

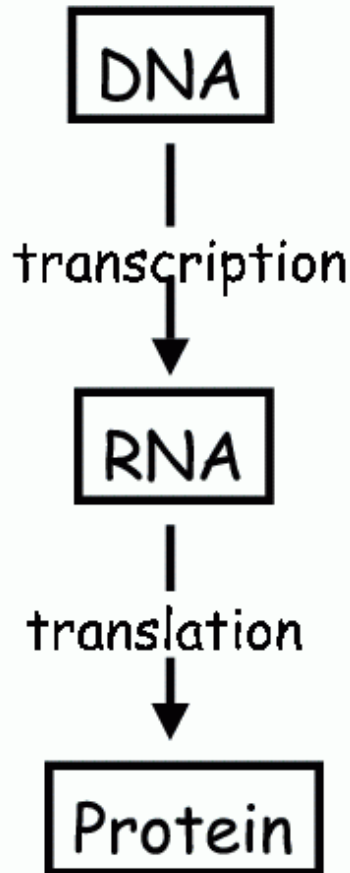
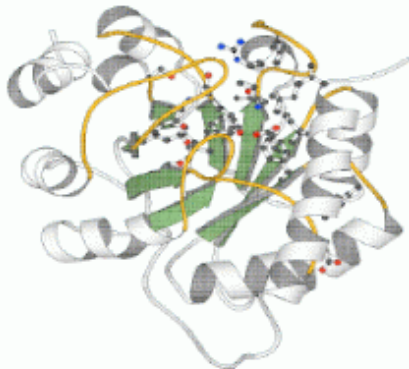
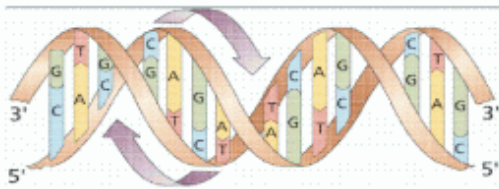
Genetic analyses for taxonomy

1. Some background details
2. Current and future sequencing
3. Collecting and storing genetic resources
- 4 Why do phylogenetic trees sometimes disagree with other datasets?

1. Some background details

DNA is transcribed into mRNA, which is translated into amino acids

Central Dogma: DNA → RNA → Protein



CCTGAGCCAACTATTGATGAA



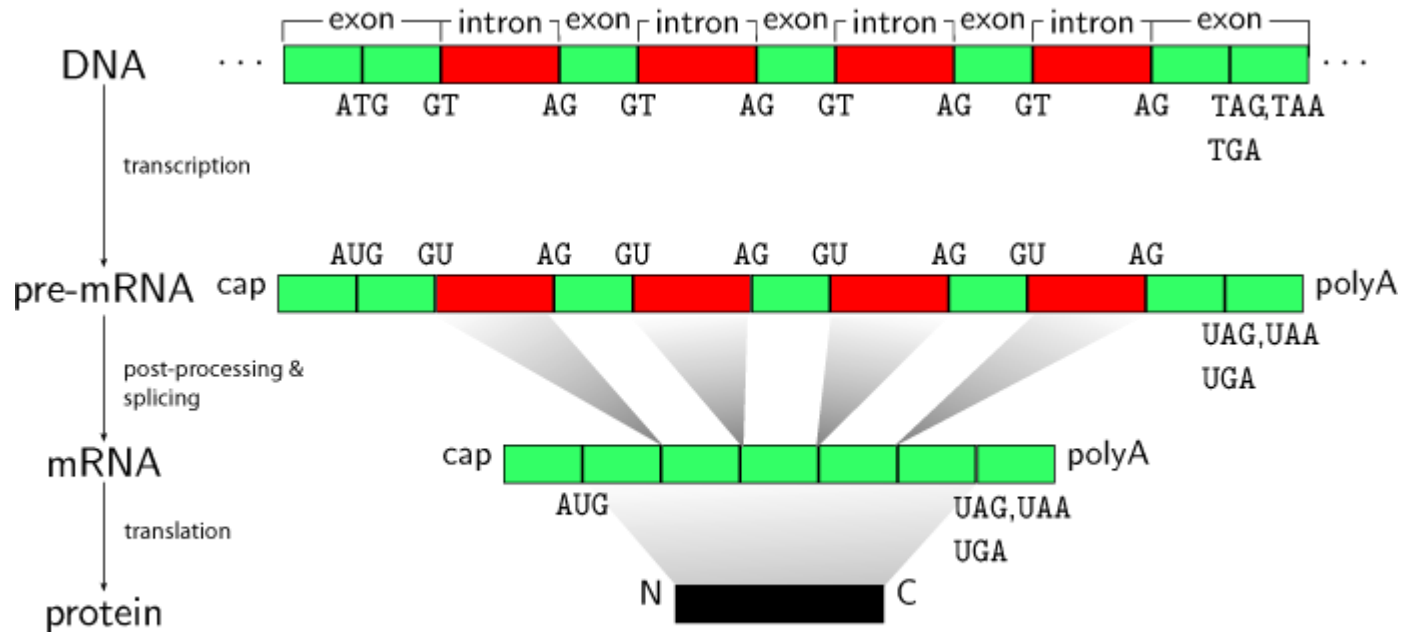
CCUGAGCCAAACUAUUGAUGAA



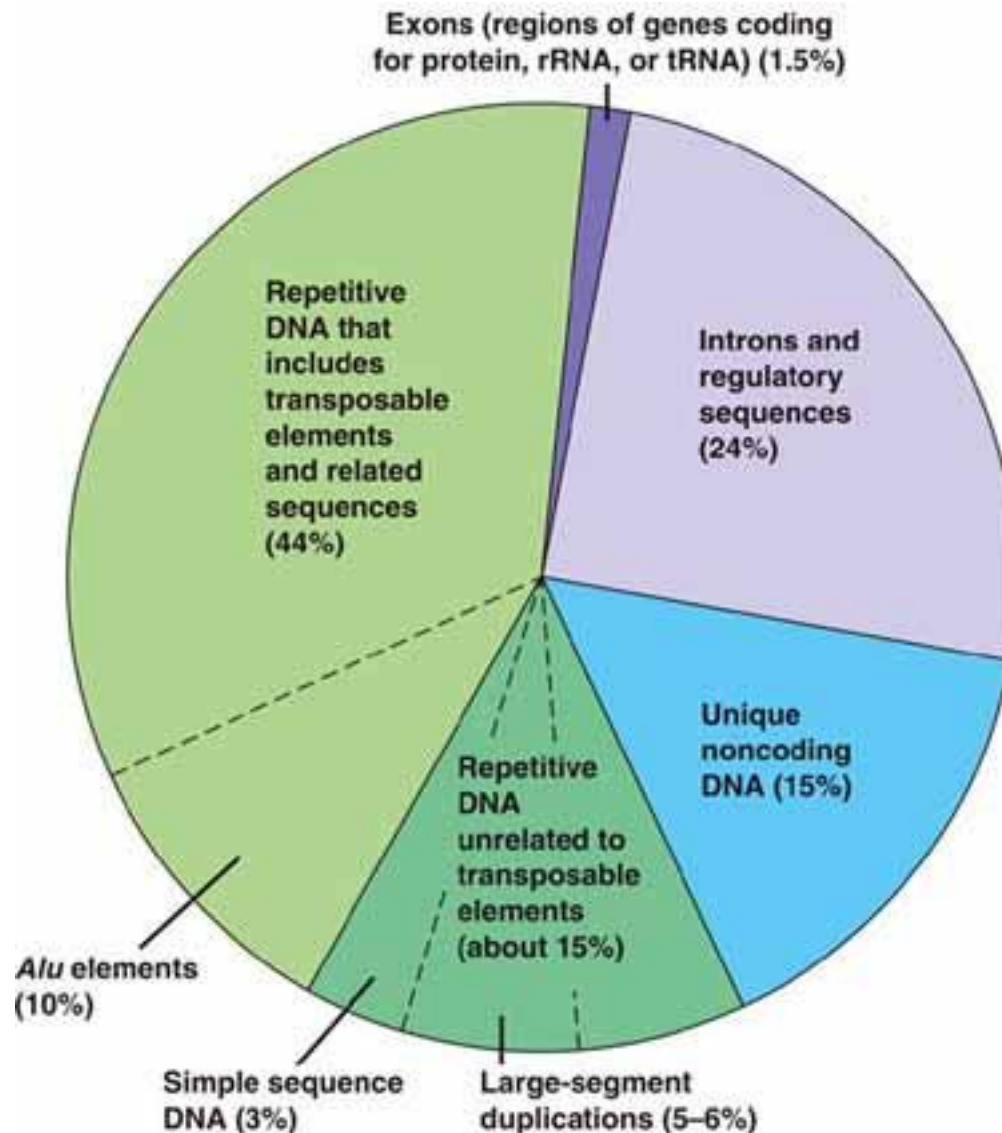
PEPTIDE

Genes are made up of introns and exons. Introns can be very long and are removed by splicing in gene expression.

This leads to concatenated exons for each gene (equivalent to the coding sequence or “CDS”)



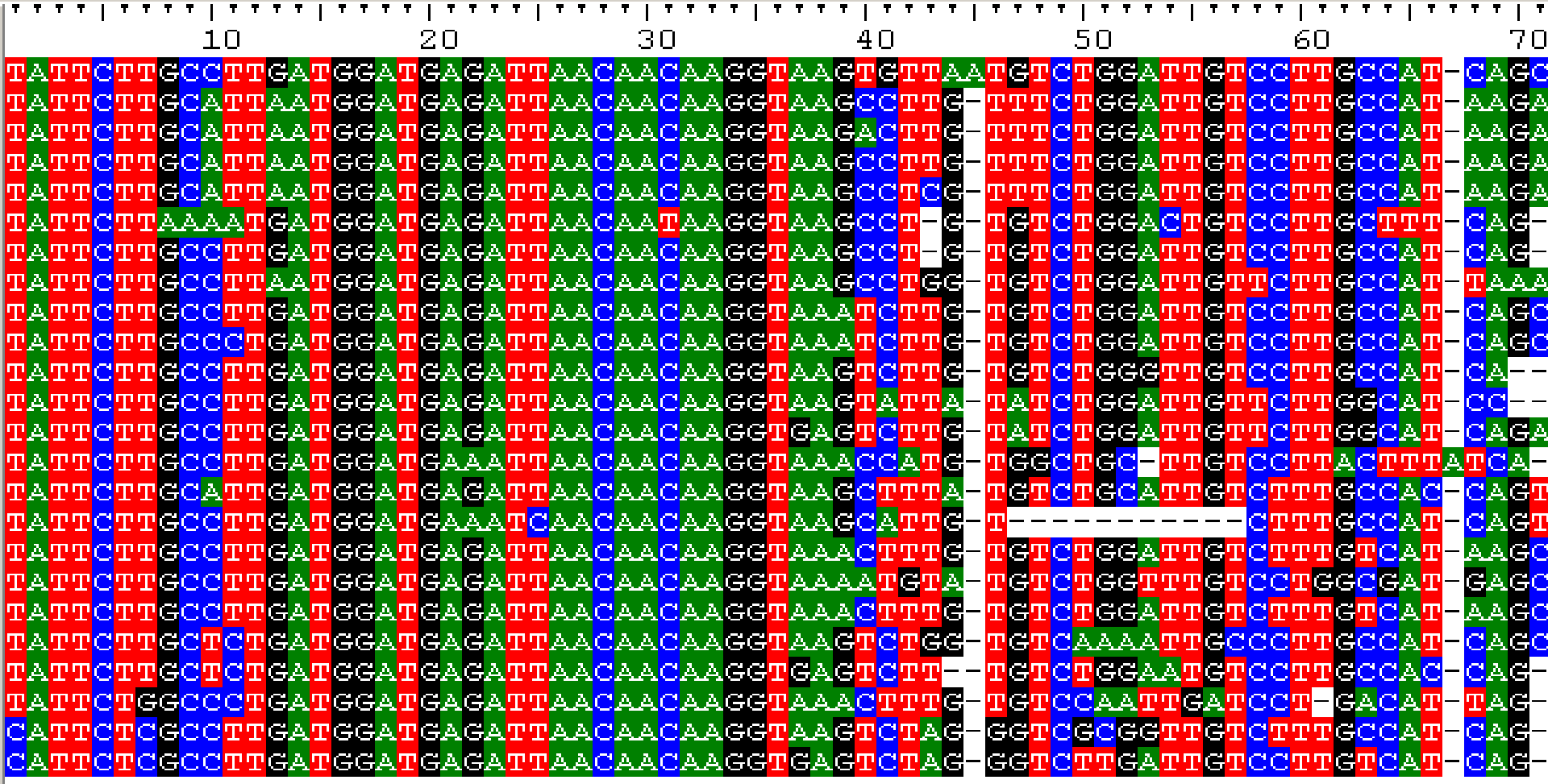
As well as introns within genes, there are large gaps of non-coding sequence between genes. In fact, only around 1% of the mammalian genome is made up of coding DNA)



Exons (coding parts of the gene) tend to be relatively conserved across taxa. Introns are more variable. Below we can see part of an exon and an intron.

The screenshot displays the BioEdit Sequence Alignment Editor interface. The main window shows a multiple sequence alignment of 24 sequences, all labeled 'TMC1' followed by a species name. The sequences are color-coded by nucleotide: Adenine (A) in blue, Thymine (T) in red, Cytosine (C) in green, and Guanine (G) in yellow. The alignment is viewed in a window titled 'C:\Documents and Settings\Kalina D\Desktop\Tmc1_steve intexons.fas' with 24 total sequences. The sequence mask is set to 'None' and the numbering mask is also 'None'. The start ruler is at position 1. The alignment shows a highly conserved region (exon) from approximately position 10 to 65, followed by a highly variable region (intron) from approximately position 65 to 90. The species listed on the left include Horse, Human, Chimp, Orang, Macaque, Bushbaby, Mouselemur, Tarsier, BottDolphin, Cow, Alpaca, Dog, Cat, Tenrec, Armadillo, Sloth, Hare, Pika, Rabbit, Shrew, Kangrat, Rockhyrax, Rat, and Mouse.

Species	10	20	30	40	50	60	70	80	90												
TMC1 Horse	TATTC	TTGCC	TTGAT	GGATG	GAGATT	AACA	CAAGGT	AAGTGT	AA	TGTC	TGGATT	GTCC	TTGCC	AT	CAGCAT	GTTACT					
TMC1 Human	TATTC	TTGC	ATTAA	TGGAT	GAGATT	AACA	CAAGGT	TAAG	CCTTG	TTTCT	TGGATT	GTCC	TTGCC	AT	AAAGAT	GTTGTT					
TMC1 Chimp	TATTC	TTGC	ATTAA	TGGAT	GAGATT	AACA	CAAGGT	TAAG	ACTTG	TTTCT	TGGATT	GTCC	TTGCC	AT	AAAGAT	GTTGTT					
TMC1 Orang	TATTC	TTGC	ATTAA	TGGAT	GAGATT	AACA	CAAGGT	TAAG	CCTTG	TTTCT	TGGATT	GTCC	TTGCC	AT	AAAGAT	GTTGTT					
TMC1 Macaque	TATTC	TTGC	ATTAA	TGGAT	GAGATT	AACA	CAAGGT	TAAG	CCTTG	TTTCT	TGGATT	GTCC	TTGCC	AT	AAAGAT	GTTGTT					
TMC1 Bushbaby	TATTC	TTT	AAAA	TGAT	GGAT	GAGATT	AACA	TAAGGT	TAAG	CCTTG	G	TGTC	TGGATT	GTCC	TTGCC	TTT	CAG	---	TGTT		
TMC1 Mouselemur	TATTC	TTGC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CCTTG	G	TGTC	TGGATT	GTCC	TTGCC	AT	CAG	---	TGTT			
TMC1 Tarsier	TATTC	TTGC	TTAA	TGGAT	GAGATT	AACA	CAAGGT	TAAG	CCTTG	GG	TGTC	TGGATT	GTCC	TTGCC	AT	TAAGAT	GTTGTT				
TMC1 BottDolphin	TATTC	TTGC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	TGGATT	GTCC	TTGCC	AT	CAGCAT	GTTGTT					
TMC1 Cow	TATTC	TTGCC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	TGGATT	GTCC	TTGCC	AT	CAGCAT	GTTGTT					
TMC1 Alpaca	TATTC	TTGC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	TGGATT	GTCC	TTGCC	AT	CA	---	---	---			
TMC1 Dog	TATTC	TTGCC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TATCT	TGGATT	GTCC	TTGCC	AT	CC	---	ATGT	---	GTT		
TMC1 Cat	TATTC	TTGC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAGT	CTTG	TATCT	TGGATT	GTCC	TTGCC	AT	CAGAGT	GTTATT					
TMC1 Tenrec	TATTC	TTGC	TTGAT	GGAT	GAAAT	AACA	CAAGGT	TAAG	CCATG	TGGCT	GC	TTGTC	TTAC	TTTATCA	---	---	---	---	GTGTT		
TMC1 Armadillo	TATTC	TTGC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	TGCATT	GTCC	TTGCC	AC	CAGTGT	ATTGTT					
TMC1 Sloth	TATTC	TTGCC	TTGAT	GGAT	GAAAT	AACA	CAAGGT	TAAG	CTTG	T	---	---	---	---	C	TTTGCC	AT	---	---	TTTATT	
TMC1 Hare	TATTC	TTGC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	TGGATT	GTCC	TTGTC	CAT	AA	GCAT	AT	---	---	TTT	
TMC1 Pika	TATTC	TTGC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	TGGATT	GTCC	TTGCC	AT	GAGCAT	GTTCTT					
TMC1 Rabbit	TATTC	TTGCC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	TGGATT	GTCC	TTGTC	CAT	AA	GCAT	AT	---	---	TTT	
TMC1 Shrew	TATTC	TTGCT	CTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	AAAA	TTGCC	TTGCC	AT	CAGCT	GGAT	GTT				
TMC1 Kangrat	TATTC	TTGCC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAGT	CTTG	TGTC	TGGAA	TTGCC	TTGCC	AC	CAG	---	---	---	---	TTTAC	
TMC1 Rockhyrax	TATTC	TTGCC	TTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	TGTC	CAA	TTGAT	CTTG	GA	CAT	TAG	---	---	---	TTTAA	
TMC1 Rat	CATTC	TCG	CTTGAT	GGAT	GAGATT	AACA	CAAGGT	TAAG	CTTG	GGT	TCG	GGTTG	CTTG	TTGCC	AT	CAG	---	---	---	---	TGTTCAA
TMC1 Mouse	CATTC	TCG	CTTGAT	GGAT	GAGATT	AACA	CAAGGT	TAGT	CTAG	GGT	CTTGAT	TTGTC	TTGTC	CAT	CAG	---	---	---	---	---	TGTTCAA



