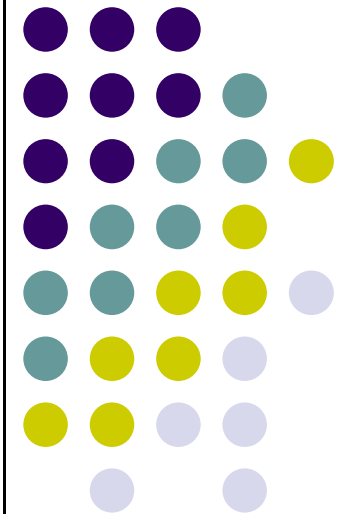


Introduction to Phylogenetics

Lars S Jermiin – lars.jermiin@usyd.edu.au
Room 320, Heydon-Laurence Bldg A08



COMMONWEALTH OF AUSTRALIA

Copyright Regulation

WARNING

This material has been reproduced and communicated to you by or on behalf of the University of Sydney pursuant to Part VB of the *Copyright Act 1968 (the Act)*.

The material in this communication may be subject to copyright under the Act. Any further reproduction or communication of this material by you may be the subject of copyright protection under the Act.

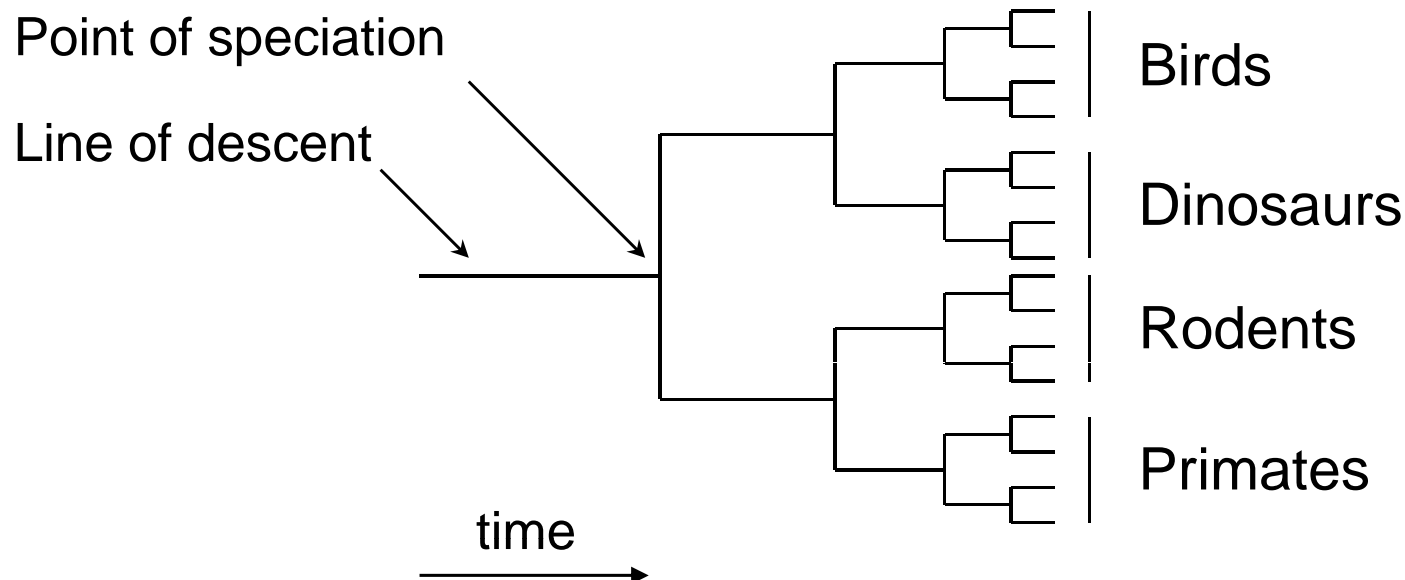
Do not remove this notice.

The Aims of This Lecture

- To outline –
 - the basis of phylogenetic research
 - the relevance of phylogenetic methods
 - the properties of a phylogenetic tree
 - the origin of the phylogenetic discipline
 - the nature of phylogenetic data

Basis for Phylogenetics

- The evolution of all species
 - is the product of an accumulation of mutations in germ-line DNA, genetic drift, natural selection and changes in the environment
 - can be represented by a **tree-like pattern of descent**



What is Phylogenetic Research?

- It is the study of evolutionary pattern and process, where
 - the **evolutionary pattern** corresponds to the tree-like pattern of descent
 - the **evolutionary process** represents the mechanisms by which mutations accumulate in germ-line DNA

