genome blasts
- Transcriptome-transcriptome blasts
- Pulling out a known set of genes using e.g. `tblastn`

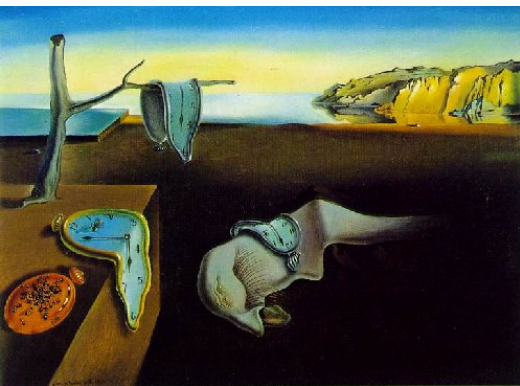
2696 genes!



Kawahara & Breinholt 2014: PRS
doi: 10.1098/rspb.2014.0970

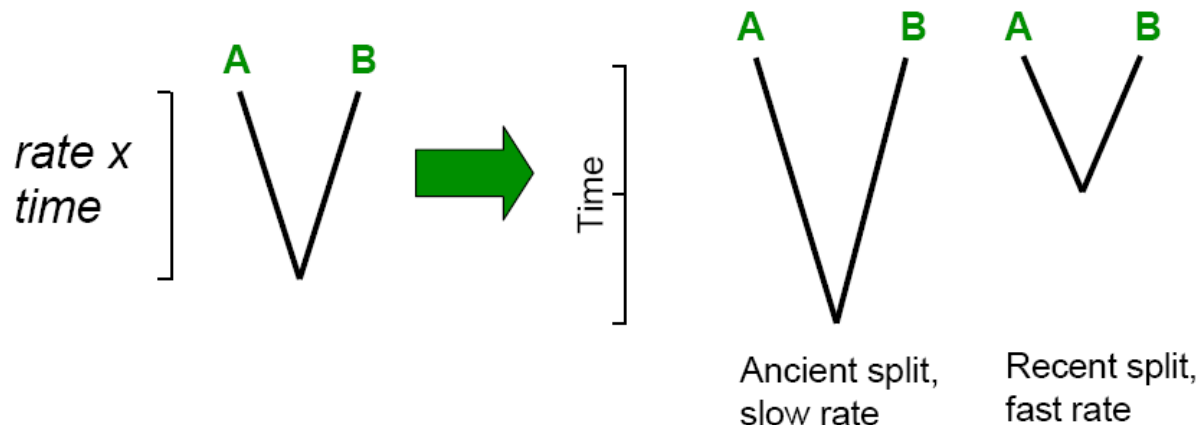
Molecular clocks for dating

- Evolutionary dynamics of nucleotides fairly well known now
- Relaxed clock: rate allowed to vary among branches
- Calibrating the clocks (fossils) gives us absolute times of divergence



Calibrating Estimates of Rates and Divergence Times

- Why do we need to calibrate?
- Phylogenetic methods usually estimate trees with branch lengths measured in substitutions per site
 - Substitutions per site = rate x time



Separating rate and time

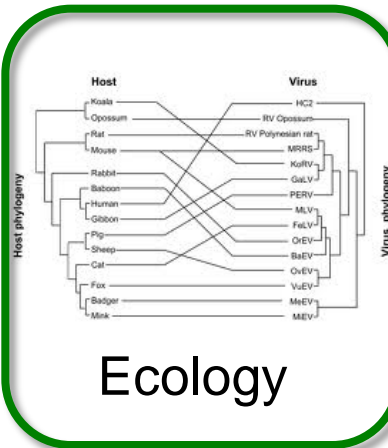
- Information about rate
 - Substitution rate obtained from an independent study
- Information about time:



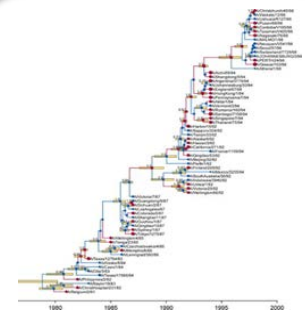
Fossil record



Biogeography



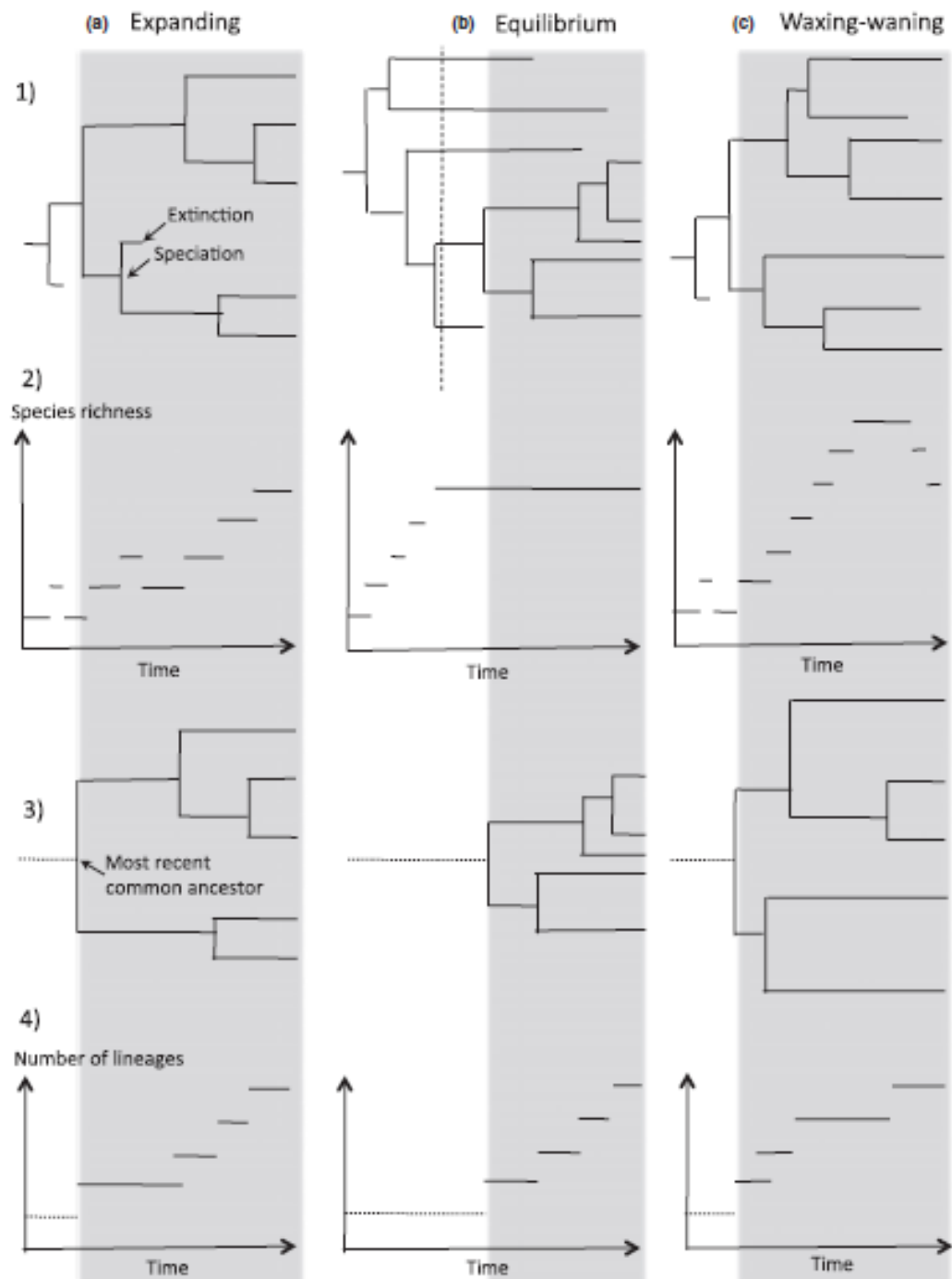
Ecology