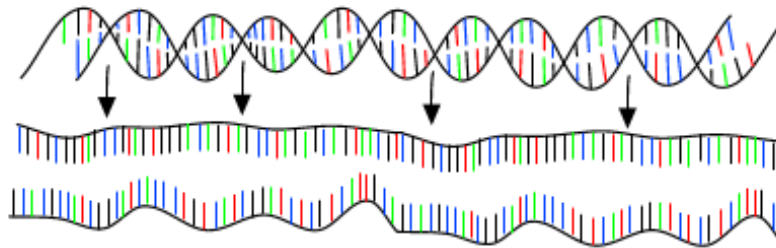
Exonic region

Intronic region

We use PCR to amplify DNA

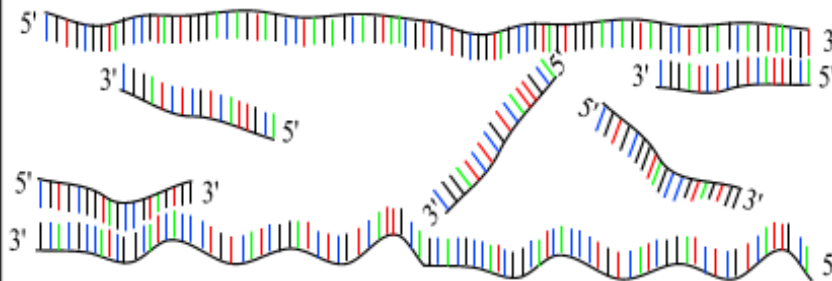
PCR : Polymerase Chain Reaction

30 - 40 cycles of 3 steps :



Step 1 : denaturation

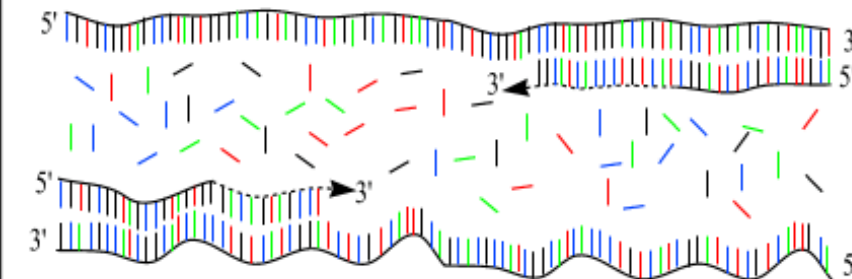
1 minut 94 °C



Step 2 : annealing

45 seconds 54 °C

forward and reverse primers !!!



Step 3 : extension

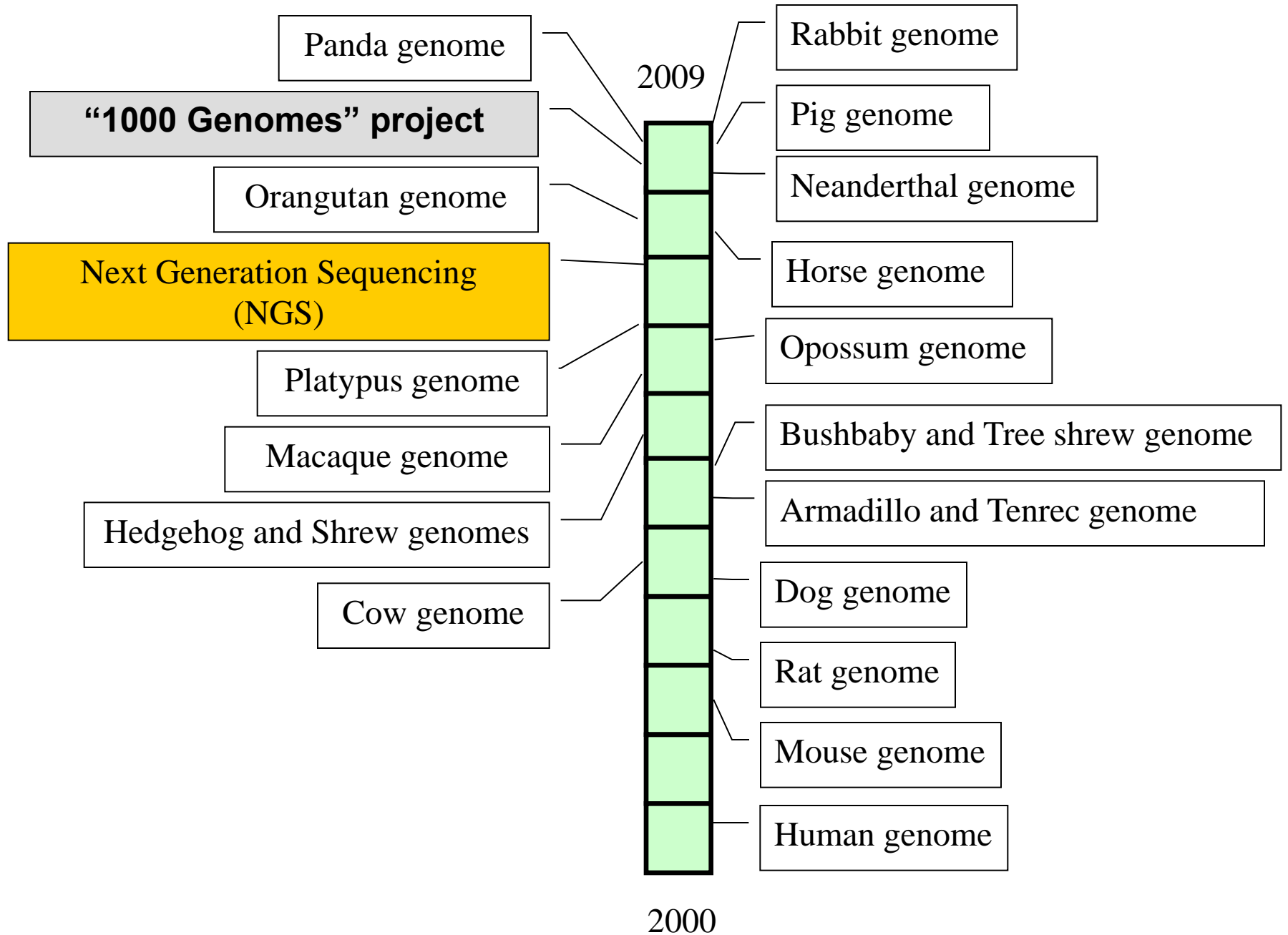
2 minutes 72 °C
only dNTP's

2. Current and future sequencing

Until 2007 we all used Sanger sequencing.
Several large genome projects were conducted at huge expense

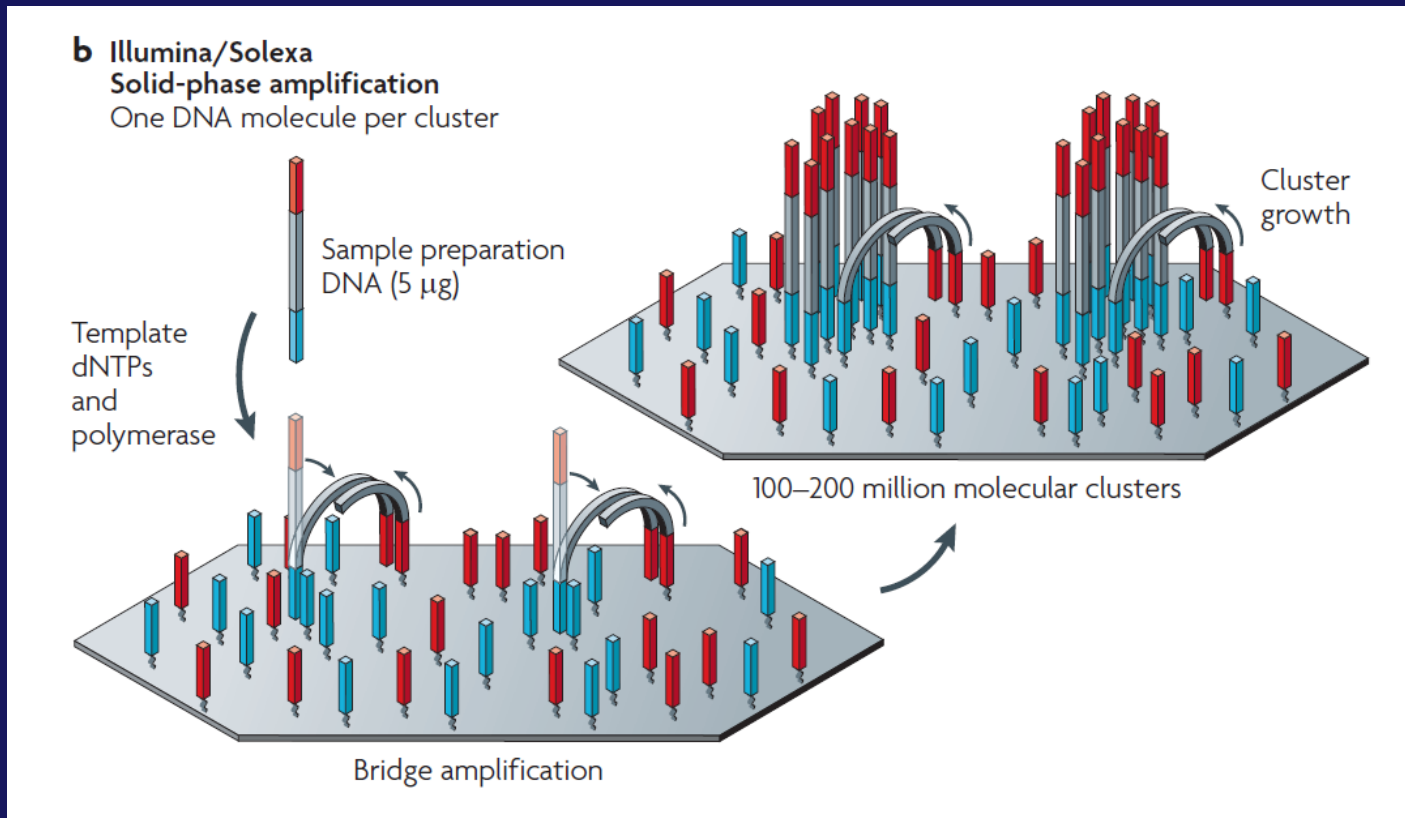
In 2007 several companies released technologies, termed “Next Generation Sequencing”
(shotgun sequencing of small reads)

This means more and more genomes are now being produced

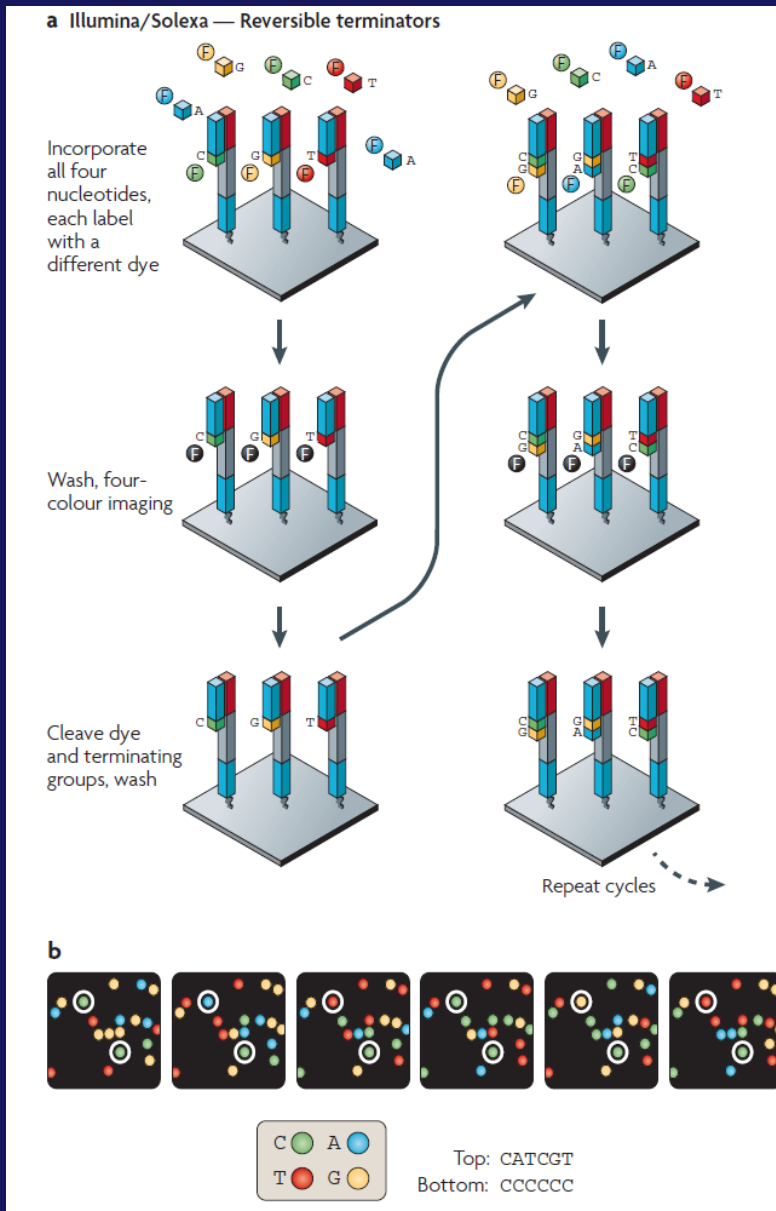


Next generation sequencing

- Based on shotgun sequencing
- Adaptors containing universal priming sites are ligated to ends of the DNA fragment
- DNA templates amplified clonally to get clusters



- Fluoro-labelled dNTPS washed over the strand
- Each time a photo is taken



NGS by Illumina



2 years

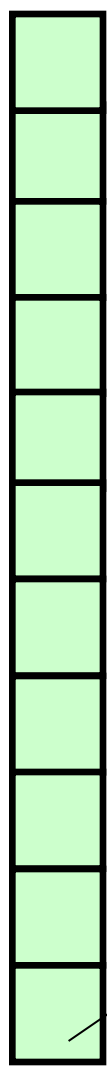


Up to 6.5 Gb per day
640 million paired-end reads

Up to 25 Gb per day
2 billion paired-end reads

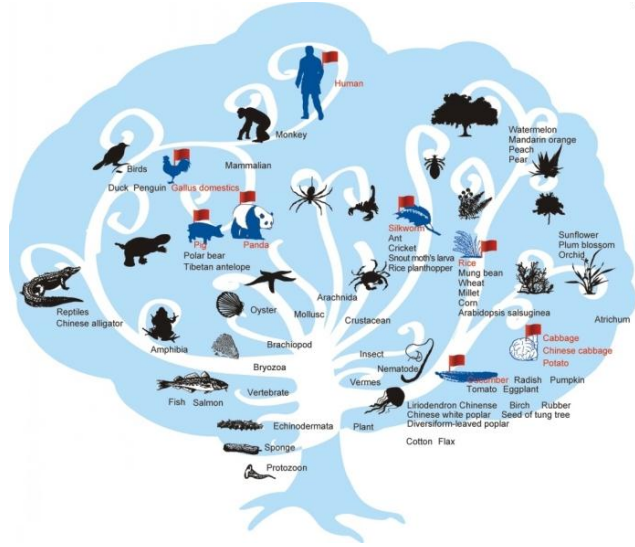
~60x coverage of a human genome in a single run for under \$10,000

2019



2010

**BGI
launches
"1000
Genome"
initiative**



In Process
Tibetan antelope
Polar bear
Camel
Puma

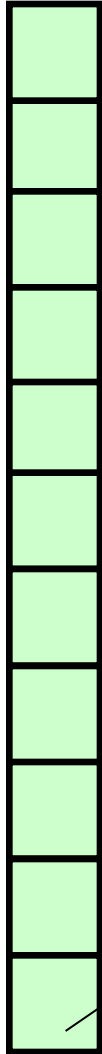
Next
Hyrax
Potto
Wombat
Chinese dolphin
Donkey
Porpoise
Asian lion
Beluga whale
Giraffe
Aardwolf
Whale
Mole-rat
Hamster