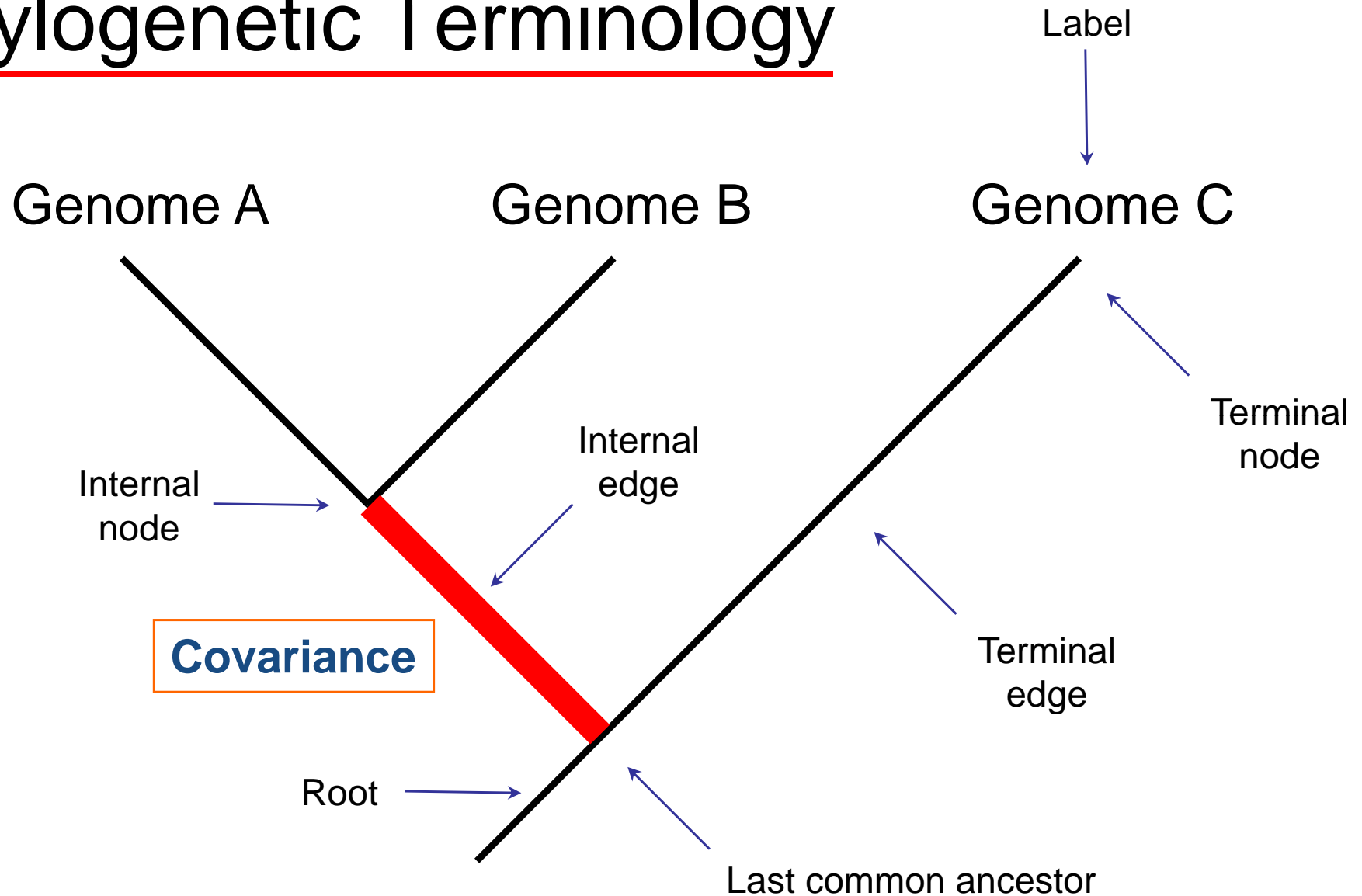
Why Phylogenetic Research?

- Because every naturally occurring piece of DNA has an evolutionary history
- Because knowledge of the evolution and function of such DNA may
 - improve management of public health
 - improve management of our resources
 - guide scientific research and development

Take-home Message #1

- In order to conduct good scientific research, it is often necessary to use the evolutionary history – i.e. the **phylogeny**

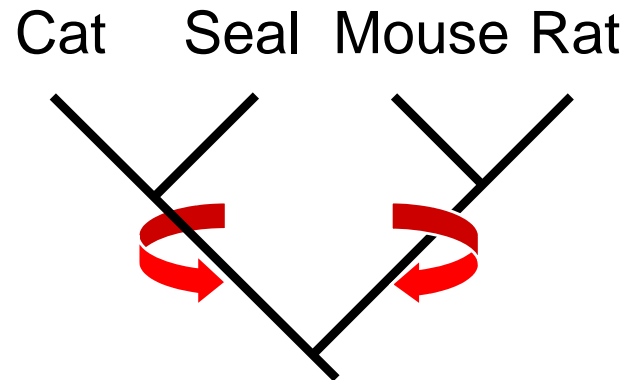
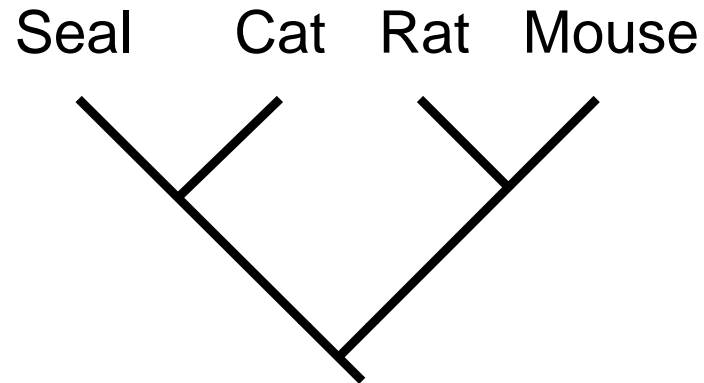
Phylogenetic Terminology



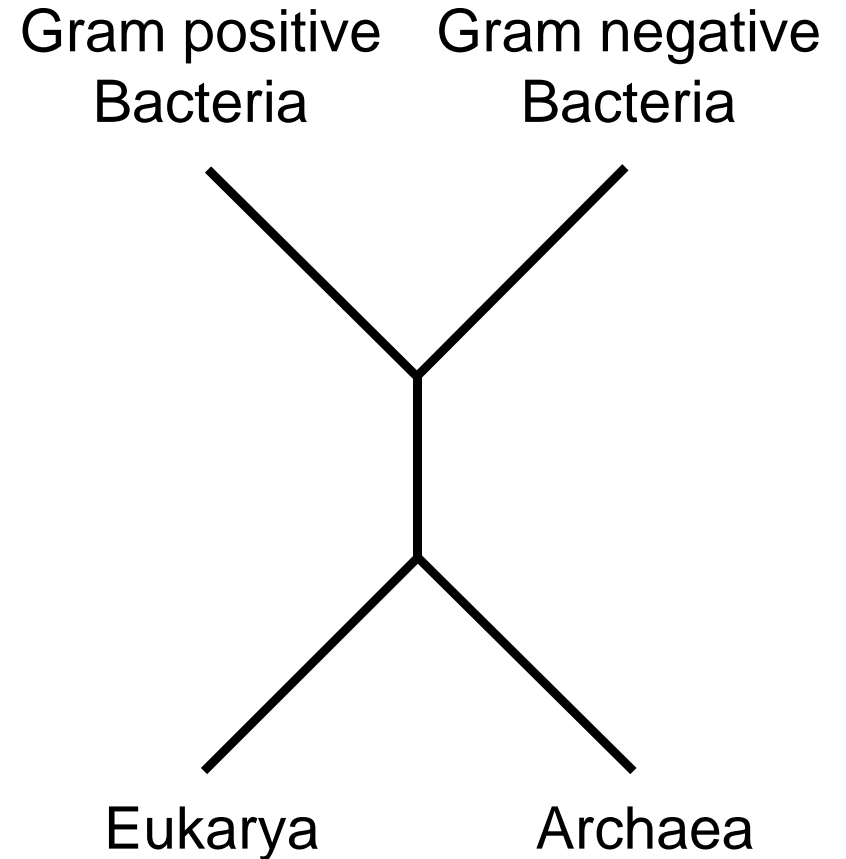
Note:

- The length of each edge may be proportional to evolutionary time!