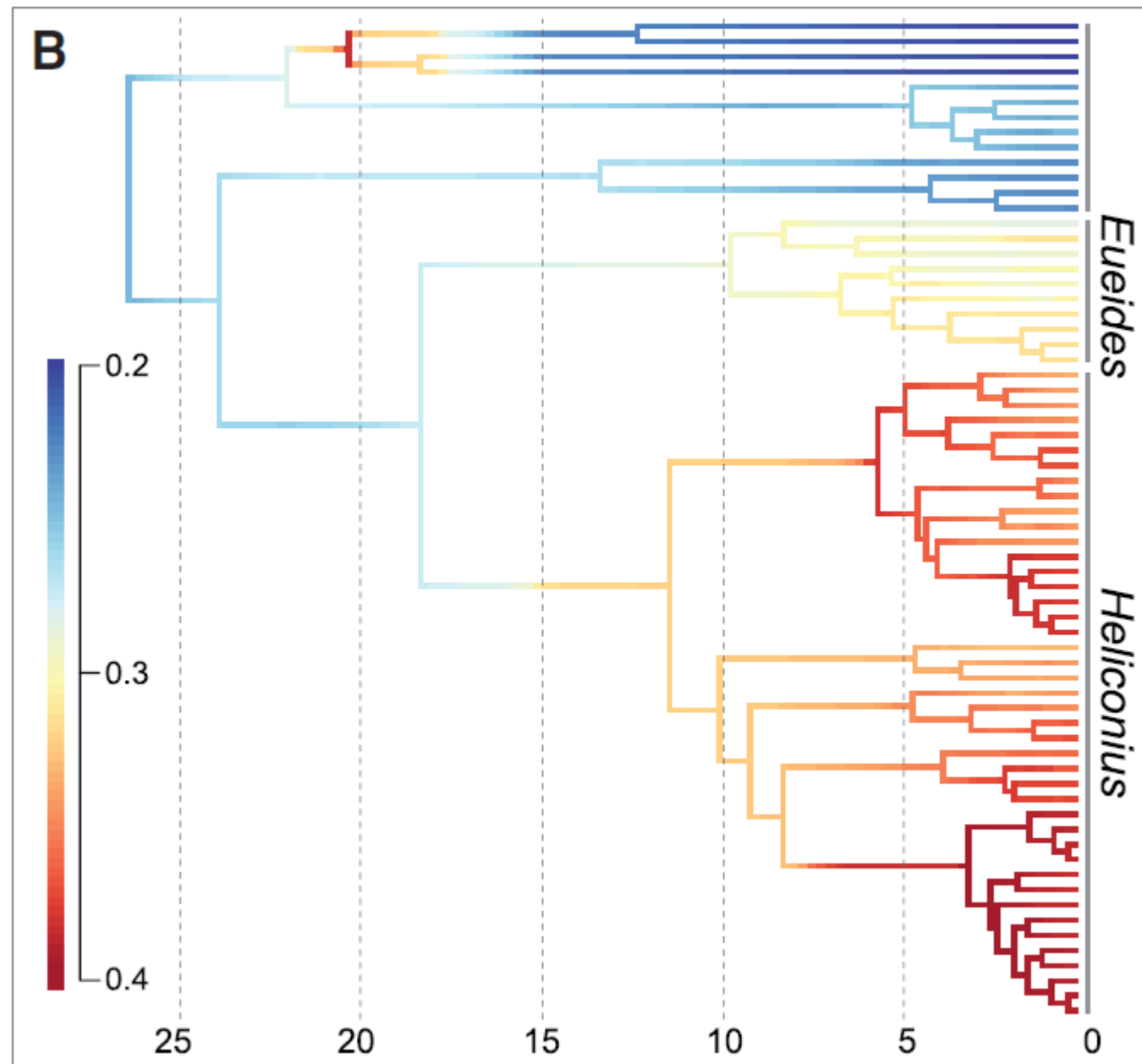


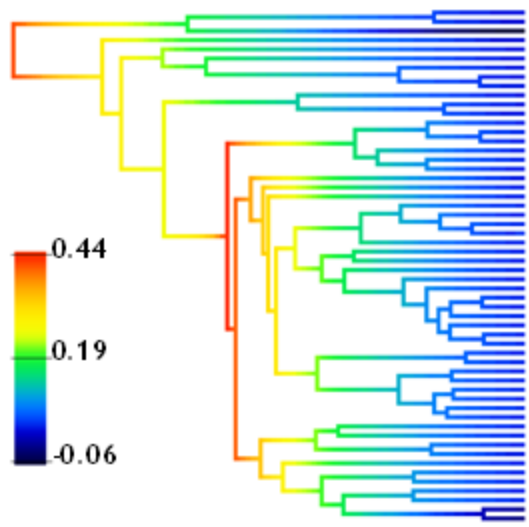
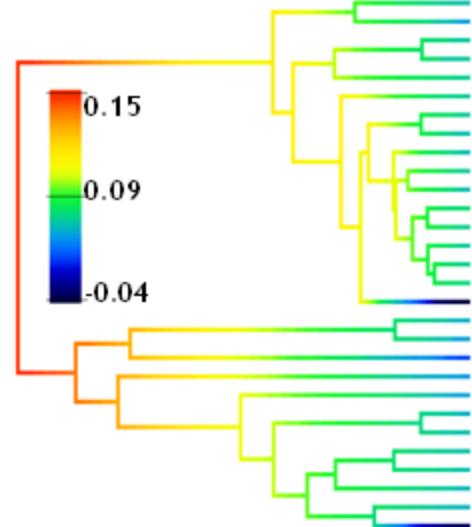
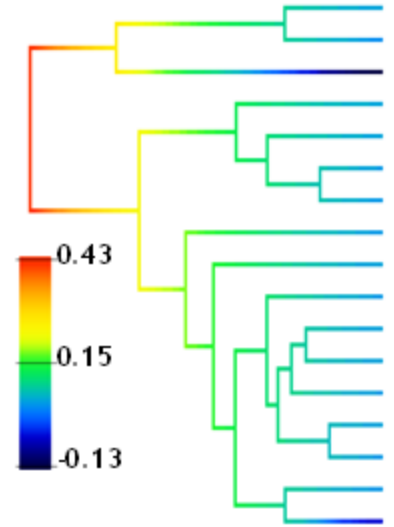