Tasmanian devil genome

Gorilla and Gibbon genome

Myotis genome

2019



2010

**BGI
launches
“1000
Genome”
initiative**

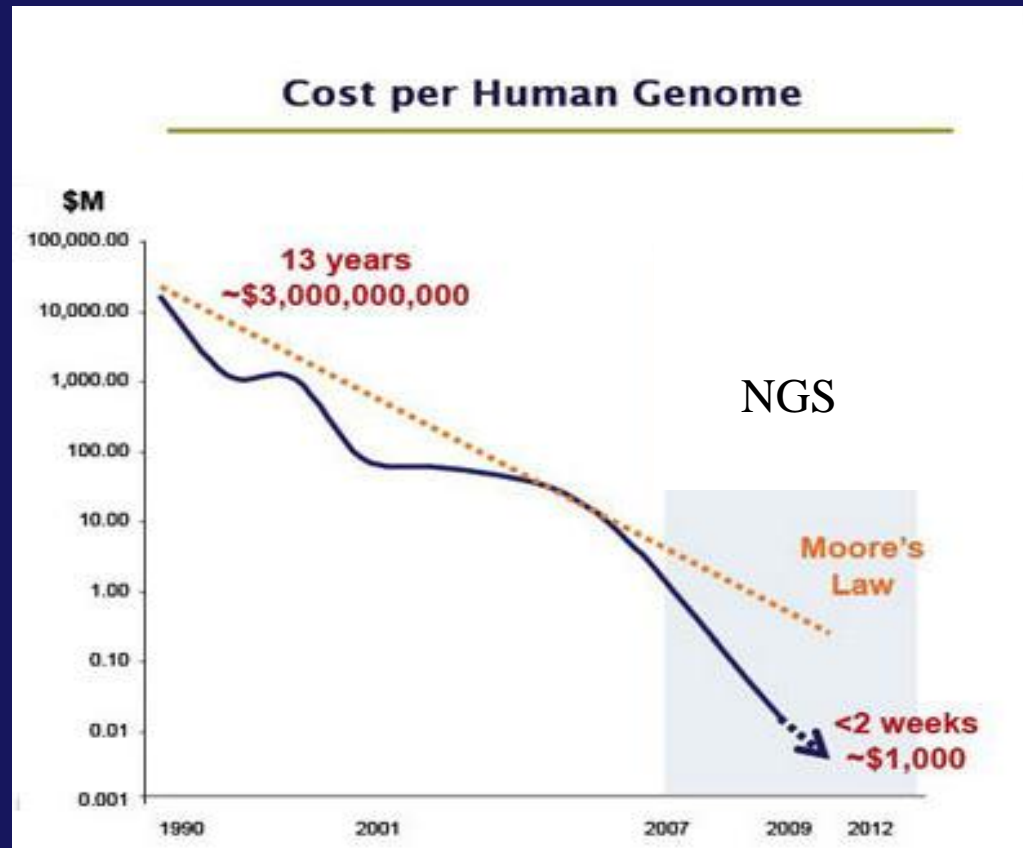
Tasmanian devil genome

Gorilla and Gibbon genome

Myotis genome

- Other existing bat “genome” projects**
- Hipposideros armiger*
 - Rhinolophus ferrumequinum*
 - Rhinolophus sinicus*
 - Rhinolophus affinis*
 - Rhinolophus yunanensis*
 - Megaderma lyra*
 - Eidolon helvum*
 - Pteronotus parnellii*
 - Pteropus vampyrus*
 - Tadarida brasiliensis*
 - Eptesicus fuscus*
 - And many more.....*

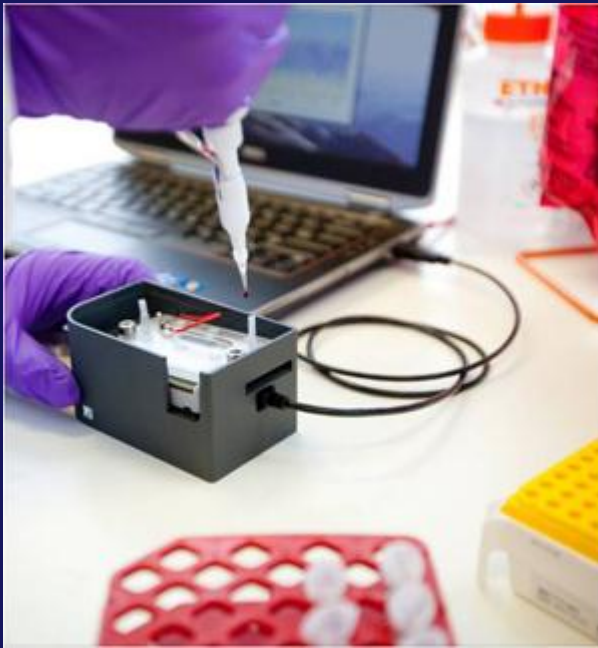
Affordability?



- Gene sequencers outpace microchips
- Numerous companies promise genomes for 1K USD within just 2-5 years
- This means these methods will be within our financial reach.
- Some predict a genome will be less than 100 dollars in a few years.

Within one or two years from now

- Single molecule approaches (this could mean our museum samples are useful for genome sequencing)
- Ultra portable sequencers (could be useful for field work)



Oxford Nanopore's "minION"

So how will we benefit from these methods?

Development of new DNA markers

Examples of using genome comparative data

Future of phylogenomics and population genomics

Development of new DNA markers I

- Microsatellite discovery by mining published genome data

Shikano *et al.* (2010) *BMC Genomics* **11**: 334

Sequenced genomes for microsatellite marker development in nine-spined sticklebacks

- Microsatellite development by low coverage genome sequencing

Abdelkrim (2009) *BioTechniques* **46**: 185-192

blue duck DNA → 454 sequencing → 17215 reads → >200 loci → 24 primer sets

Table 2. Summary Results of the Development of Microsatellite Markers Following the Genomic Approach on Three Other Species

Species	Class	Number of reads	Minimum number of repeats*	Number of microsatellites detected	Number of potential primer pairs
<i>Maritrema novaezealandensis</i> ¹	Trematoda	31120	5,4,4	676	46
<i>Motuweta isolata</i> ²	Insecta	52059	4,4,4	472	134
<i>Powelliphanta augusta</i> ²	Gastropoda	35196	6,6,6	2541	170

*For di-, tri-, and tetranucleotide, respectively

¹Unpublished data provided by Yuri Springer (University of Otago)

²Unpublished data provided by Thomas Buckley (Landcare Research)

Development of new DNA markers II

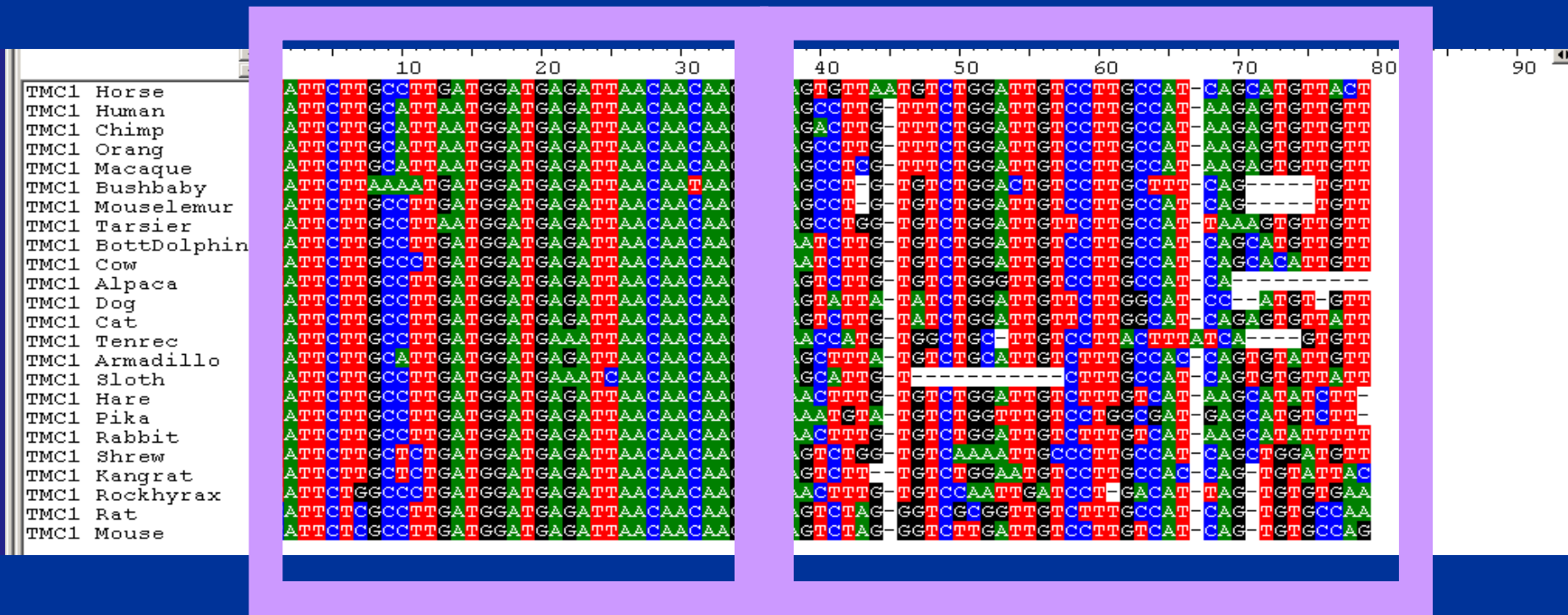
- Non-coding and coding DNA from mining genome data

“Transmembrane channel-like protein 1” gene

	10	20	30	40	50	60	70	80	90
TMC1 Horse	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTGTTAA	TGTCCTGGATTGTCCCTTGCCAT	CAGCATGTTACT						
TMC1 Human	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGCCTTG	TTTCTGGATTGTCCCTTGCCAT	AAGAGTGTGTT						
TMC1 Chimp	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGACTTG	TTTCTGGATTGTCCCTTGCCAT	AAGAGTGTGTT						
TMC1 Orang	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGCCTTG	TTTCTGGATTGTCCCTTGCCAT	AAGAGTGTGTT						
TMC1 Macaque	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGCCTTCG	TTTCTGGATTGTCCCTTGCCAT	AAGAGTGTGTT						
TMC1 Bushbaby	TATTCCTTAAAAATGATGGATGAGATTAAACAACAAGGTAAGCCTTG	TGTCCTGGACTGTCCCTTGCTTT	CAG-----TGTT						
TMC1 Mouselemur	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGCCTTG	TGTCCTGGATTGTCCCTTGCCAT	CAG-----TGTT						
TMC1 Tarsier	TATTCCTTGCCCTTAATGGATGAGATTAAACAACAAGGTAAGCCTTGG	TGTCCTGGATTGTCCCTTGCCAT	TAAAAGTGTGTT						
TMC1 BottDolphin	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTTG	TGTCCTGGATTGTCCCTTGCCAT	CAGCATGTTGTT						
TMC1 Cow	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTTG	TGTCCTGGATTGTCCCTTGCCAT	CAGCACAATTGTT						
TMC1 Alpaca	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTTG	TGTCCTGGATTGTCCCTTGCCAT	CA-----						
TMC1 Dog	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTATTA	TATCTGGATTGTCTTTGGCAT	CC--ATGT-GTT						
TMC1 Cat	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTGAGTCTTTG	TATCTGGATTGTCTTTGGCAT	CAGAGTGTATT						
TMC1 Tenrec	TATTCCTTGCCCTTGATGGATGAAAATAACAACAAGGTAAGCCATG	TGGCTGC-TTGTCTTACTTTATCA	-----GTGTT						
TMC1 Armadillo	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGCCTTTA	TGTCCTGCATTGTCTTTGCCAC	CAGTGTATTGTT						
TMC1 Sloth	TATTCCTTGCCCTTGATGGATGAAAATCAACAACAAGGTAAGCATTG	T-----CTTTGCCAT	CAGTGTGTTATT						
TMC1 Hare	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGCCTTTG	TGTCCTGGATTGTCTTTGTCA	AAGCATATCTTT						
TMC1 Pika	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTGTGTA	TGTCCTGGTTTGTCCCTGGCGAT	GAGCATGTCCTT						
TMC1 Rabbit	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGCCTTTG	TGTCCTGGATTGTCTTTGTCA	AAGCATATTTTTT						
TMC1 Shrew	TATTCCTTGCTCTGATGGATGAGATTAAACAACAAGGTAAGTCTTGG	TGTCAAAAATGCCCCGCCAT	CAGCTGGATGTT						
TMC1 Kangrat	TATTCCTTGCTCTGATGGATGAGATTAAACAACAAGGTAAGTCTTCT	TGTCCTGGAAATGCCCCGCCAC	CAG-TGTATTAC						
TMC1 Rockhyrax	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGCCTTTG	TGTCCTGGTTTGTCCCTGGCGAT	TAG-TGTGTGAA						
TMC1 Rat	CAATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTAG	GGTCGCGGTTGTCTTTGCCAT	CAG-TGTGCCAA						
TMC1 Mouse	CAATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTAG	GGTCCTTGATTGTCCCTTGTCA	CAG-TGTGCCAG						

- Non-coding and coding DNA from mining genome data

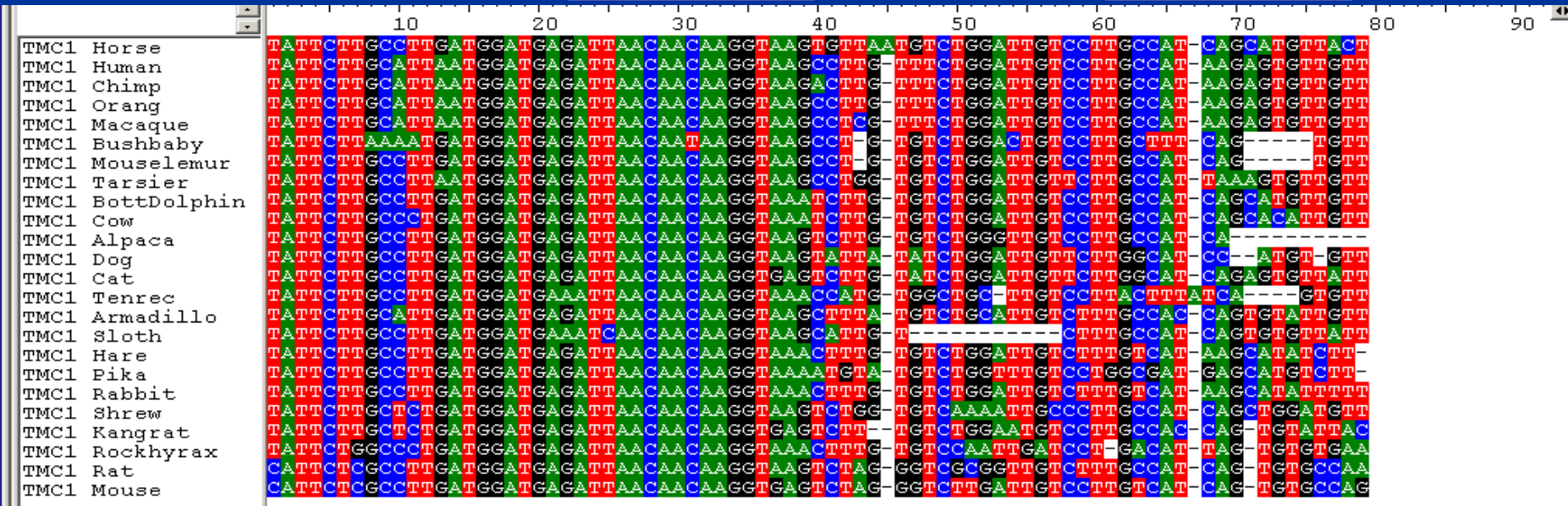
“Transmembrane channel-like protein 1” gene



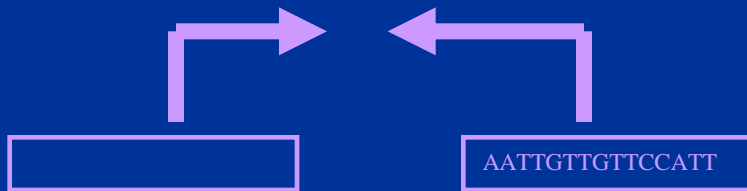
Intron

Exon

“Exon-primed intron-crossing sequences” (EPICs)



Nuclear protein coding loci (NPCL)



	10	20	30	40	50	60	70	80	90
TMC1 Horse	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTGTTAA	TGCTCGGATTGTCCCTTGCCAT	CAGCATGTTACT						
TMC1 Human	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGCCTTG	TTTCTGGATTGTCCCTTGCCAT	AAGAGTGTGTT						
TMC1 Chimp	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGACTTG	TTTCTGGATTGTCCCTTGCCAT	AAGAGTGTGTT						
TMC1 Orang	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGCCTTG	TTTCTGGATTGTCCCTTGCCAT	AAGAGTGTGTT						
TMC1 Macaque	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGCCTTCG	TTTCTGGATTGTCCCTTGCCAT	AAGAGTGTGTT						
TMC1 Bushbaby	TATTCCTTAAAAATGATGGATGAGATTAAACAACAAGGTAAGCCTTG	TGCTCGGACTGTCCCTTGCTTT	CAG-----TGT						
TMC1 Mouselemur	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGCCTTG	TGCTCGGATTGTCCCTTGCCAT	CAG-----TGT						
TMC1 Tarsier	TATTCCTTGCCCTTAATGGATGAGATTAAACAACAAGGTAAGCCTTGG	TGCTCGGATTGTCTTGCCAT	TAAAGTGTGTT						
TMC1 BottDolphin	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTTG	TGCTCGGATTGTCCCTTGCCAT	CAGCATGTTGTT						
TMC1 Cow	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTTG	TGCTCGGATTGTCCCTTGCCAT	CAGCATGTTGTT						
TMC1 Alpaca	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTTG	TGCTCGGATTGTCCCTTGCCAT	CA-----						
TMC1 Dog	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTATTTA	TATCTGGATTGTCTTGCCAT	CC--ATGT-GTT						
TMC1 Cat	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTGAGTCTTTG	TATCTGGATTGTCTTGCCAT	CAGAGTGTATT						
TMC1 Tenrec	TATTCCTTGCCCTTGATGGATGAAAATTAACAACAAGGTAAGCCATG	TGGCTGC-TTGTCTTACTTTATCA	-----GTGTT						
TMC1 Armadillo	TATTCCTTGCAATTAATGGATGAGATTAAACAACAAGGTAAGCCTTTA	TGCTCGCATTTGTCTTGCCAC	CAGTGTATTGTT						
TMC1 Sloth	TATTCCTTGCCCTTGATGGATGAAAATCAACAACAAGGTAAGCATTTG	T-----CTTGCCAT	CAGTGTGTTATT						
TMC1 Hare	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGCCTTTG	TGCTCGGATTGTCTTTGTCAAT	AAGCATATCTTT						
TMC1 Pika	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTAAATGTA	TGCTCGGTTTGTCTTGCCGAT	GAGCATGTCCTT						
TMC1 Rabbit	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGCCTTTG	TGCTCGGATTGTCTTTGTCAAT	AAGCATATTTTTT						
TMC1 Shrew	TATTCCTTGCTCTGATGGATGAGATTAAACAACAAGGTAAGTCTTGG	TGCTCAAAAATTGCCCTTGCCAT	CAGCTGGATGTT						
TMC1 Kangrat	TATTCCTTGCTCTGATGGATGAGATTAAACAACAAGGTGAGTCTTTG	TGCTCGGAAATGCTCTTGCCAC	CAG-TGTATTAC						
TMC1 Rockhyrax	TATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGCCTTTG	TGCTCAAAATTGATCCCTTGACAT	TAG-TGTGTGAA						
TMC1 Rat	CAATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTAG	GGTCGCGGTTGTCTTGCCAT	CAG-TGTGCCAA						
TMC1 Mouse	CAATTCCTTGCCCTTGATGGATGAGATTAAACAACAAGGTAAGTCTTAG	GGTCTTGATTGTCTTGCTCAAT	CAG-TGTGCCAG						

We need to be prepared for these new methods and technologies.