More on Trees

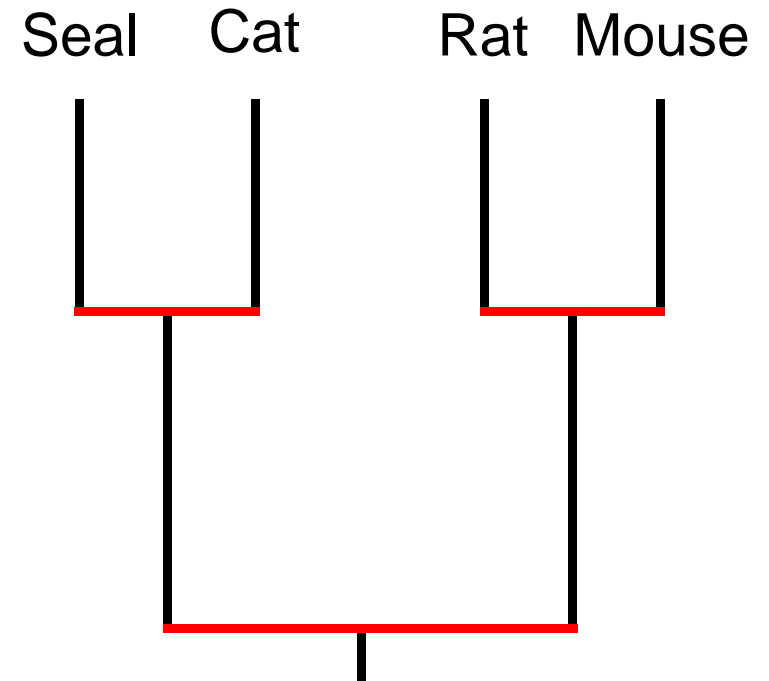
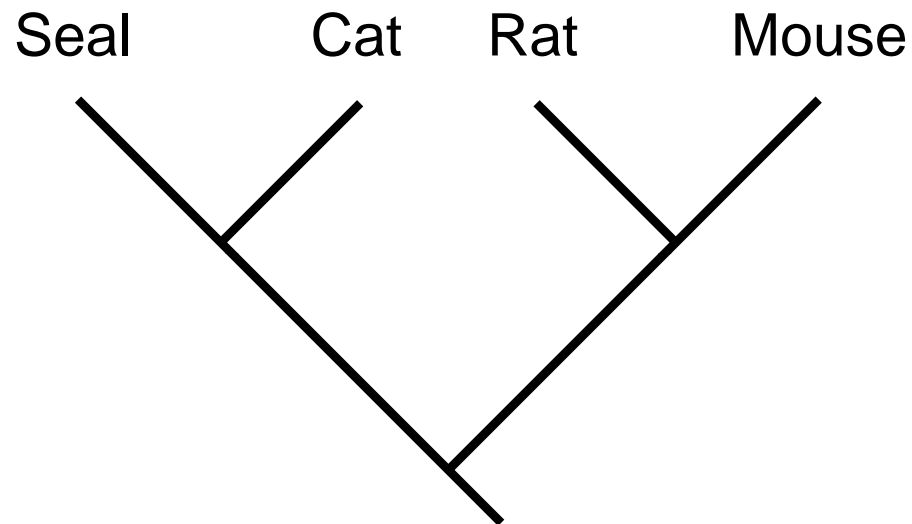


Rooted tree



Unrooted tree

More On Rooted Trees



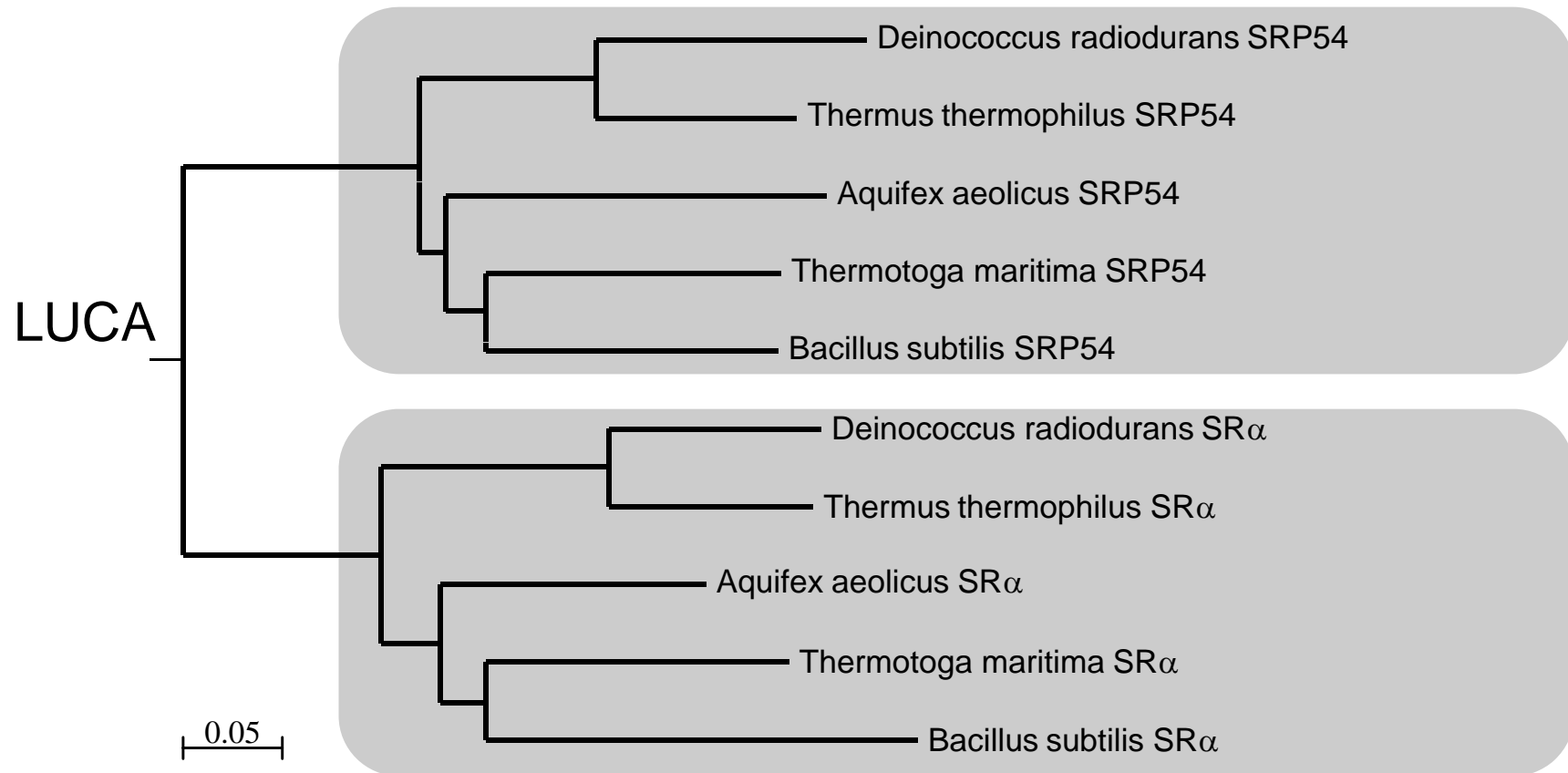
Note:

- Horizontal lines (in red) are irrelevant from a point of view of estimating time since divergence — they are only there for cosmetic reasons

Take-home Message #2

- Phylogenetic trees can be **rooted** or **unrooted**, and they can be illustrated differently, yet imply exactly the same evolutionary history

Classification of Bacteria



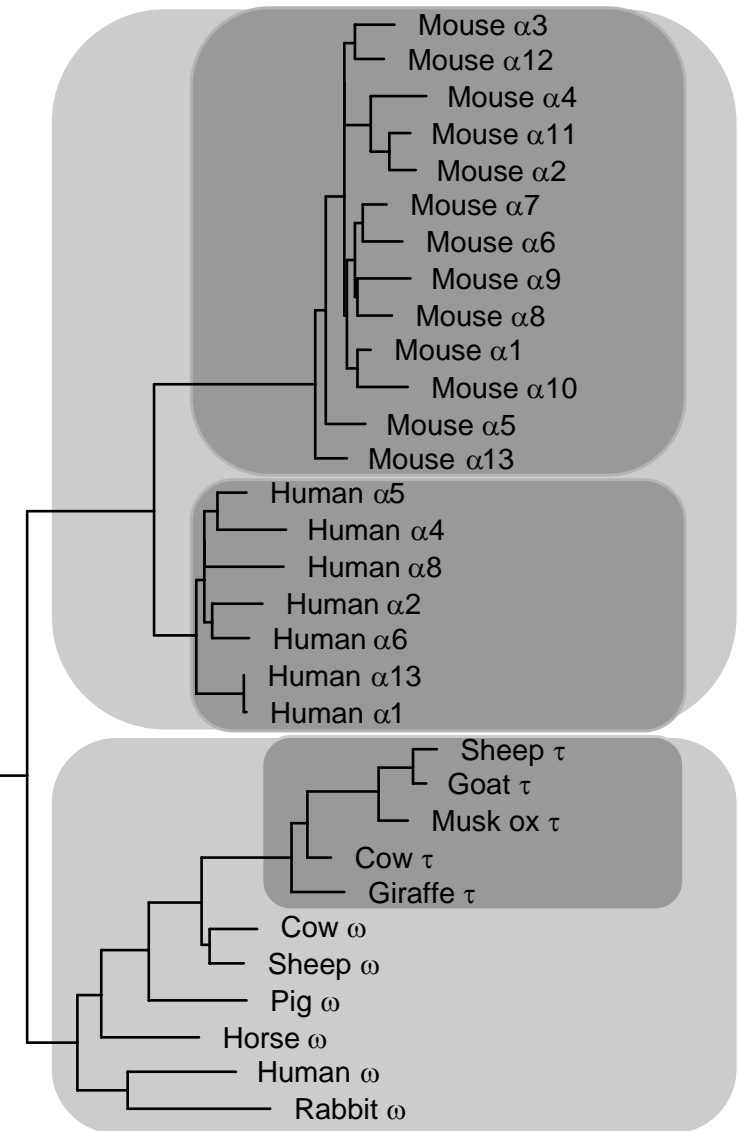
Note:

- The **length of each edge** represents the amount of **evolutionary change** along that edge and may be proportional to time (**'molecular clock'**)

Classification of Proteins

Type I Interferon — A family of cytokines with **pleiotropic activities**, including inhibition of viral replication and cell proliferation, and activation of the immune system