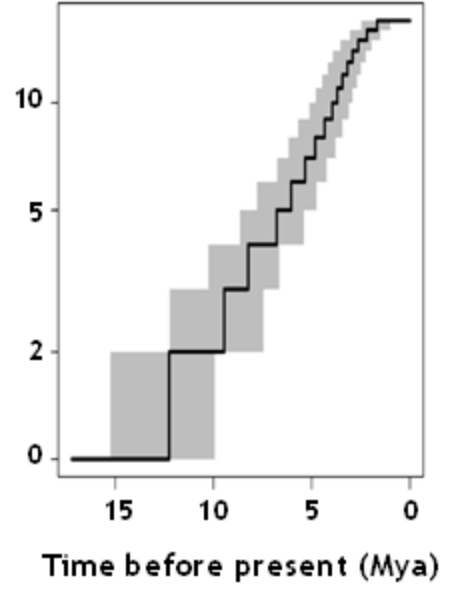
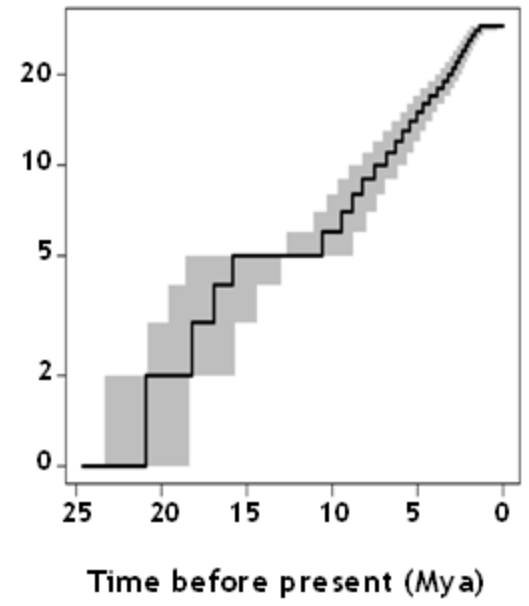
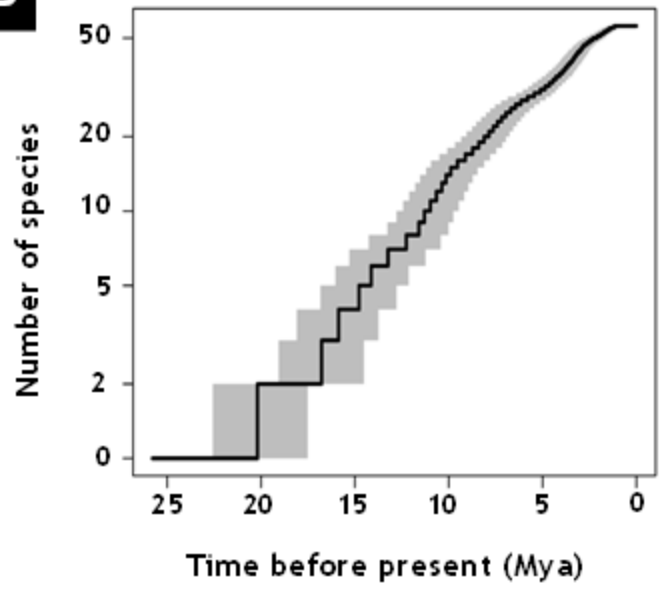
Sampling times