It is important to think about how we maximise the potential benefit of our samples, particularly for our ongoing collecting and research

Collecting and storing genetic resources

The best approaches to collecting and storing bat material for genetic analysis will depend on the needs of the samples

Very often we cannot predict the future technologies so it is important to take care now to safeguard the value of our material in the future

Most genetic analyses are based on DNA

Uses of DNA work include:

Sequencing for phylogenetic and phylogeographic analyses

Species identification via bar coding (COI) and other loci

Species ID etc

Microsatellite genotyping

Functional genes

Advantages of DNA

Relatively stable

Evenly distributed across all cells
(same result from muscle versus wing versus liver)

Advantages of using introns as a source of variation

Disadvantages of DNA

Introns and inter-genic areas can also make primer design difficult

Exonic within genes areas might be far apart from each other

Lots of Intergenic sequence

Work on RNA is becoming more important

RNA can be studied to determine expression in different tissue types

No introns or Intergenic regions, so get more gene sequence per dollar

Disadvantages of RNA

Degrades rapidly

Need more material to get enough

Need multiple tissue to obtain all genes

For amplification, need to convert to DNA first

DNA and RNA need to be collected and stored differently

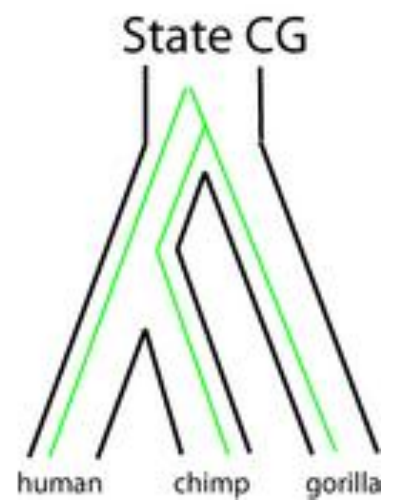
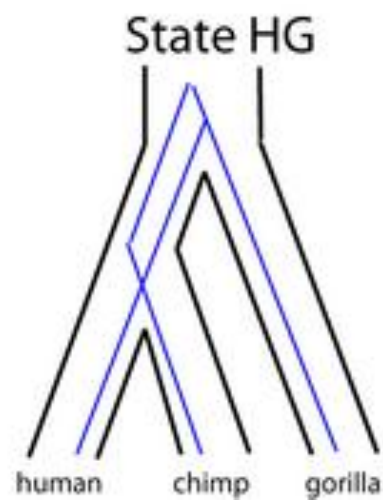
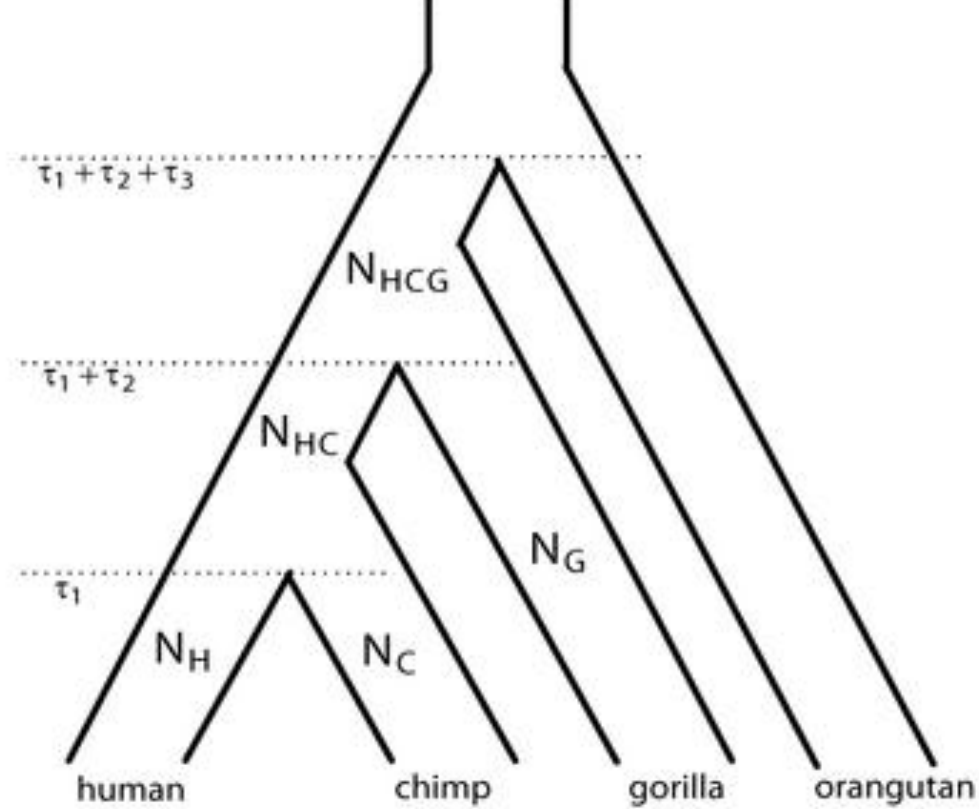
Tissue preservation methods

RNAlater	Liquid nitrogen
100% Ethanol	Dry ice
70% Ethanol	Tissue lysis buffer
Formulin	AllProtect
IMS	Silica gel
VTM	Freezing (-20)
DMSO	Freezing (-80)

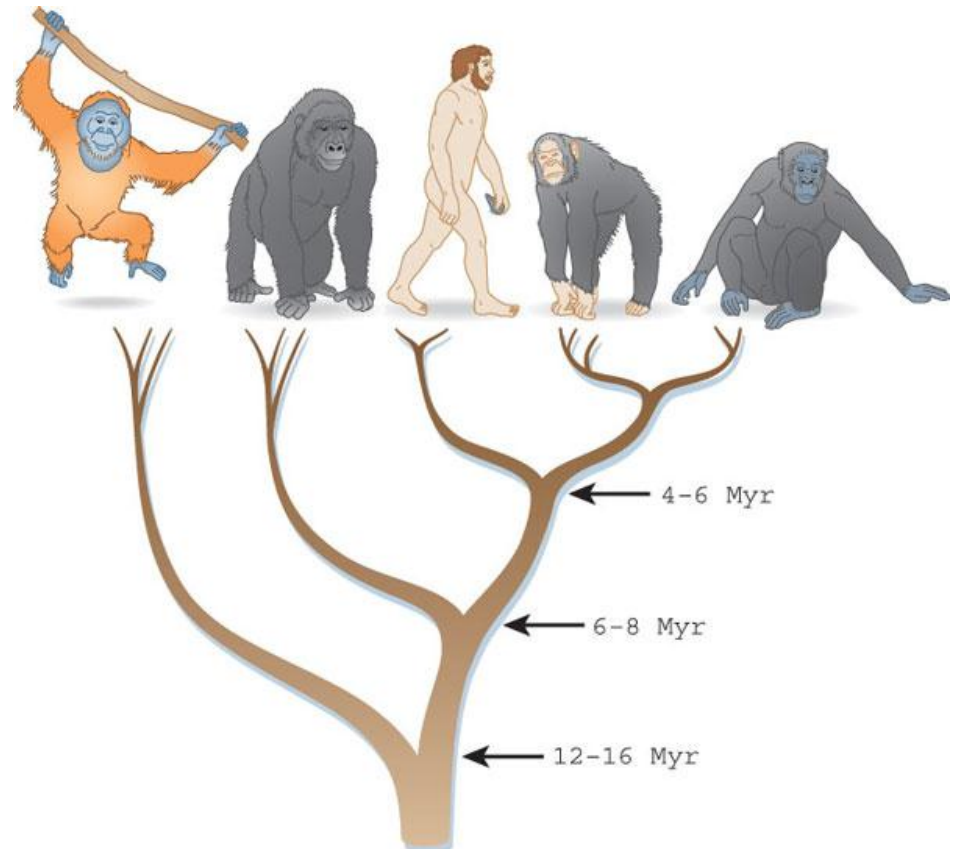
Why do phylogenetic trees sometimes disagree with other datasets?

Why do phylogenetic trees sometimes disagree with other datasets?

1. Incomplete sorting

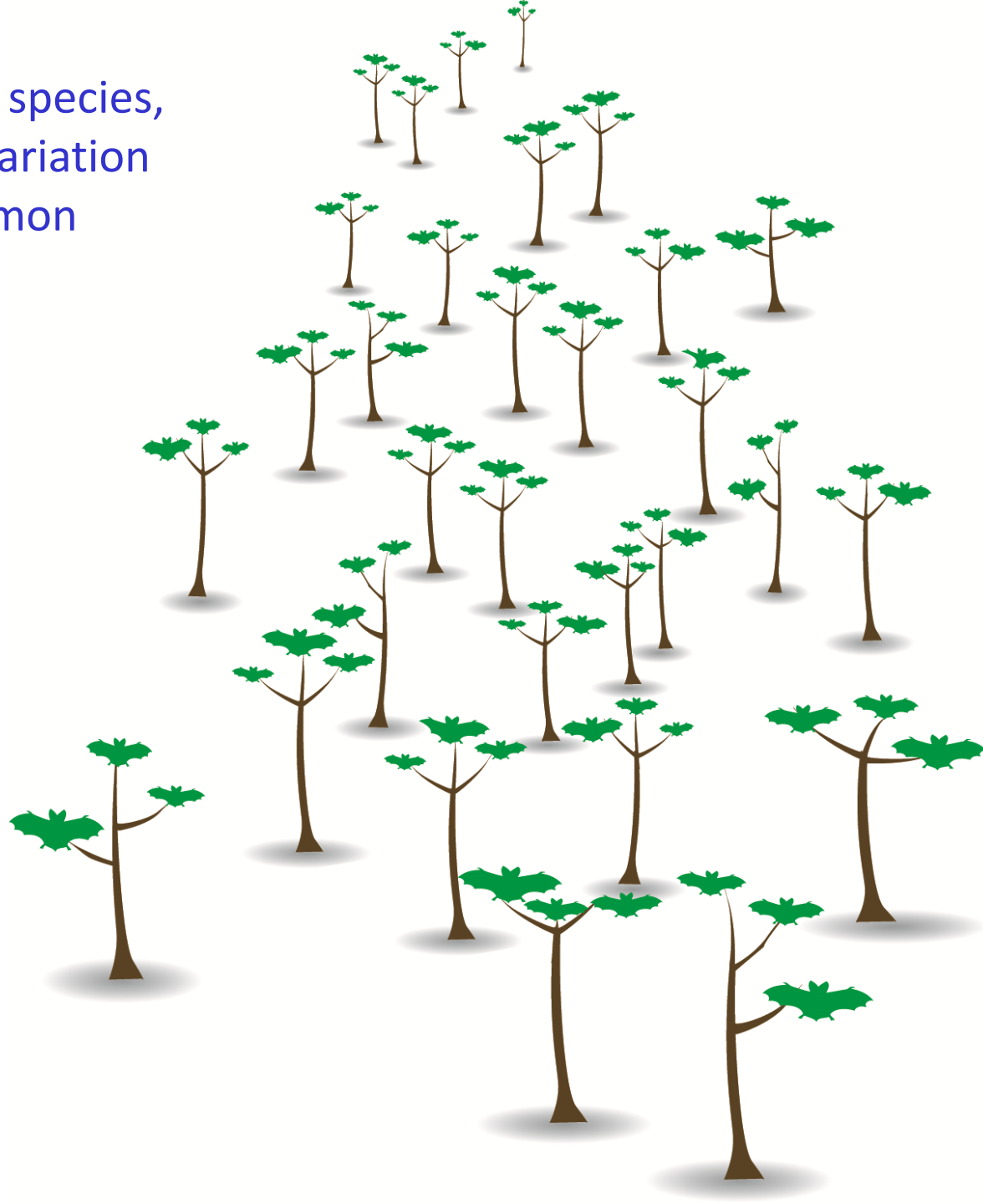


For 1% of genome, humans more closely related to orang utans than to chimps

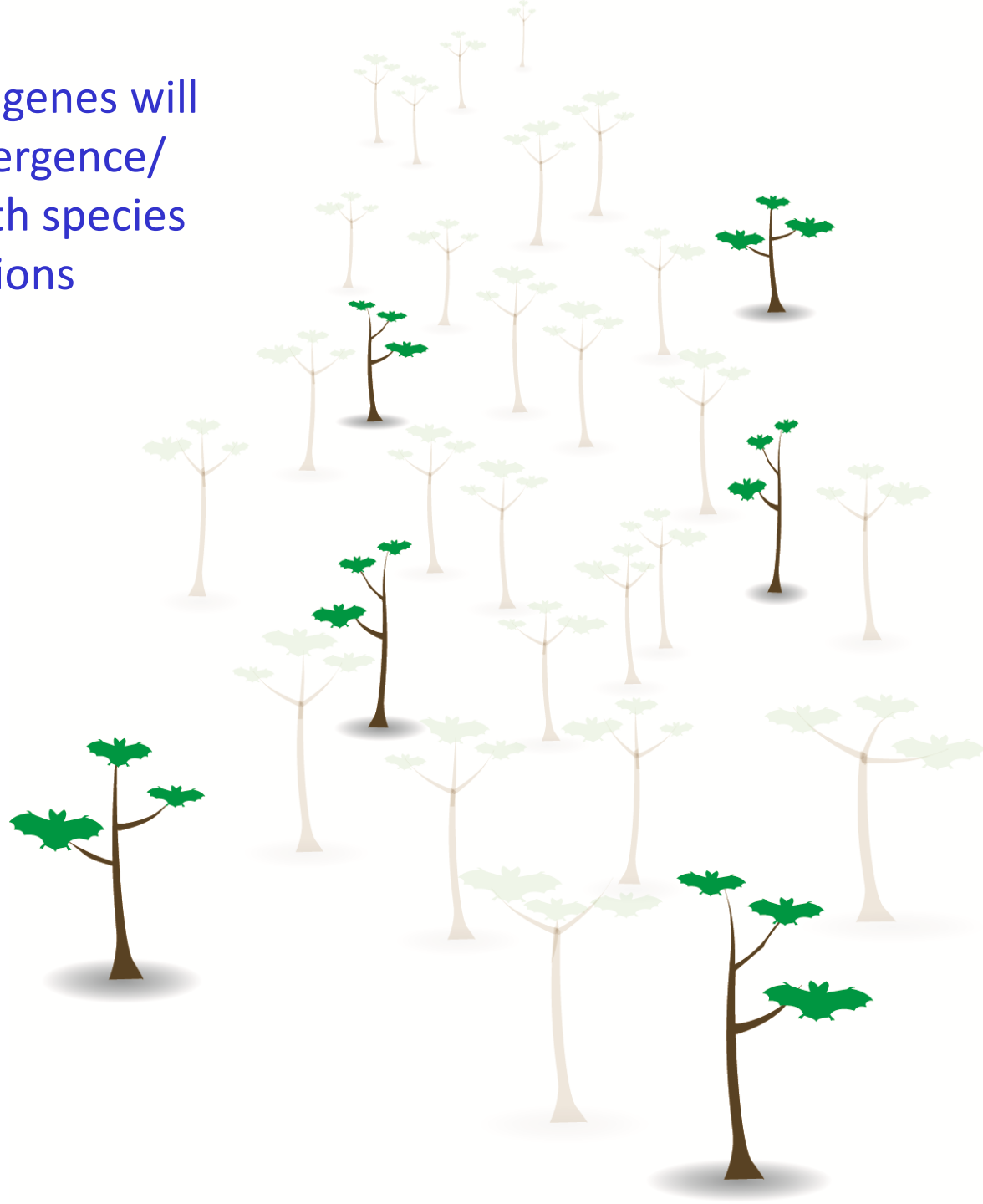


Genome Res. 2011 March; 21(3): 349–356.

For sister species,
shared variation
common



Only a few genes will
show divergence/
sorting with species
divisions



Why do phylogenetic trees sometimes disagree with other datasets?

1. Incomplete sorting
2. Long branch attraction

Why do phylogenetic trees sometimes disagree with other datasets?

1. Incomplete sorting
2. Long branch attraction
3. Introgression

Introgression

Movement of genes from one taxon to another following mating

Previous thought to be uncommon in wild mammals

Now known to be widespread in mammals, incl. bats!

Hybridization between black (*Pteropus alecto*) and grey-headed (*P. poliocephalus*).
Webb & Tidemann (1995) Australian Mammalogy, 18, 19-26.

Hybridization in Peters' tent-making bat (*Uroderma bilobatum*: Phyllostomidae).
Hoffmann et al (2003) Molecular Ecology, 12, 2981-2993.

Berthier et al (2006) Hybridization between *Myotis myotis* and *Myotis blythii*.
Proceedings of the Royal Society B: Biological Science, 273, 3101-3109.