Lineage to the other classes of Type I INFs

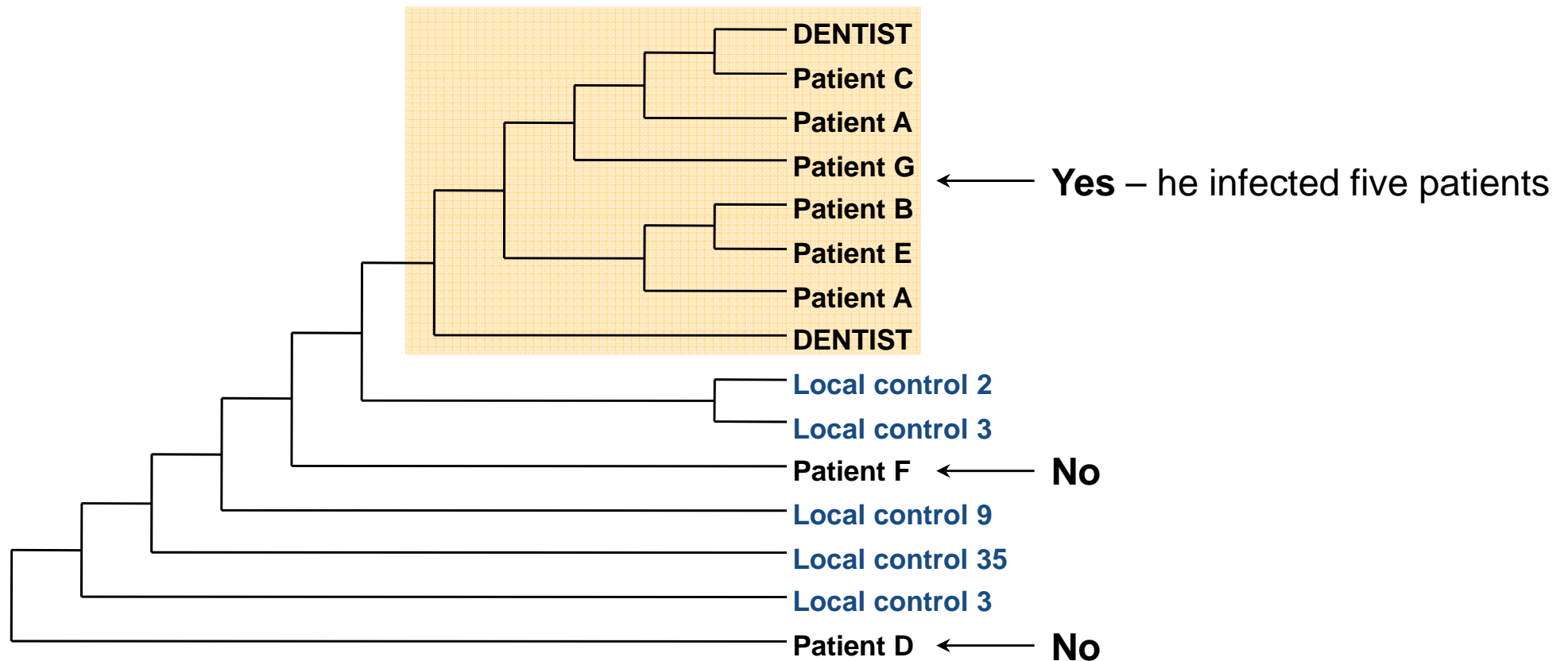


Note:

- The rate of evolutionary change may vary and be dependent on the other gene products (**natural selection**)

Forensic Analysis of HIV Infection

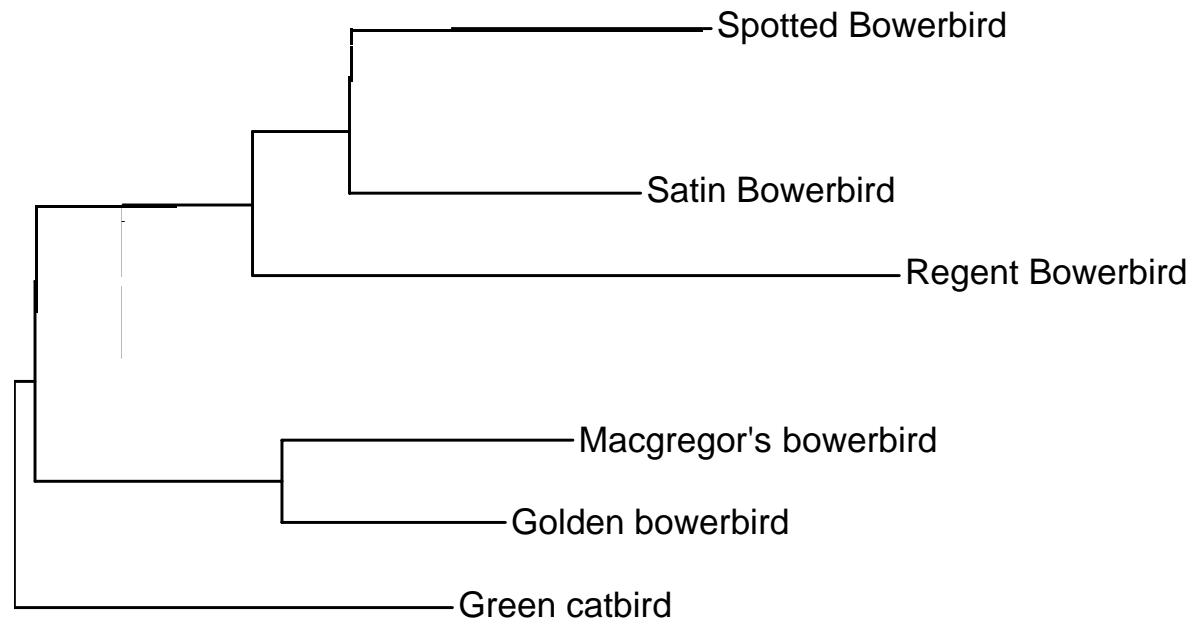
Question: Did the Florida dentist infect any of his patients with HIV?



Phylogenetic tree inferred using HIV sequences from the dentist, his patients, and five other HIV-positive people from the same geographical region

Measuring Genetic Diversity (GD)

The length of the edges in a phylogeny provides a basis for estimates of GD



Species Composition	GD	Loss
All species	0.444	
All species except Great bowerbird	0.406	~8.6%
All species except Tooth-billed bowerbird	0.336	~24.3%