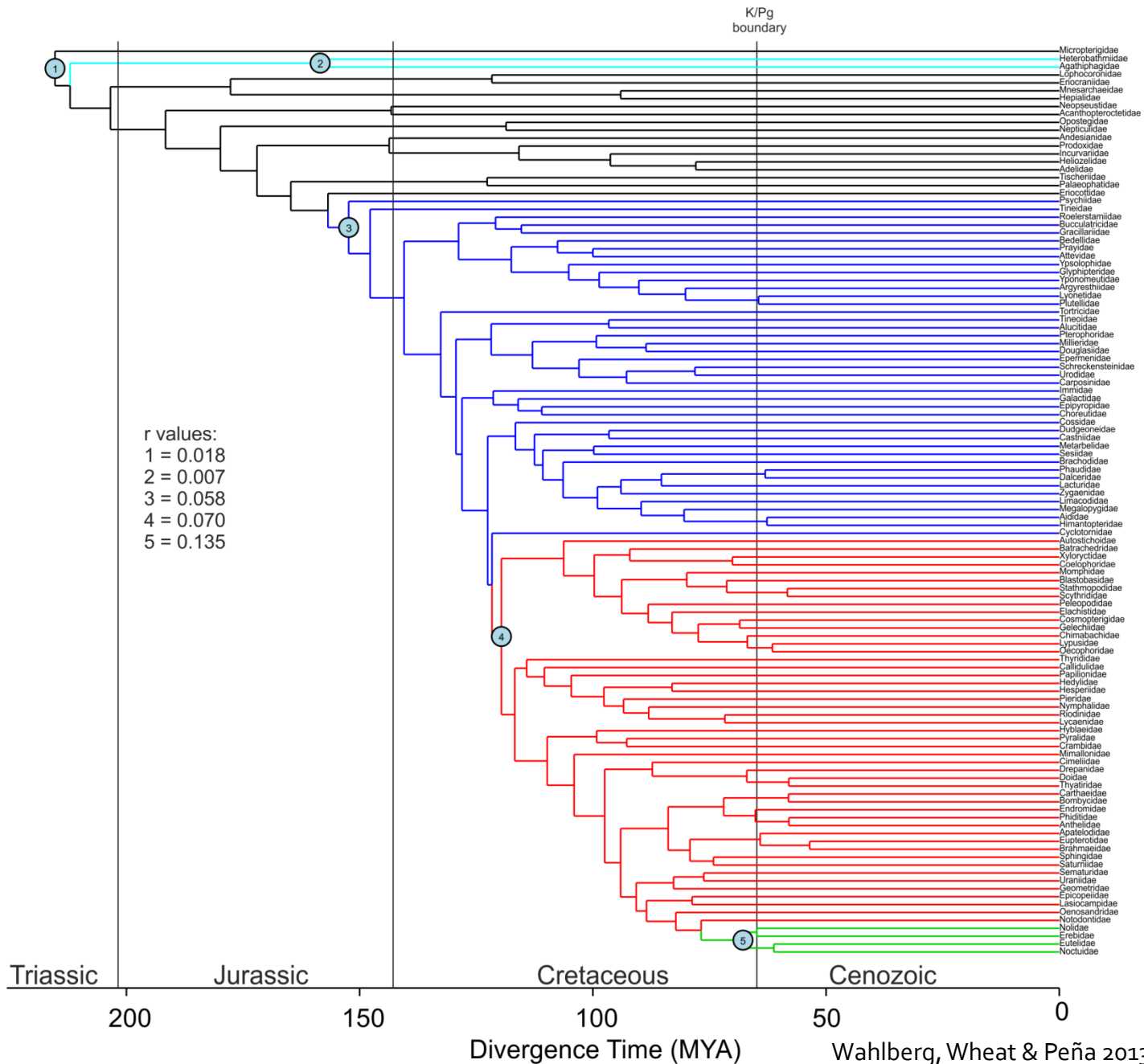
Studying diversification

- Current diversity is unevenly distributed among clades
- Why are some clades more species-rich than others?
- Just chance or some key feature that allowed a lineage to speciate more rapidly?
- Timed trees can tell us something about dynamics





A**Mydosama****Mycalesis****Lohora****B**



BIG₄ Work Packages

- Project on Lepidoptera phylogenomics utilising museum specimens
- Project on Lepidoptera diversification dynamics
- These two projects are intertwined and there are many connections to other projects in BIG₄