Hulva et al (2010) Hybridisation in the genus *Pipistrellus*.
Molecular Ecology, 19, 5417-5431.

Mao et al (2010) Historical hybridisation in *Rhinolophus pearsoni* and *R. yunanensis*. Molecular Ecology, 19, 1352-1366.

Mao et al (2010) Hybridisation in *Rhinolophus affinis* subspecies.
Molecular Ecology, 19, 2754-2769.

Nesi et al (2011) Possible introgression between *Epomophorus gambianus* and *Micropteropus pusillus*
Comptes Rendus Biologies, 334, 544-554.

Example 1: *Rhinolophus pearsoni* and *Rhinolophus yunanensis*



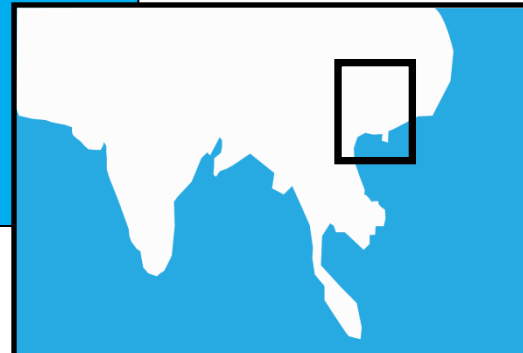
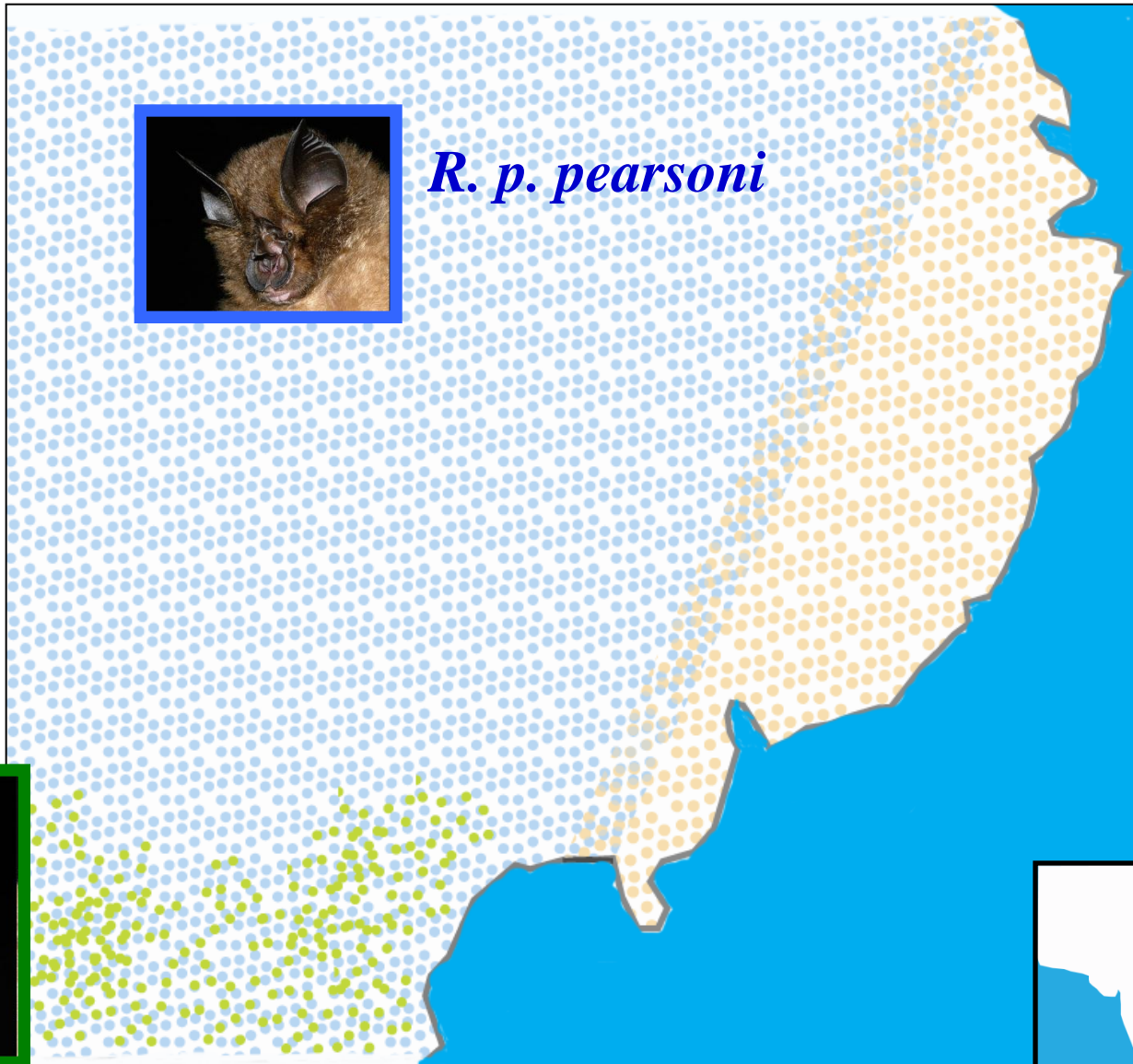
*R. p.
chinensis*