Tracing Changes in Peptides

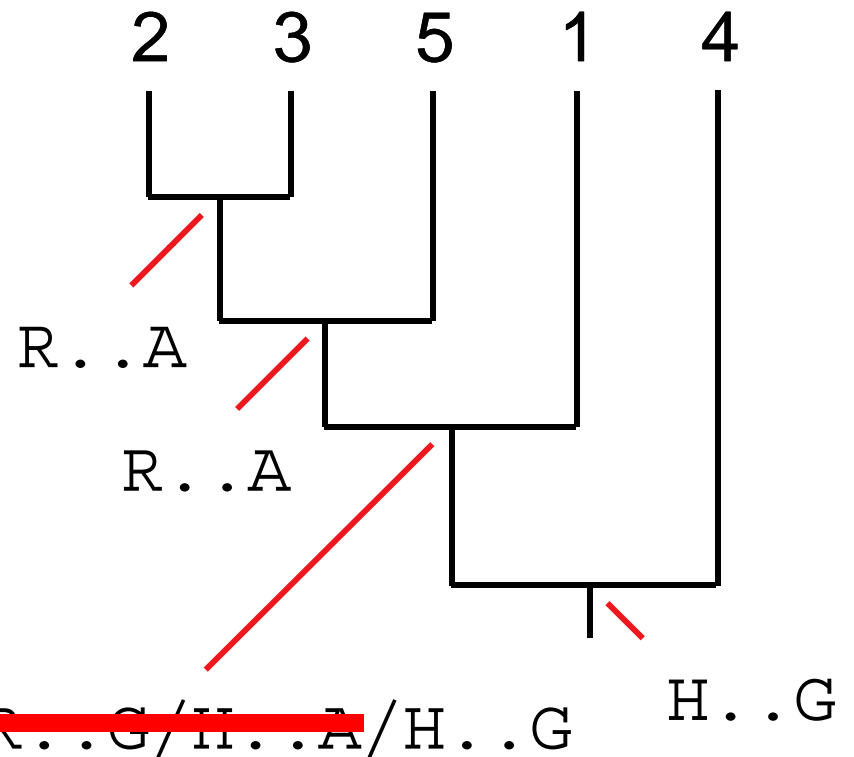
Sequence 1 .H..G.
 Sequence 2 .R..A.
 Sequence 3 .R..A.
 Sequence 4 .H..G.
 Sequence 5 .R..A.



~~Consensus .R..A.~~

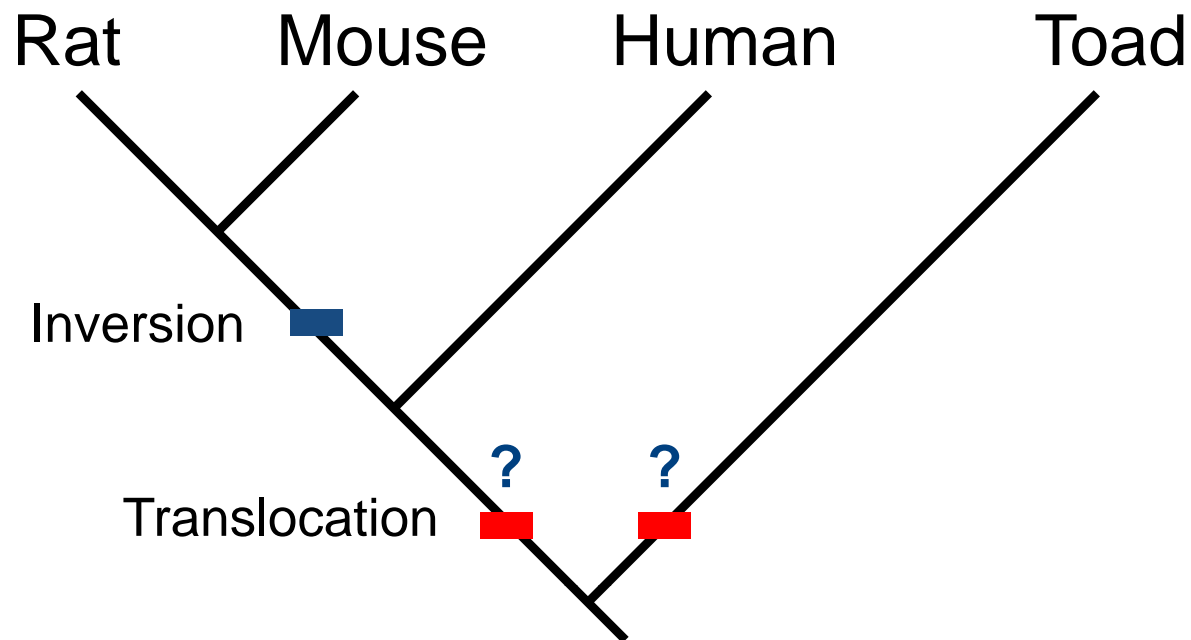
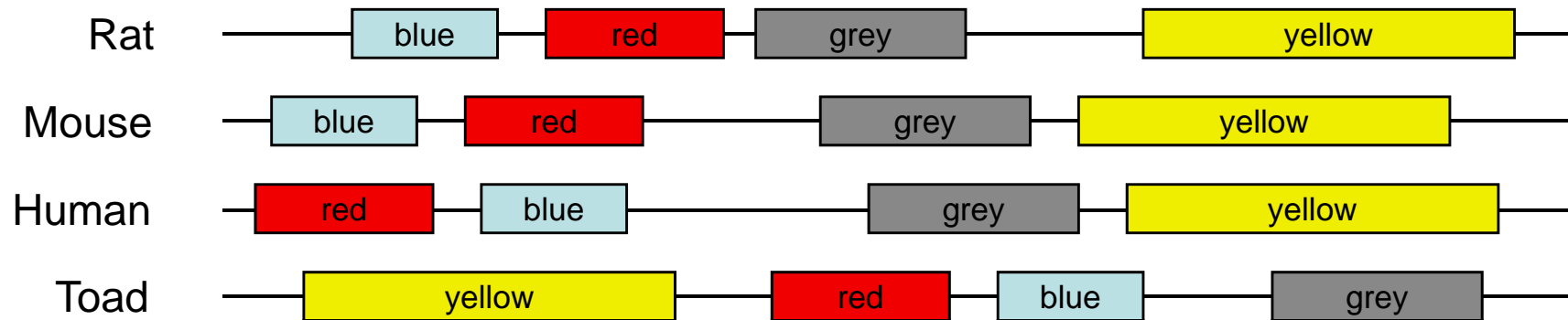


Ancestral .H..G.



Tree-dependent Inference of Ancestral Sequences

Tracing Changes in Gene Order



Take-home Message #3

- When comparing traits from more than two related taxa, we must consider the **covariance** caused by the evolutionary history

A Fundamental Problem...

The phylogeny is rarely known, implying that it must be inferred by using phylogenetic methods and data

Phylogenetics in a Broad Context

Life Sciences

Evolutionary theory
Physical theory
etc.

Statistical Science

Information theory
Bayesian theory
Set theory
etc.

Phylogenetic
inference

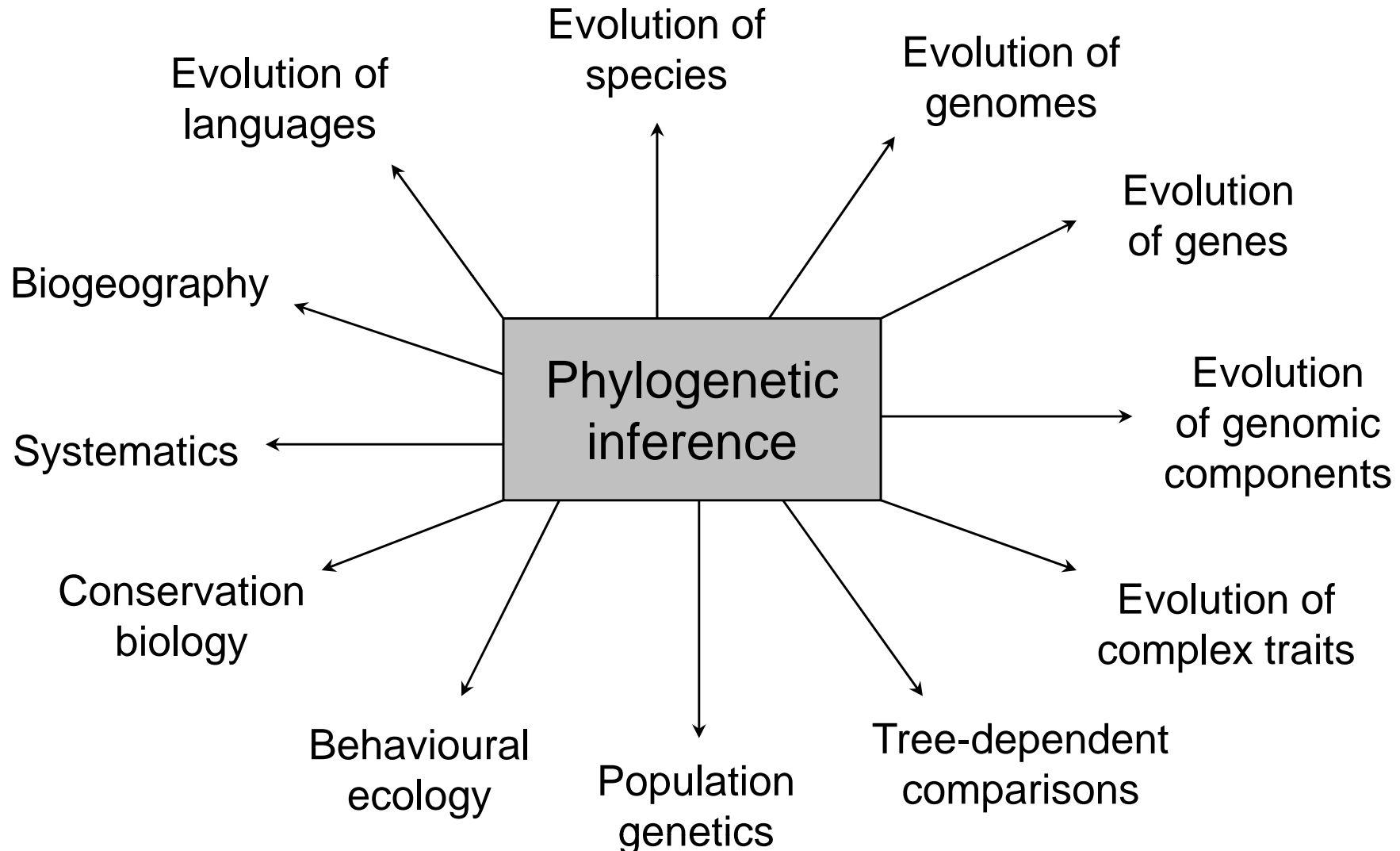
Computer Science

Graph Theory
Algorithms
etc.

Mathematical Science

Differential equations
Linear algebra
etc.