R. p. pearsoni



R. yunanensis



Example 1: *Rhinolophus pearsoni* and *Rhinolophus yunanensis*



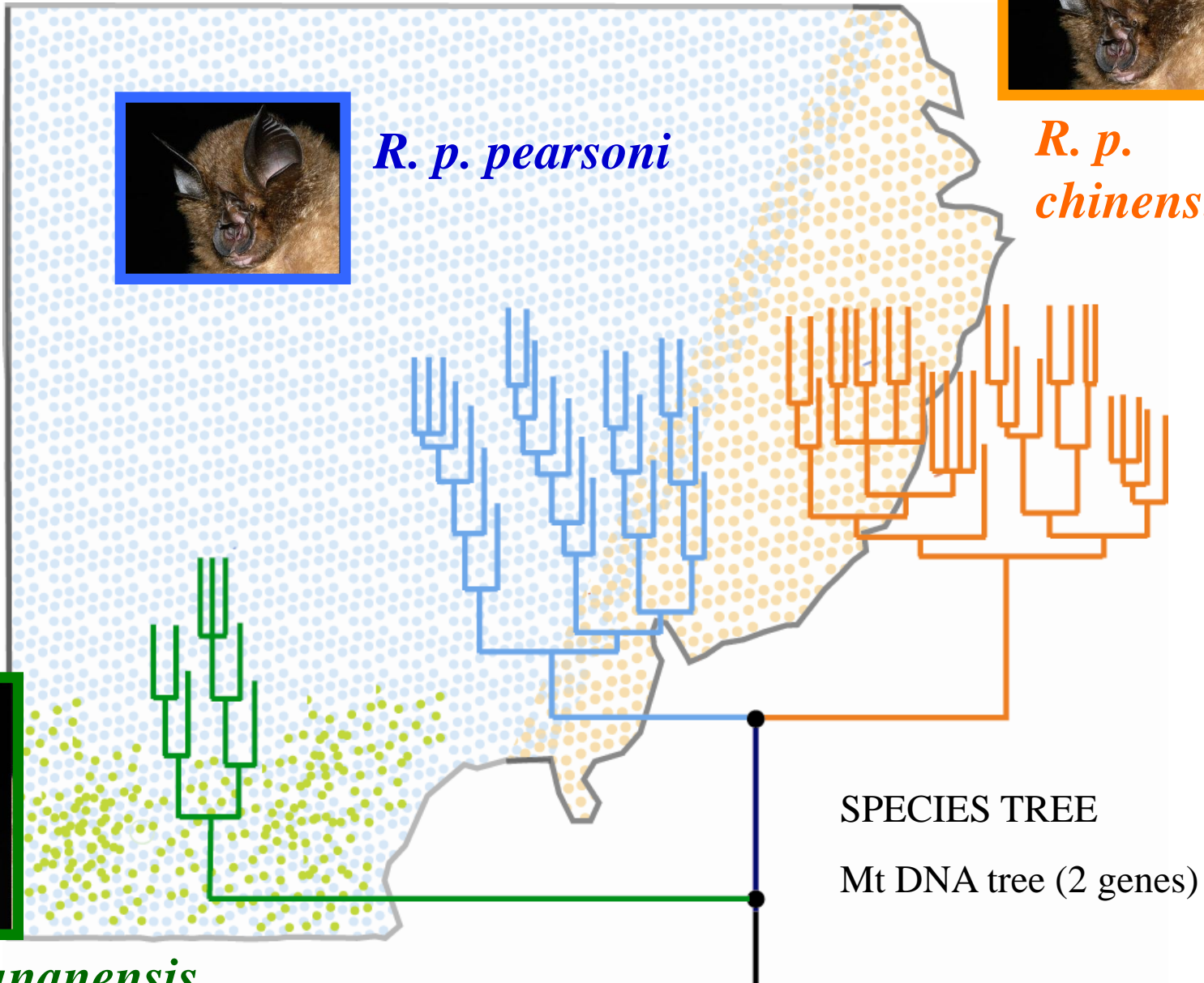
*R. p.
chinensis*



R. p. pearsoni



R. yunanensis



Example 1: *Rhinolophus pearsoni* and *Rhinolophus yunanensis*



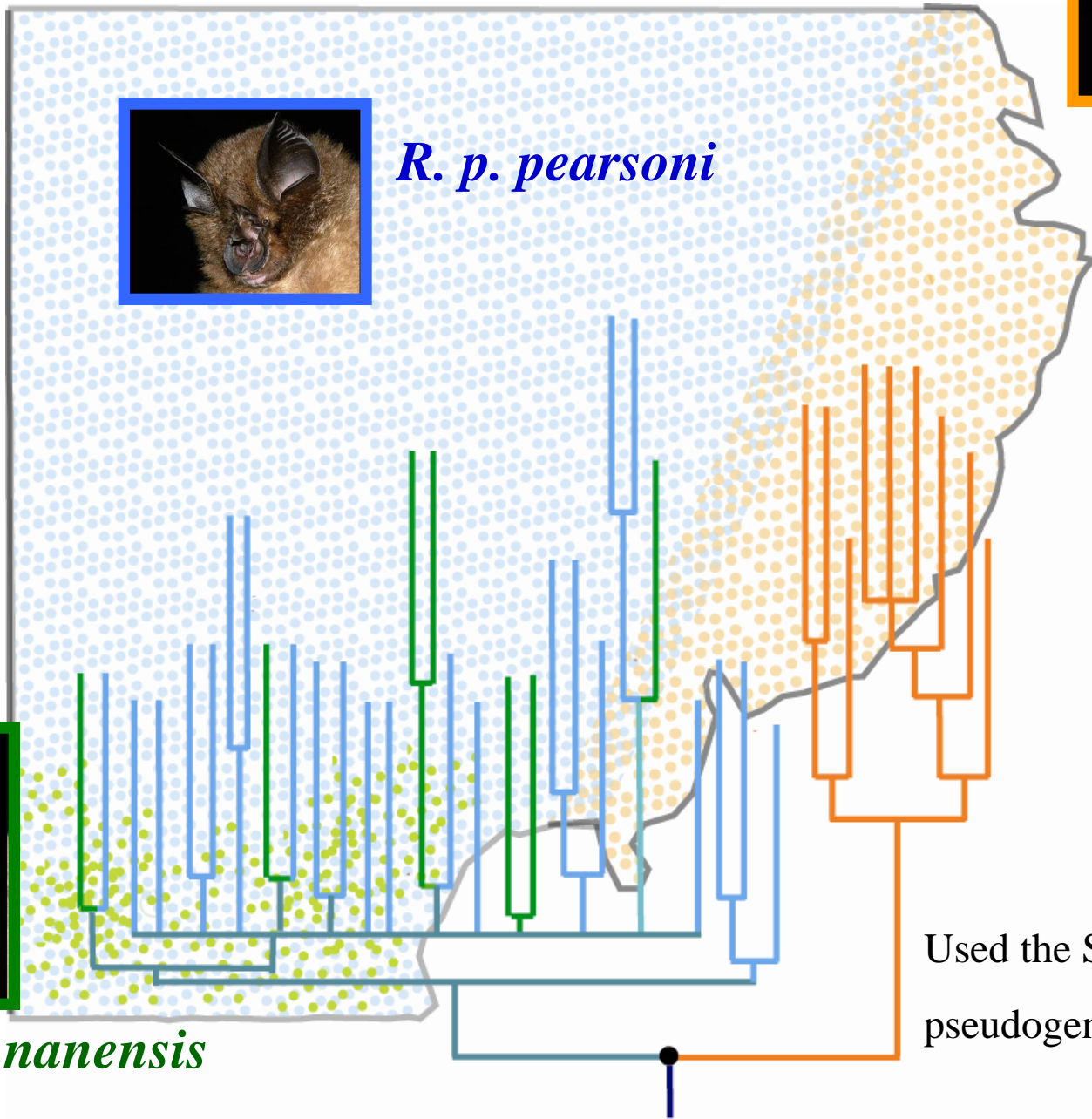
*R. p.
chinensis*



R. p. pearsoni



R. yunanensis

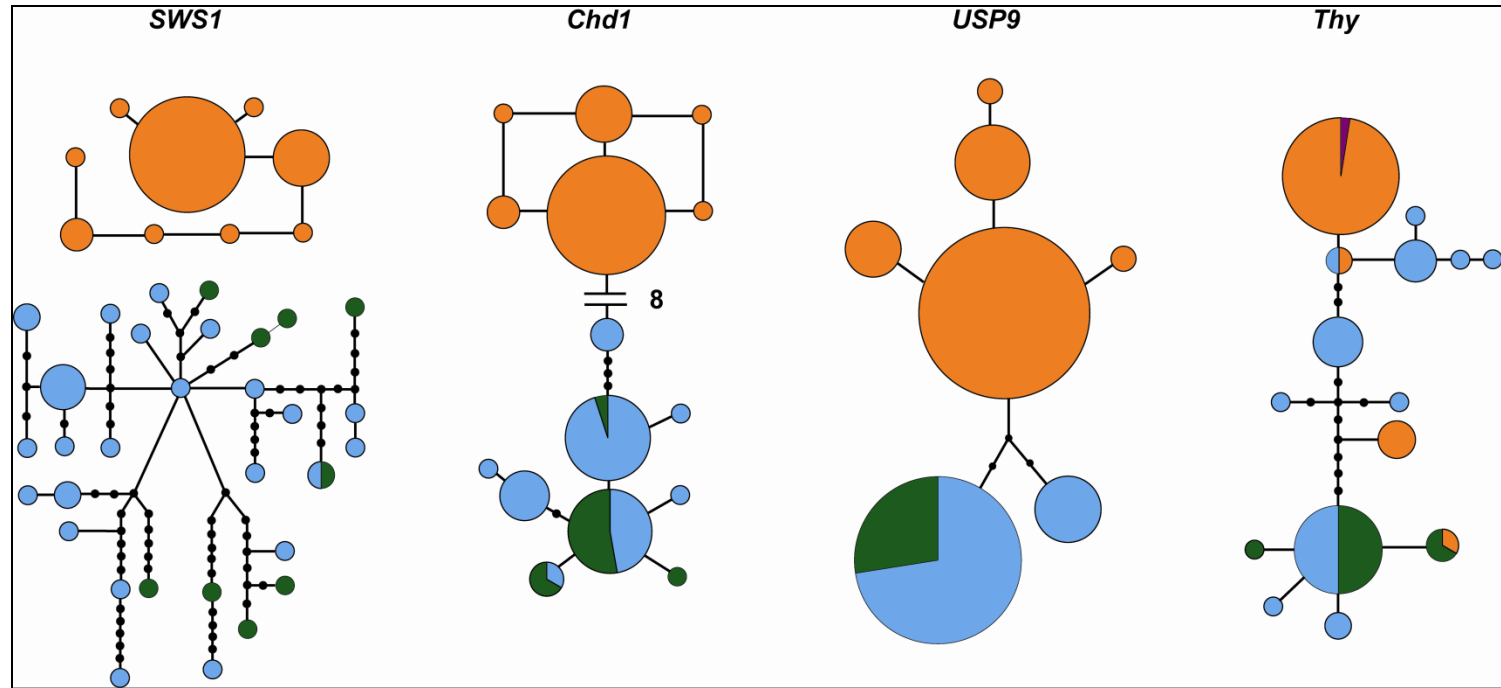


Used the SWS1
pseudogene

Nuclear intron networks



R. p. chinensis

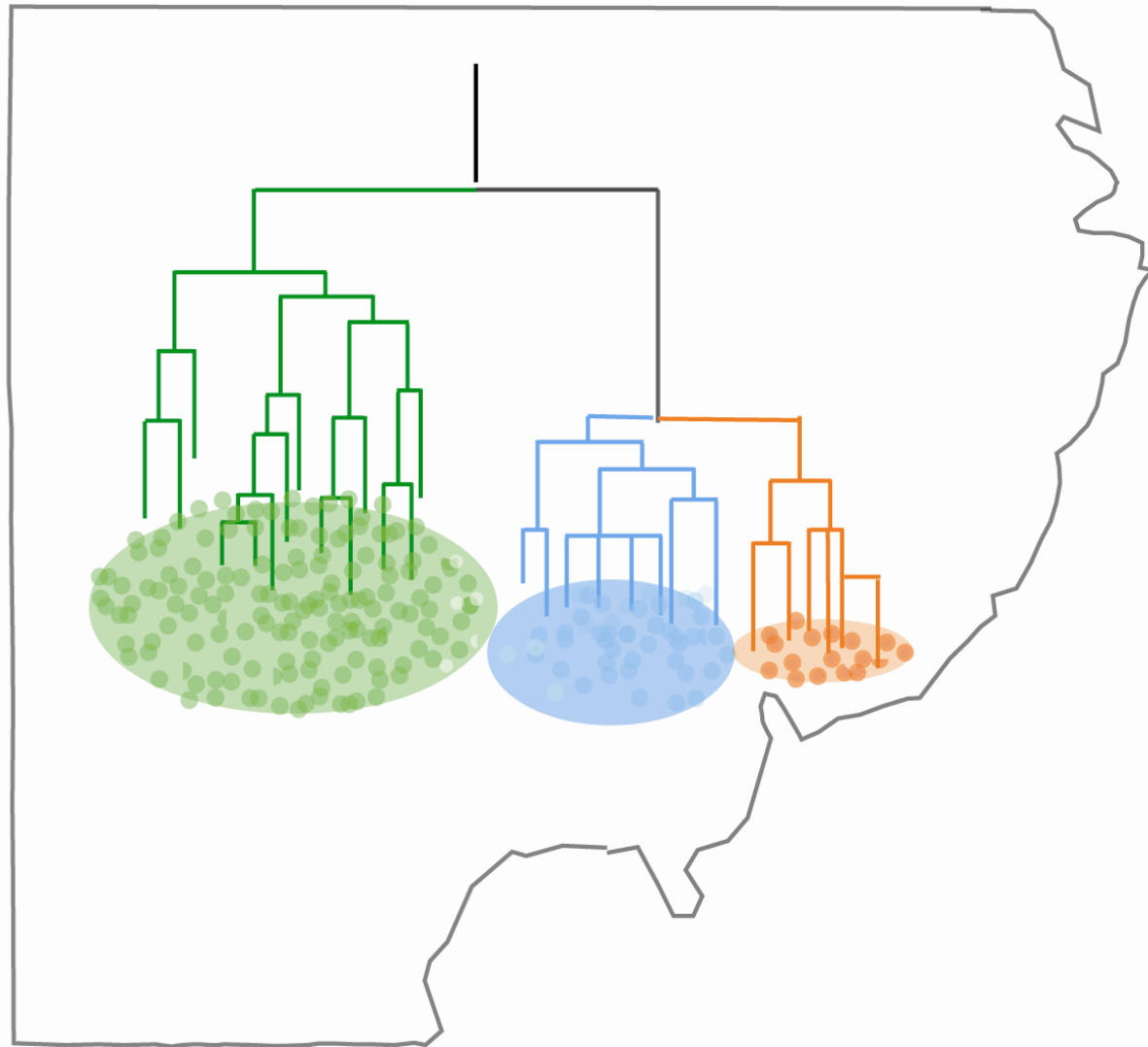


R. yunanensis

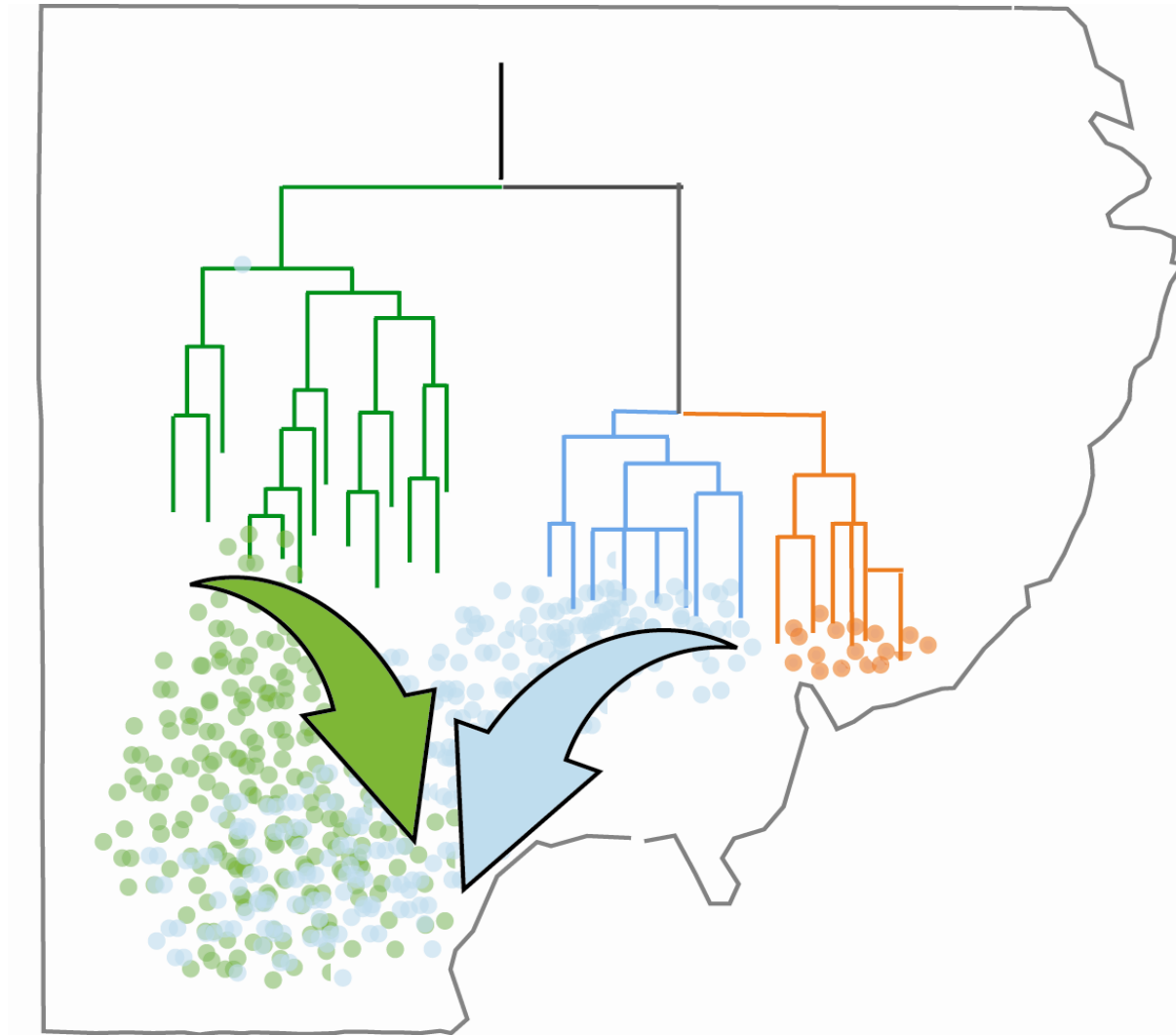


R. p. pearsoni

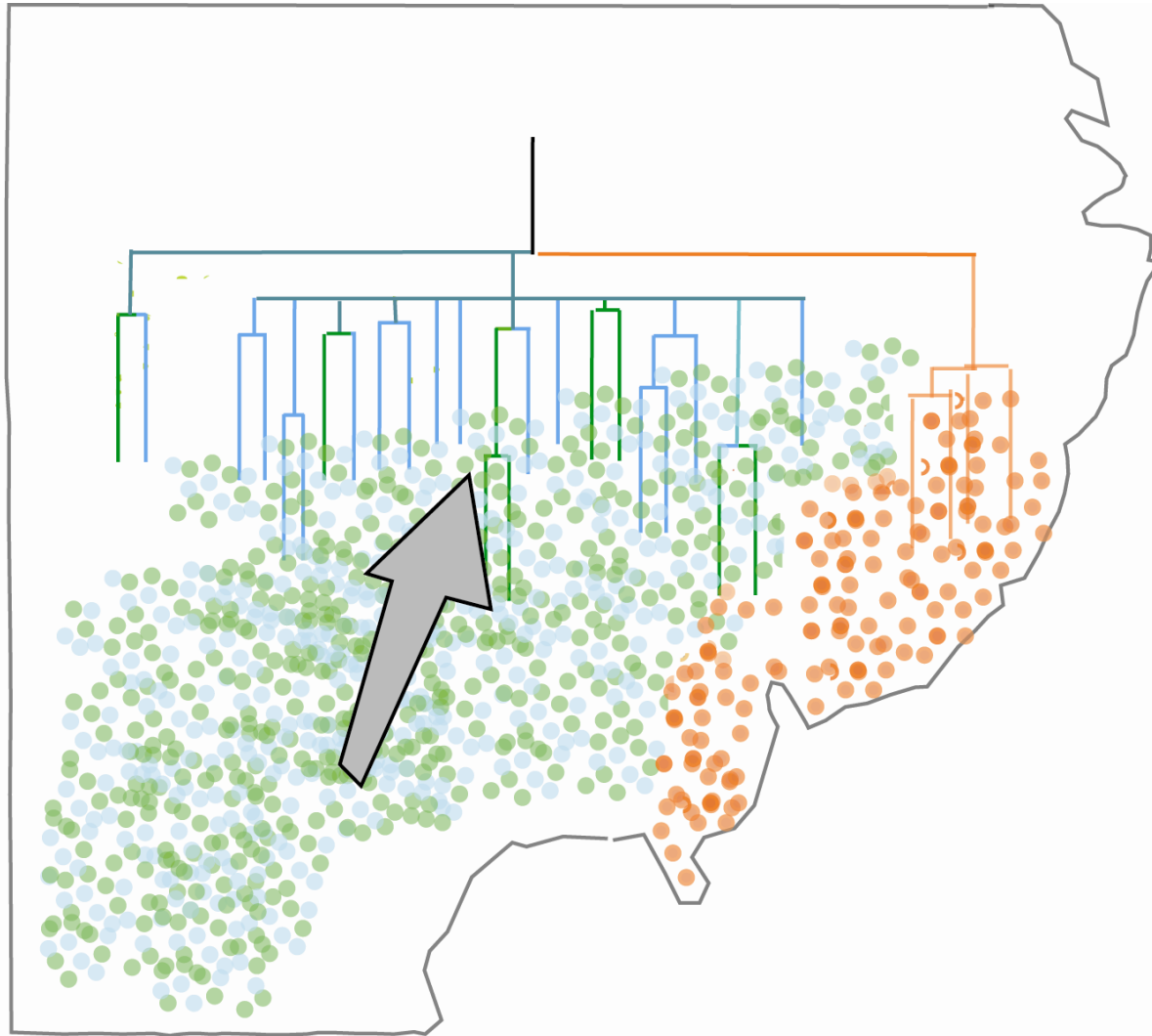
History of the nuclear genes studied



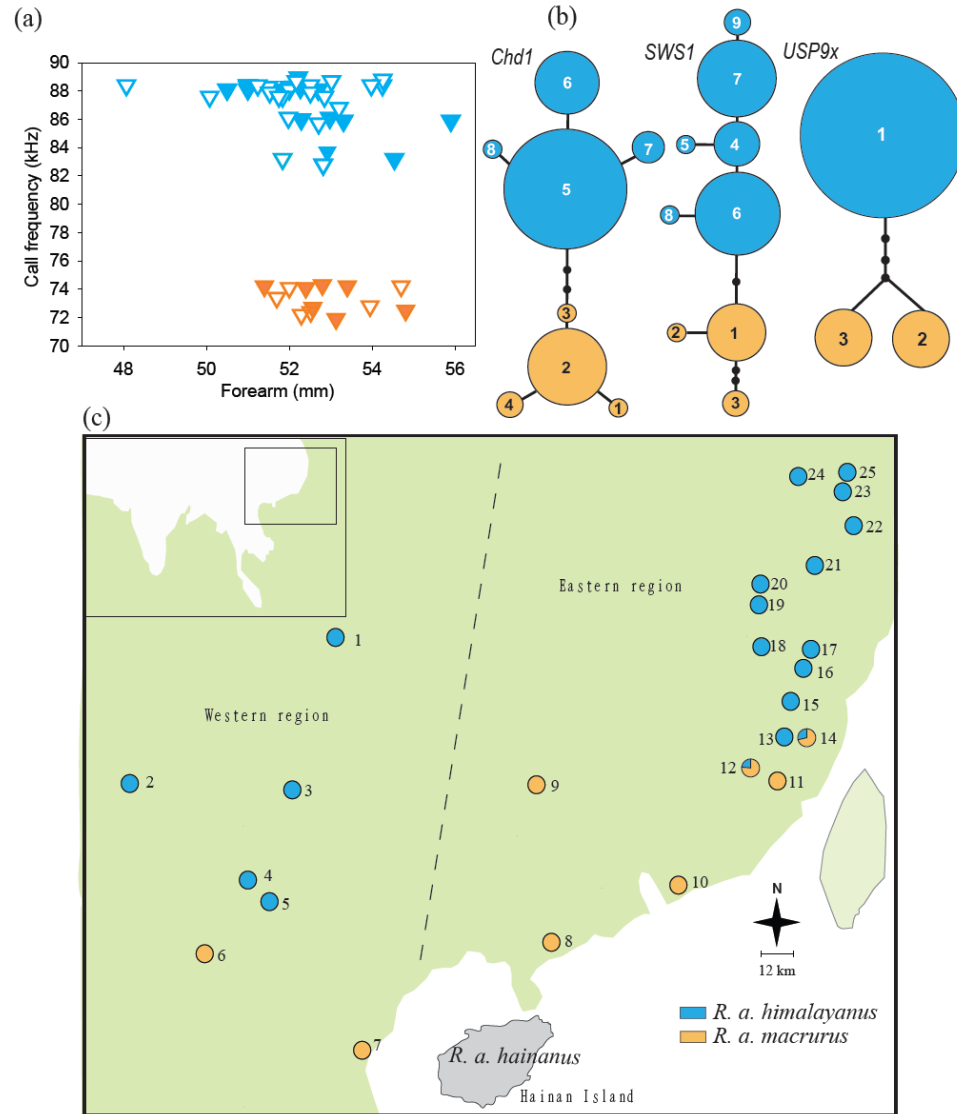
History of the nuclear genes studied



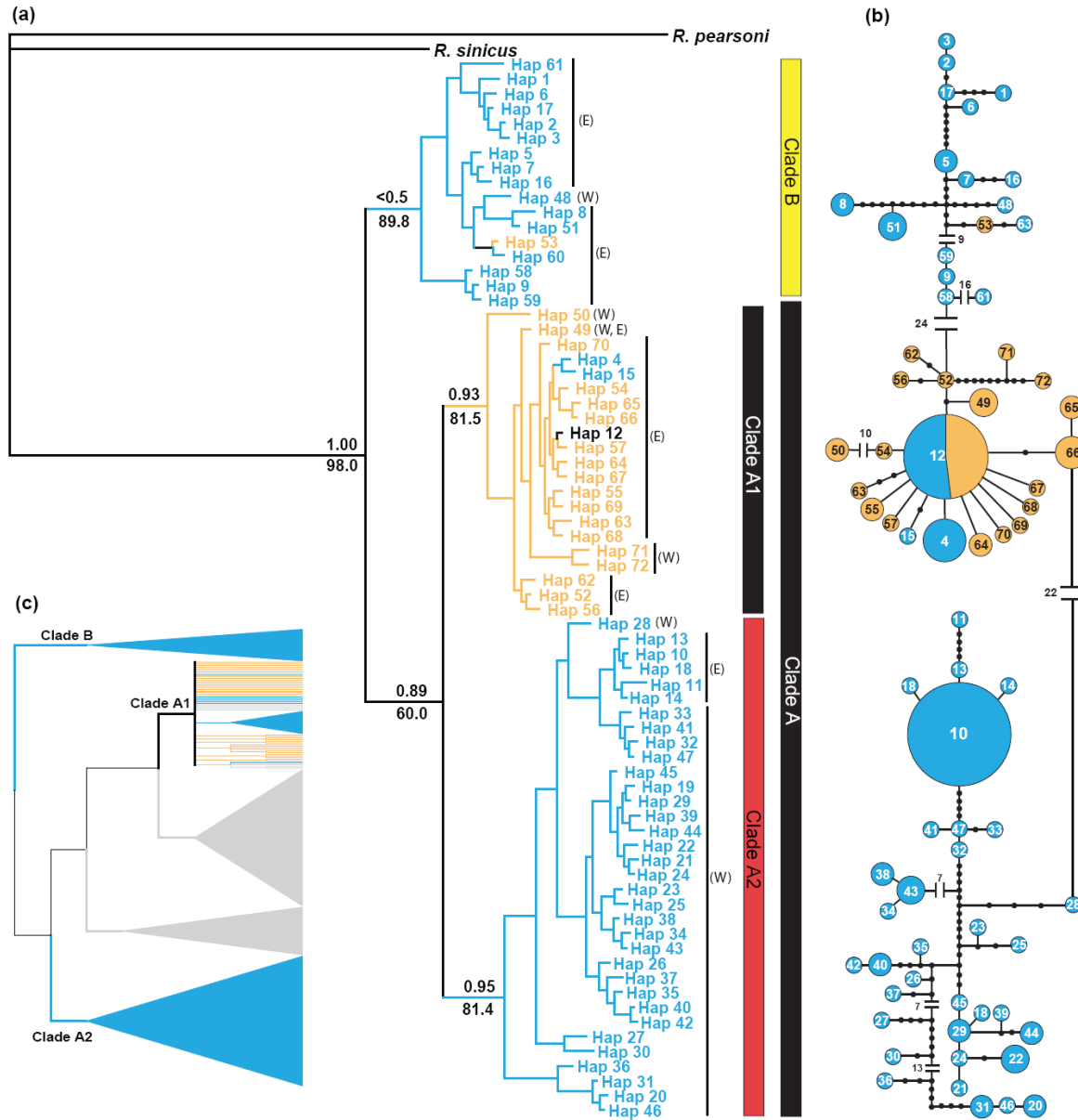
History of the nuclear genes studied



Example 2: *Rhinolophus affinis himalayanus* and *R. a. macrurus*



Example 2: *Rhinolophus affinis himalayanus* and *R. a. macrurus*



Detecting Introgression

Taxa must have a contact zone or have been in contact in the past

Often a geographical pattern

More commonly detected in mtDNA (barcoding caveat)

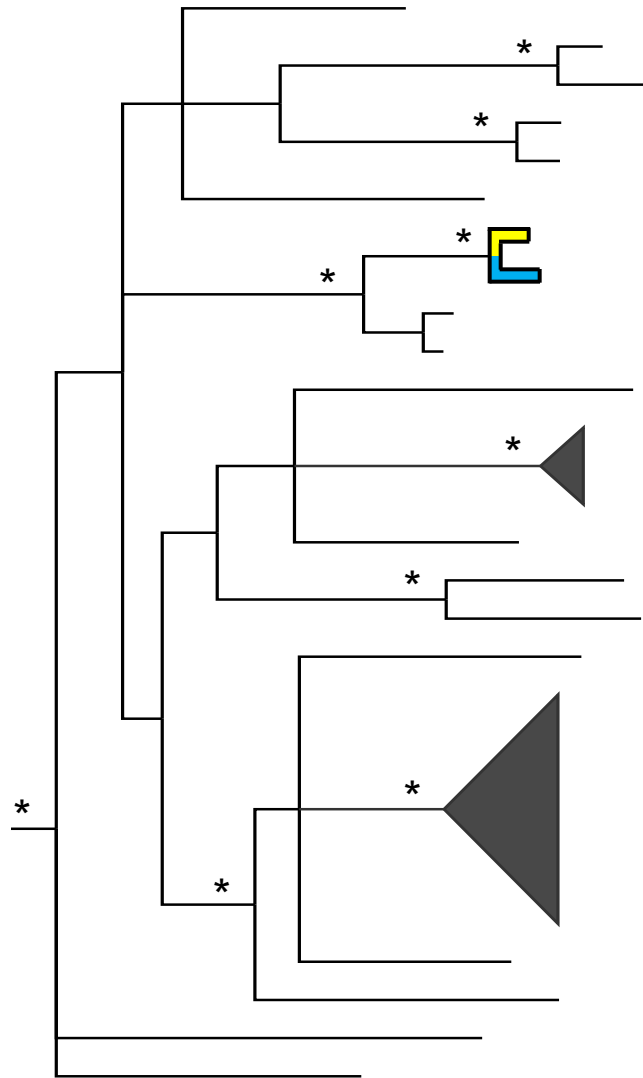
More common where one taxon has undergone population expansion

Neutral genes typically flow from resident taxon to the invading taxon

Why do phylogenetic trees sometimes disagree with other datasets?

1. Incomplete sorting
2. Long branch attraction
3. Introgression
4. Homoplasy

Bayesian tree of *Murina* based on mtDNA COI (637 bps)



M. gracilis



M. recondita

* posterior probability > 0.95



Murina gracilis (○)

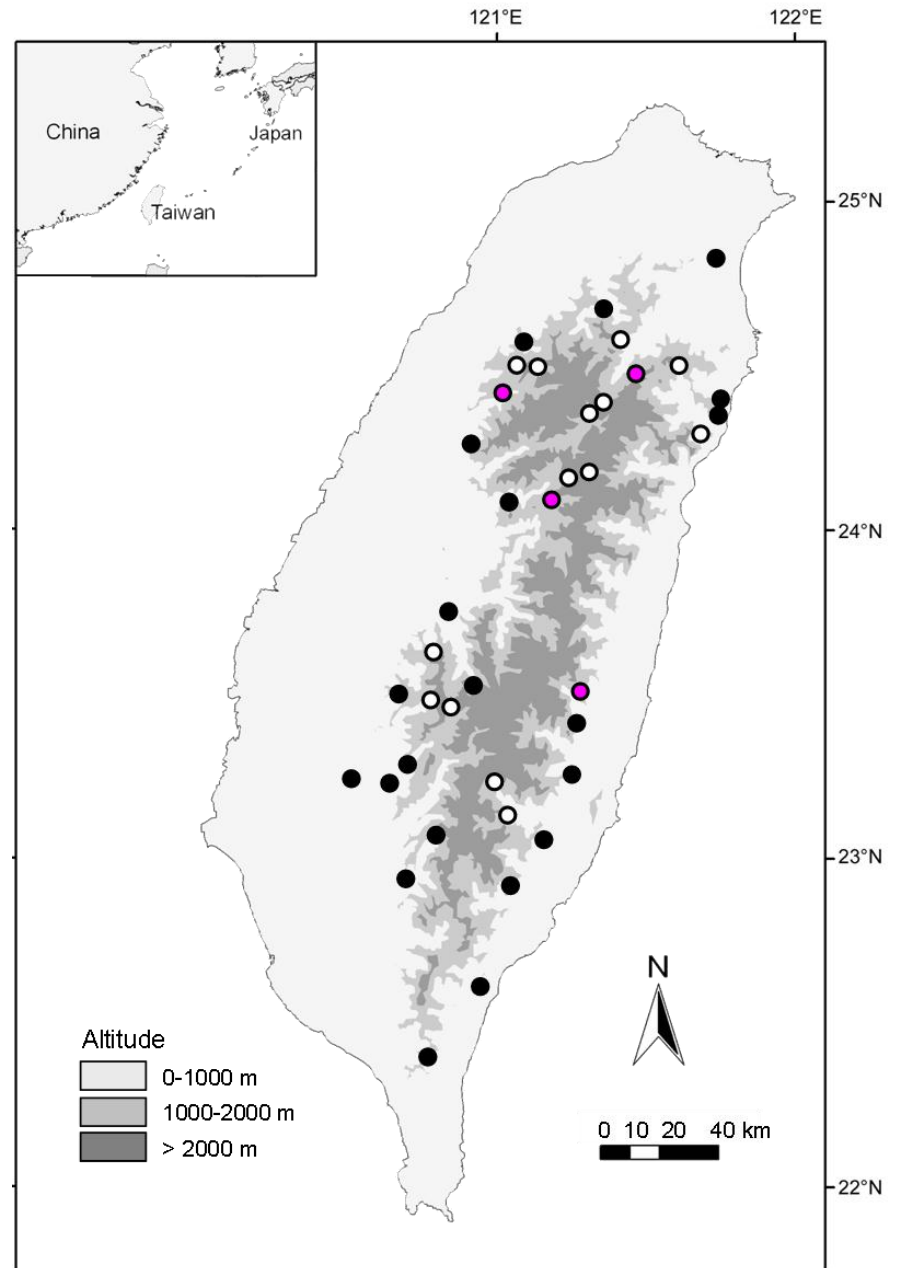
≥ 1500 m ASL



M. recondita (●)

≤ 1500 m ASL

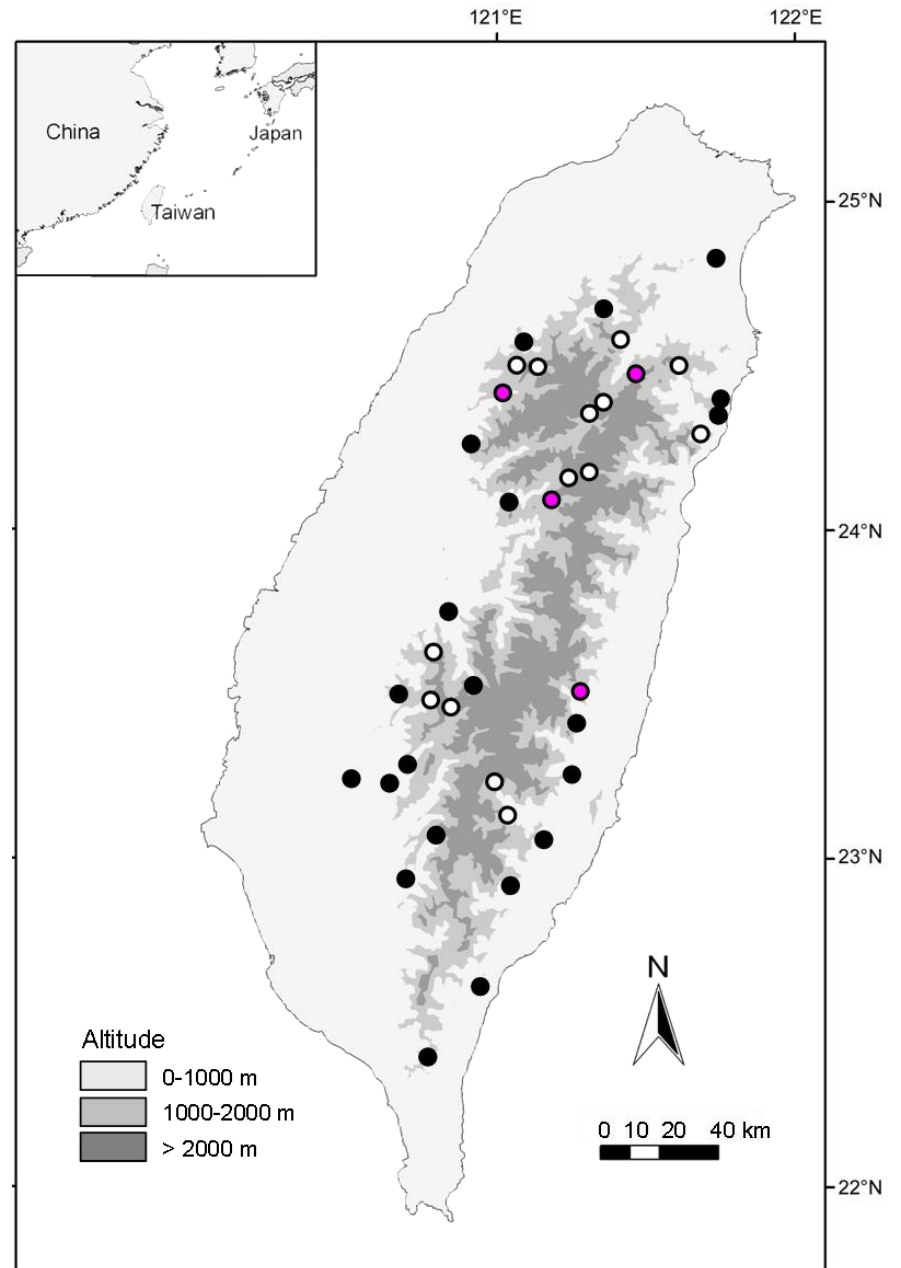
Both species (●)



Input: 106 *M. gracilis*

144 *M. recondita*

14 microsatellite loci



A – 8 repeats

Forward primer

... GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTTCGCACTTTAACGATACGG...
... CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGAAGAAGCGTGAAAATTGCTATGCC...

Reverse primer

B – 7 repeats

Forward primer

... GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTTCGCACTTTAACGATACGG...
... CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGCGTGAAAATTGCTATGCC...

Reverse primer

C – 9 repeats

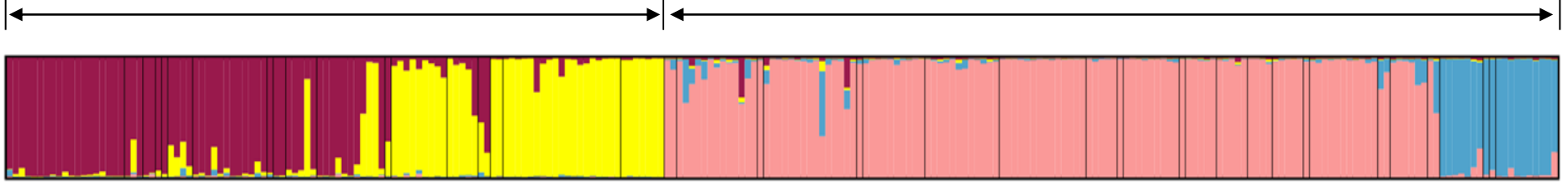
Forward primer

... GCTCCAGGCTTAGACTTCTTCTTCTTCTTCTTCTTCTTCTTCTTTCGCACTTTAACGATACGG...
... CGAGGTCCGAATCTGAAGAAGAAGAAGAAGAAGAAGAAGCGTGAAAATTGCTATGCC...

Reverse primer

M. gracilis

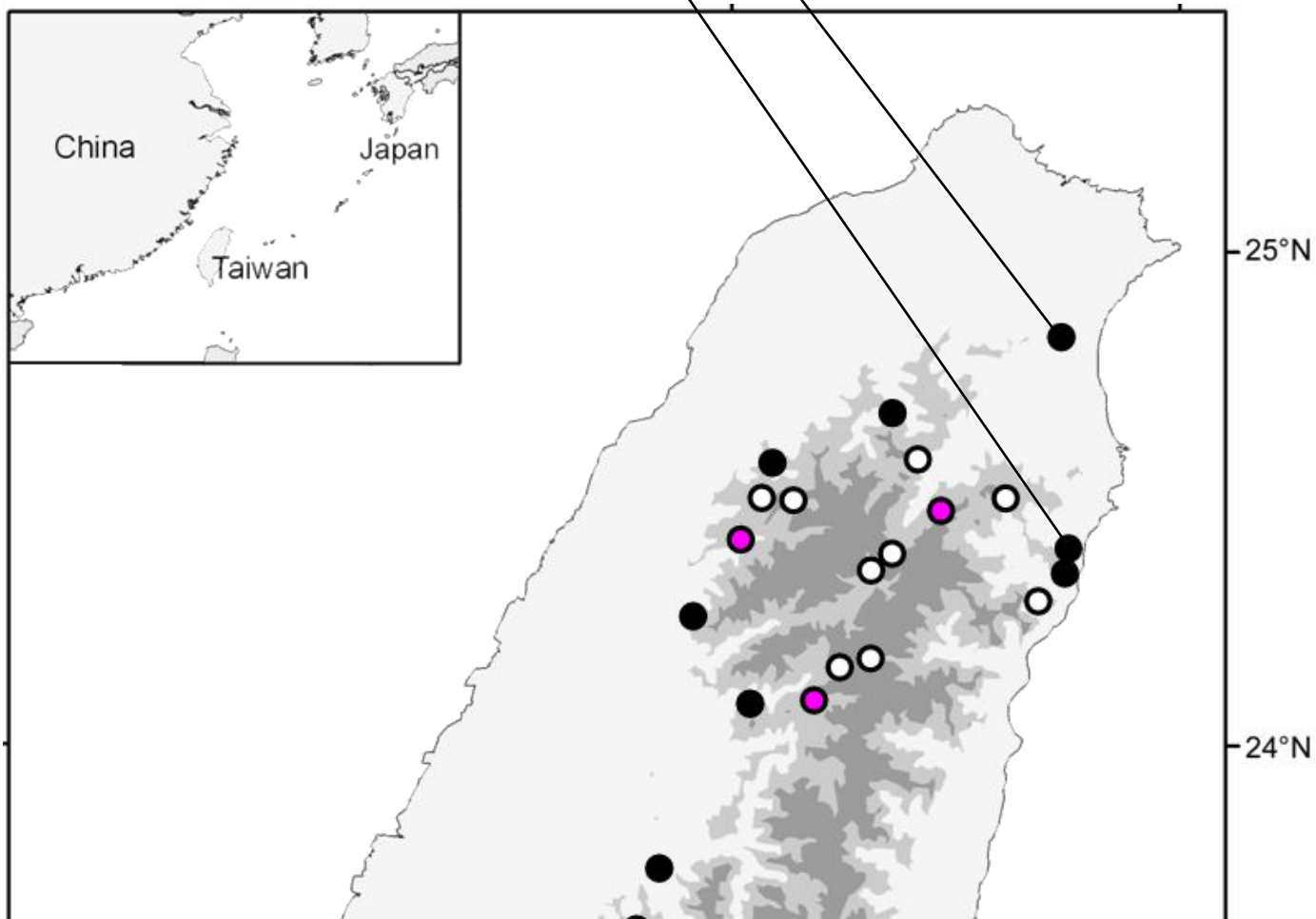
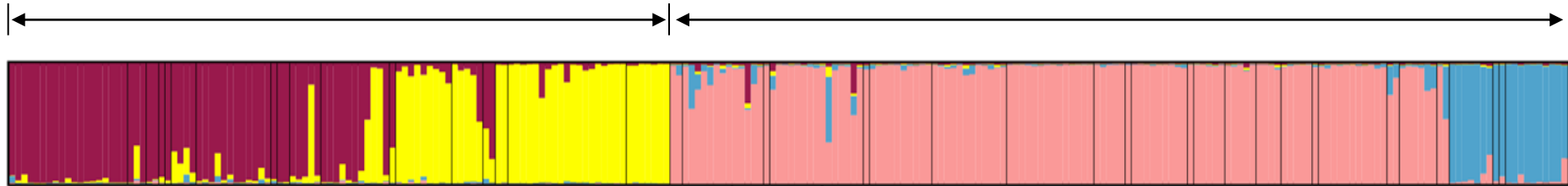
M. recondita



K = 4

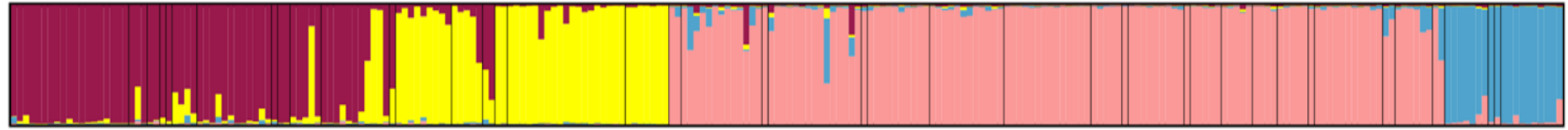
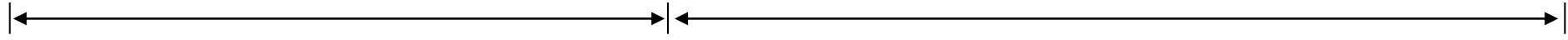
M. gracilis