Phylogenetics in a Narrow Context



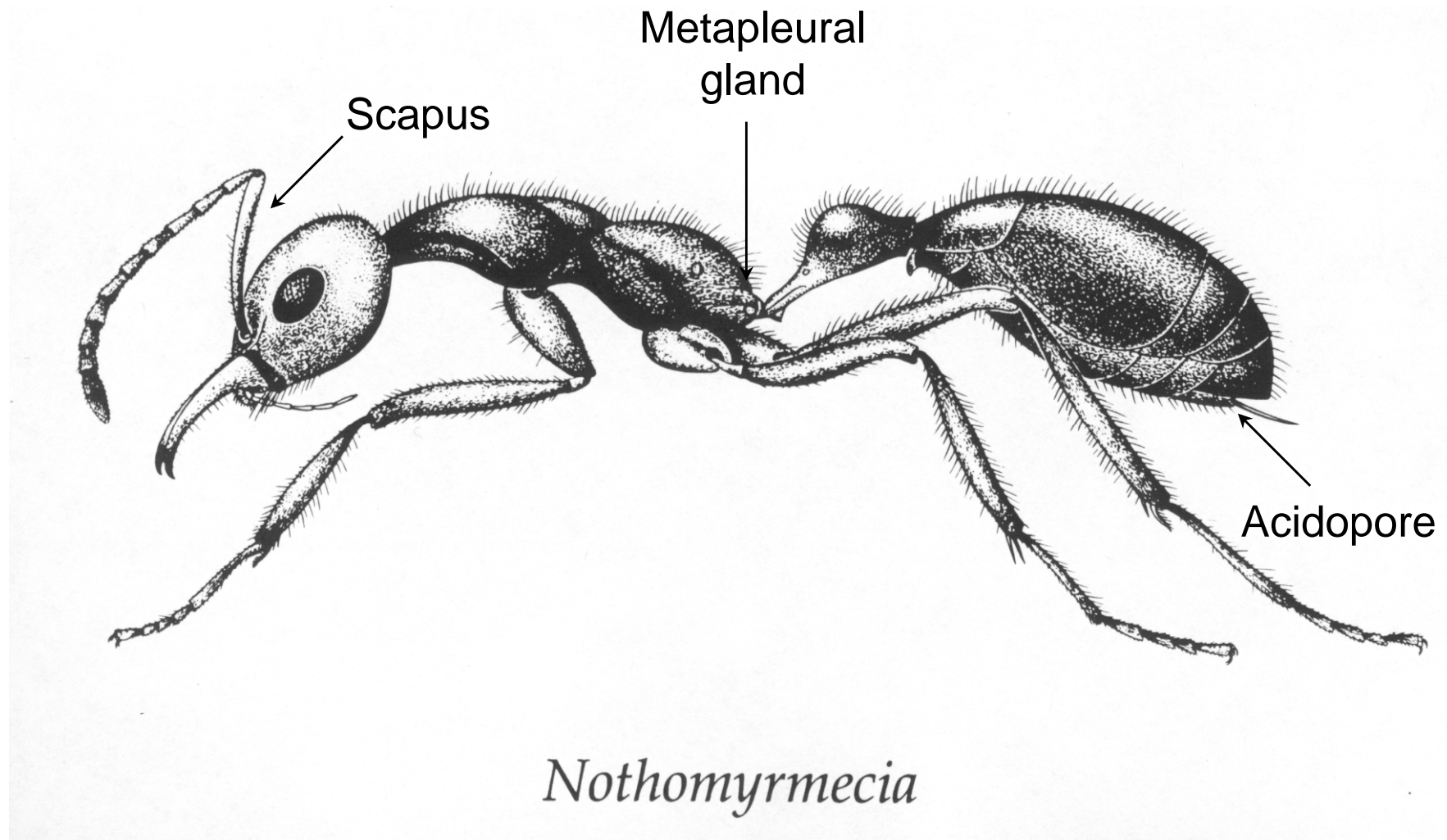
Take-home Message #4

- Phylogenetics draws heavily on the mathematical, statistical and computational sciences, and offers intelligent solutions to many areas of Life Science

Phylogenetic Data

- Obtained from the behaviour or language
- Obtained from the morphology
 - External features
 - Internal features
 - Cellular features
- Obtained from the genome
 - Continuous genetic features
 - e.g. DNA-DNA hybridisation data
 - Discrete genetic features
 - e.g. nucleotide or amino acid sequence data
 - e.g. gene order data

Example – Morphological Data



Example – Morphological Traits of Ants

Trait Number	Description of Trait
1	Acidopore present (1) or absent (0)
2	Metapleural gland present (1) or absent (0)
3	Antennae with scapus elongated (1) or short (0)

Data Matrix:	Trait Number															
	1	2	3	4	5	6	7	8	9	1	1	2	2	
Subfamily Name										0	1	6	7	
Nothomyrmeciinae	0	1	1	1	0	0	0	0	0	0	0	0	0	
..	
Formicinae	1	1	1	0	0	0	0	0	0	0	0	1	0	

Morphological Traits – Problems

- Expert knowledge is needed to characterize traits as ‘good’ or ‘bad’ phylogenetic characters
- The genetic basis for traits may not be known, so the relative importance of different traits remains unknown
- It is difficult to obtain large phylogenetic data sets
- It is difficult to obtain homologous traits from very diverged species (*e.g.*, rice & yellow-finned tuna)

Example – Molecular Data

- Banding patterns on chromosomes
- Order of genes on chromosomes
- DNA-DNA hybridisation data
- Repeated DNA, such as satellites, minisatellites and microsatellites
- Transposable elements (e.g. SINEs and LINEs)
- Nucleotide and amino acid sequences, including coding and non-coding DNA

Example – Mutations in DNA

```
GCTGTTGCTTGCTGAACTGATACGTG (original)
  |           | | |
GCTATTGCTTGTTAAACCGATACGTG (4 transitions)
  || | | | | |
GCGTTTTCTTGTA AAAAAGGATCCCTG (8 transversions)
                |||||
GCGTTTTCTTGTCCTTTTTTATCCCTG (9 bp inversion)
|||||||
-----TCCTTTTTTATCCCTG (11 bp deletion)
  |||||
--CCAAGT--TCCTTTTTTATCCCTG (7 bp insertion)
```

Note:

- The ‘-’ is used to maintain the alignment for illustrative purposes

Mutations in Protein-coding DNA

Synonymous			Non-synonymous									
Met	Ala	Thr	Ser	Gly	Pro	Trp	Tyr	Val	Asp	Asn	Arg	...
ATG	GCG	ACC	ACT	GGA	CCA	TGG	TAT	GTA	GAC	AAT	AGG	...
ATG	GCG	ACA	ACT	GGA	CCA	TGG	TAT	CTA	GAC	AAT	AGG	...
Met	Ala	Thr	Ser	Gly	Pro	Trp	Tyr	Leu	Asp	Asn	Arg	...
			Non-sense									
Met	Ala	Thr	Ser	Gly	Pro	Trp	Tyr	Val	Asp	Asn	Arg	...
ATG	GCG	ACA	ACT	GGA	CCA	TGG	TAT	CTA	GAC	AAT	AGG	...
ATG	GCG	ACA	ACT	GGA	CCA	TGG	TAG	CTA	GAC	AAT	AGG	...
Met	Ala	Thr	Ser	Gly	Pro	Trp	*					

Note:

- Insertions and deletions may lead to major changes in the C-terminal end of the encoded polypeptide chain!

Molecular Data – Problems

- It is difficult to select data that match the question of interest
- Molecular data often violate the assumptions on which the phylogenetic methods are based
- A thorough phylogenetic study is an intellectually and computationally demanding task

Take-home Message #5

- Be careful if using morphological, behavioural, linguistic or molecular data to infer phylogeny!

Suggested Literature

- M Pagel (1999). *Nature* 401, 877-884 – Good introduction to the field
- RDM Page, EC Holmes, *Molecular Evolution*, 1998.
 - Chapter 1 — General introduction (light reading)
 - Chapter 2 (Sections 2.1, 2.2, 2.4) — important reading
 - Chapter 3 — covers a lot of the background information that is assumed knowledge as well as different forms of data that may be used for phylogenetic inference (light reading)