M. recondita

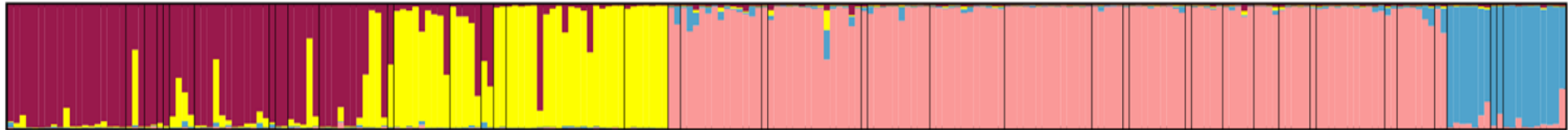


M. gracilis

M. recondita



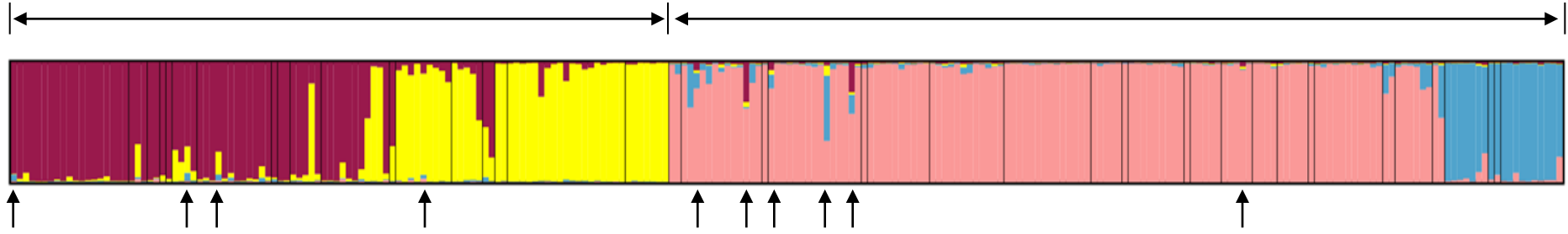
14 loci



13 loci: without locus "A9"

M. gracilis

M. recondita

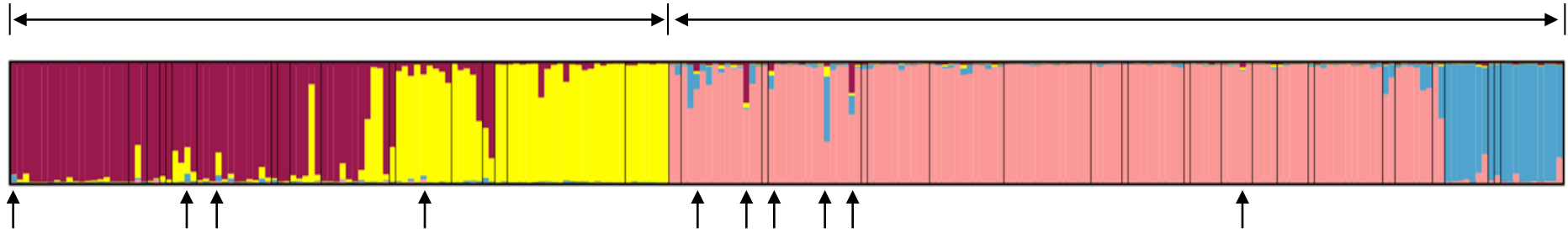


Recent hybridization? Or homoplasy?



M. gracilis

M. recondita



Recent hybridization? Or homoplasy?

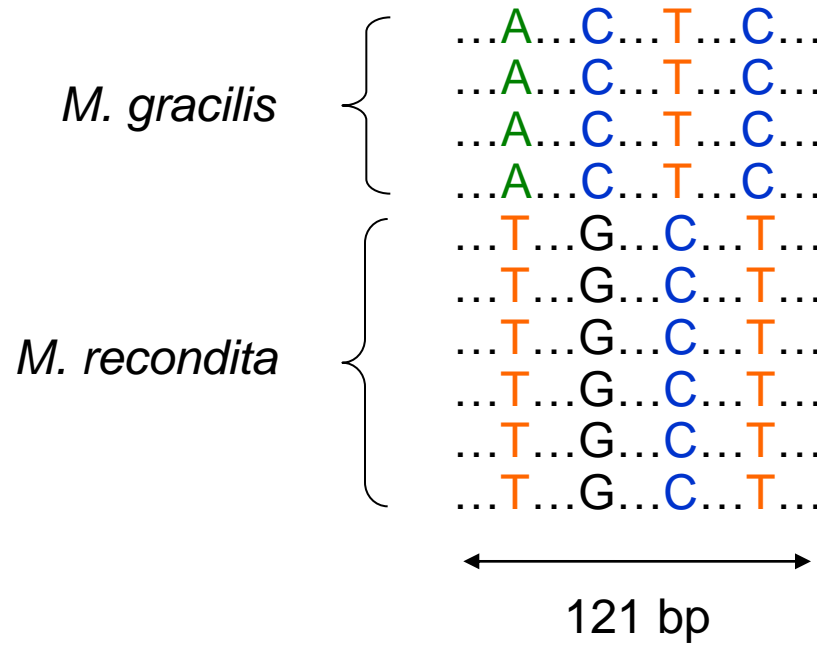
Sequencing the flanking regions of loci showing mixed ancestry for some individuals (↑):

Predictions:

If hybridization: some *M. gracilis* will phylogenetically group with *M. recondita*, and vice versa.

If homoplasy: samples of different species will be phylogenetically separate

Flanker sequencing result for locus "A9"



Flanker sequencing result for locus “A9”

M. gracilis {
...A...C...T...C...
...A...C...T...C...
...A...C...T...C...
...A...C...T...C...
M. recondita {
...T...G...C...T...
...T...G...C...T...
...T...G...C...T...
...T...G...C...T...
...T...G...C...T...
...T...G...C...T...

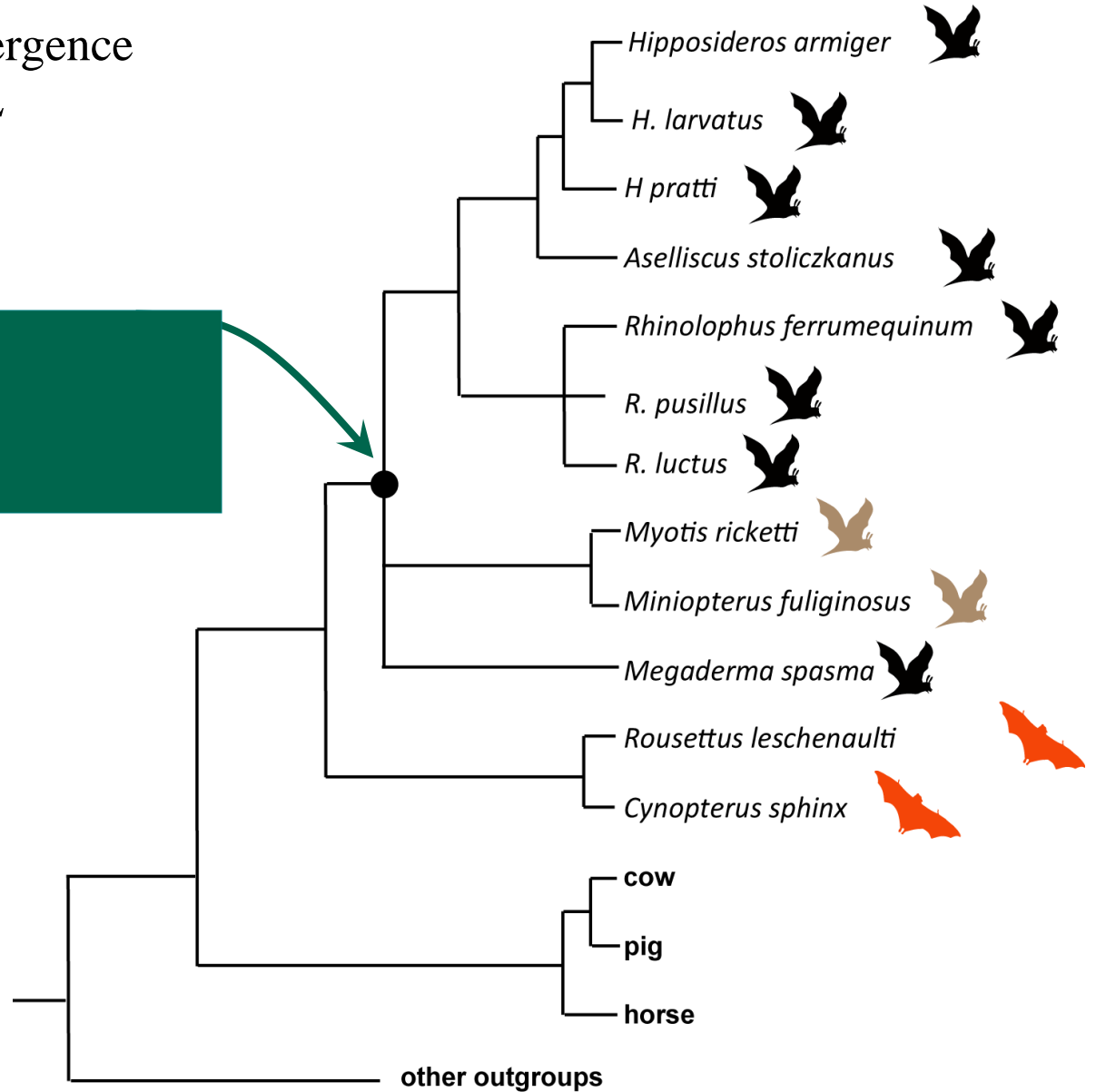
The sequencing result for another locus “A122” (209 bp) is also consistent with the prediction of the allele size homoplasy

Why do phylogenetic trees sometimes disagree with other datasets?

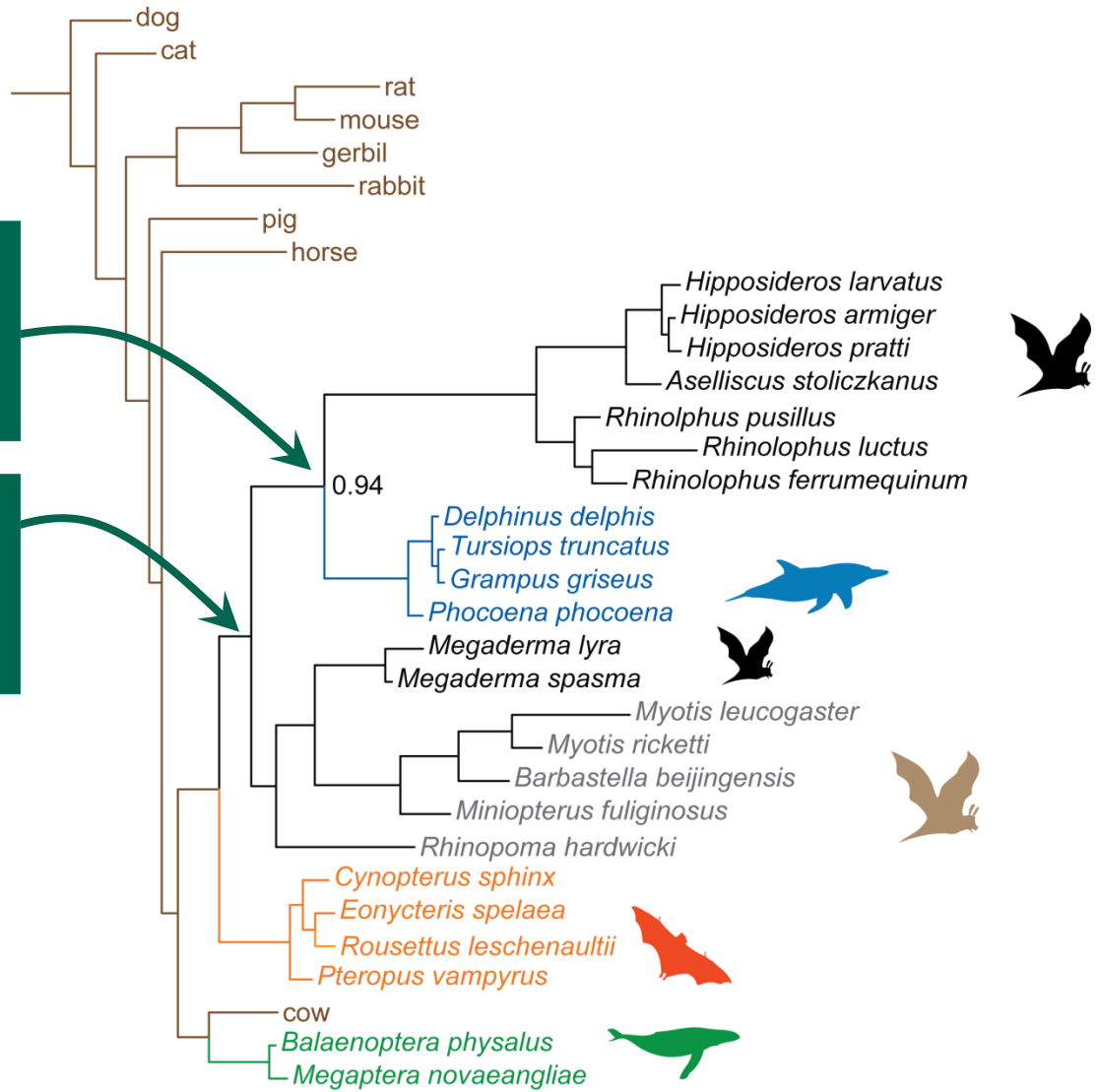
1. Incomplete sorting
2. Long branch attraction
3. Introgression
4. Homoplasy
5. Adaptive convergence

Example of adaptive convergence
Prestin gene tree using ML

Echolocating bats form
monophyletic clade
(node BPP > 0.9)



Prestin gene tree using ML

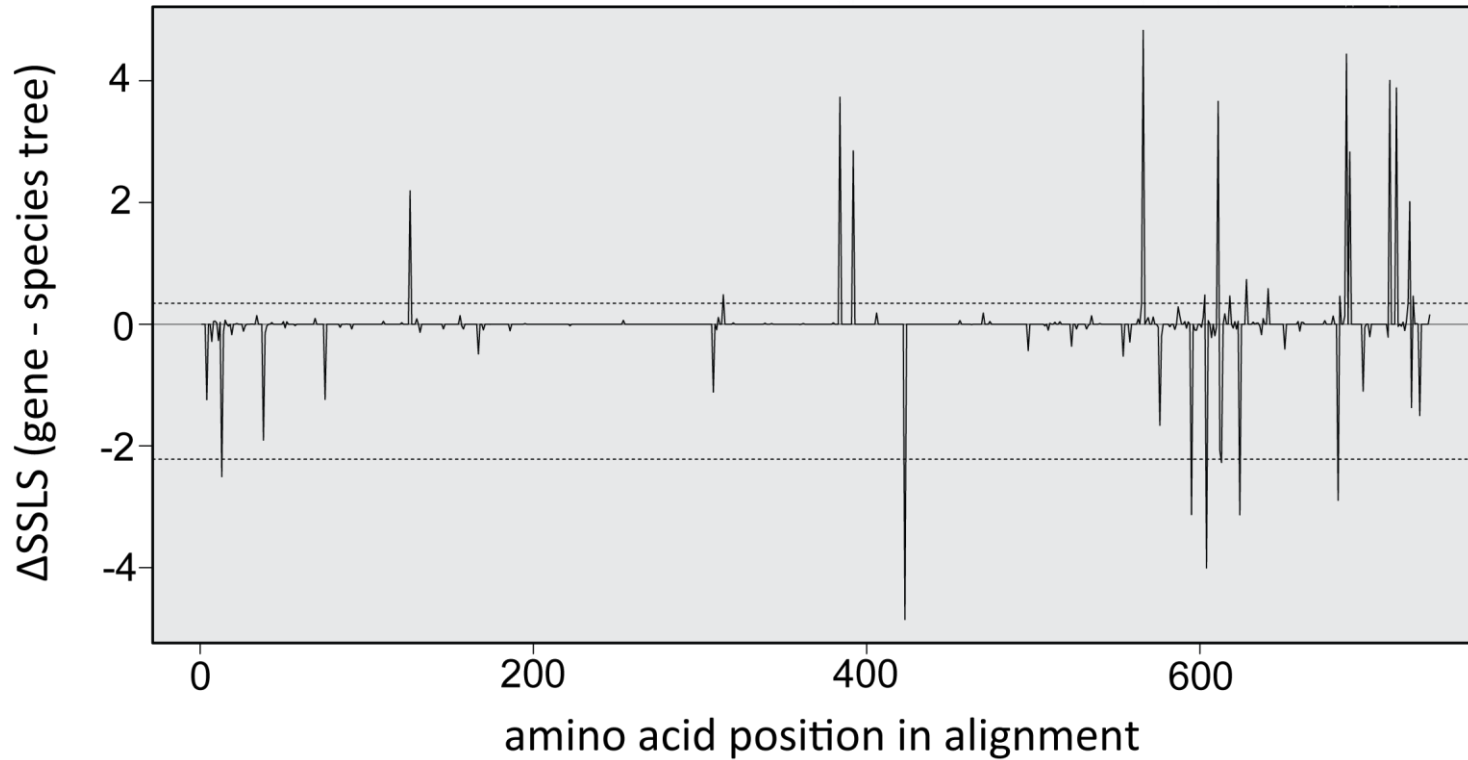


Dolphins & horseshoe bats form monophyletic clade
BPP > 0.9

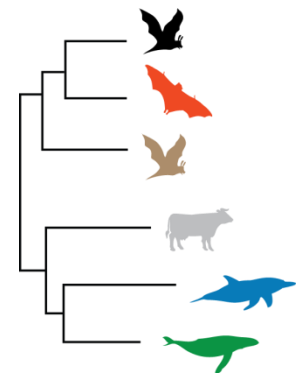
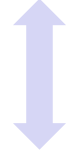
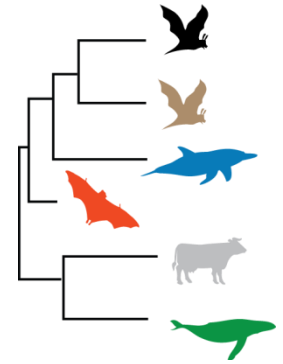
Echolocating bats & cetaceans form monophyletic clade
BPP > 0.65

0.02

Revealing site-wise convergence



Support for “wrong tree”)



Support for correct tree

Results from 1200 genes

