

The Evolution of Transposon Families

Beth Hellen

-
1. Introduction to Transposons
 2. Analysing Class II transposons as a population
 3. Analysing the GC environment of Alu elements

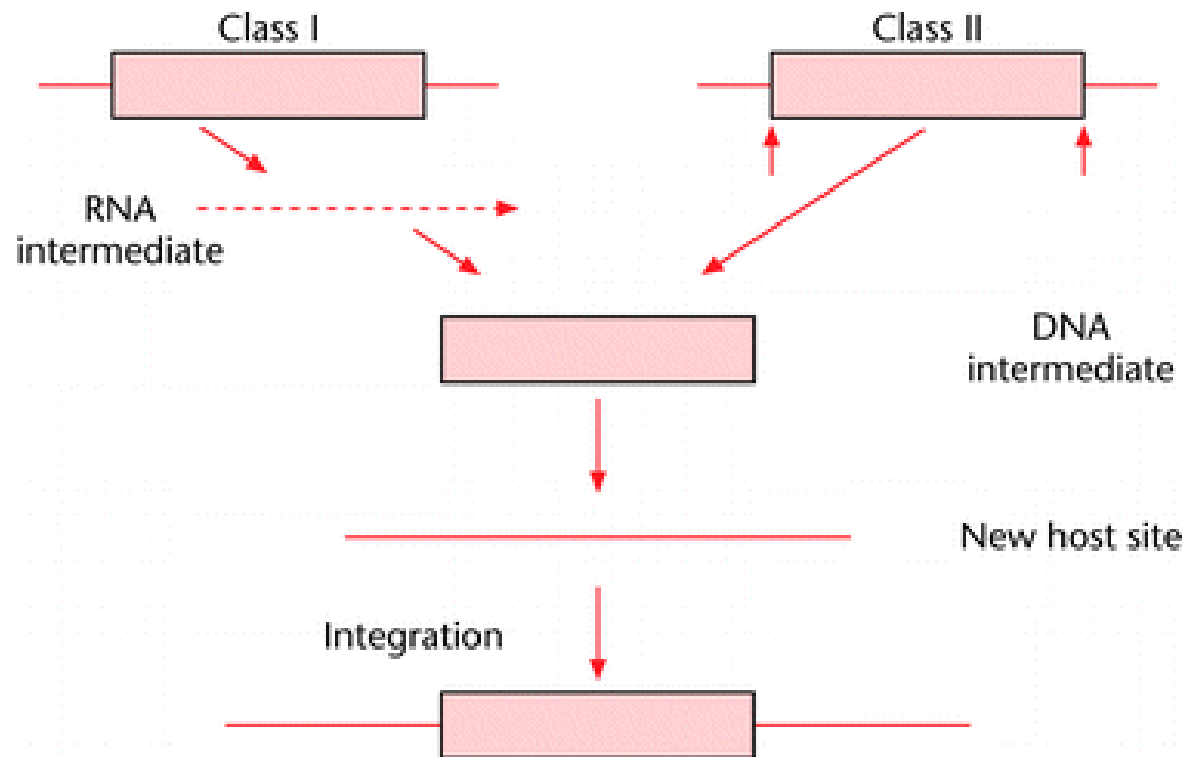


Discovery of Transposons

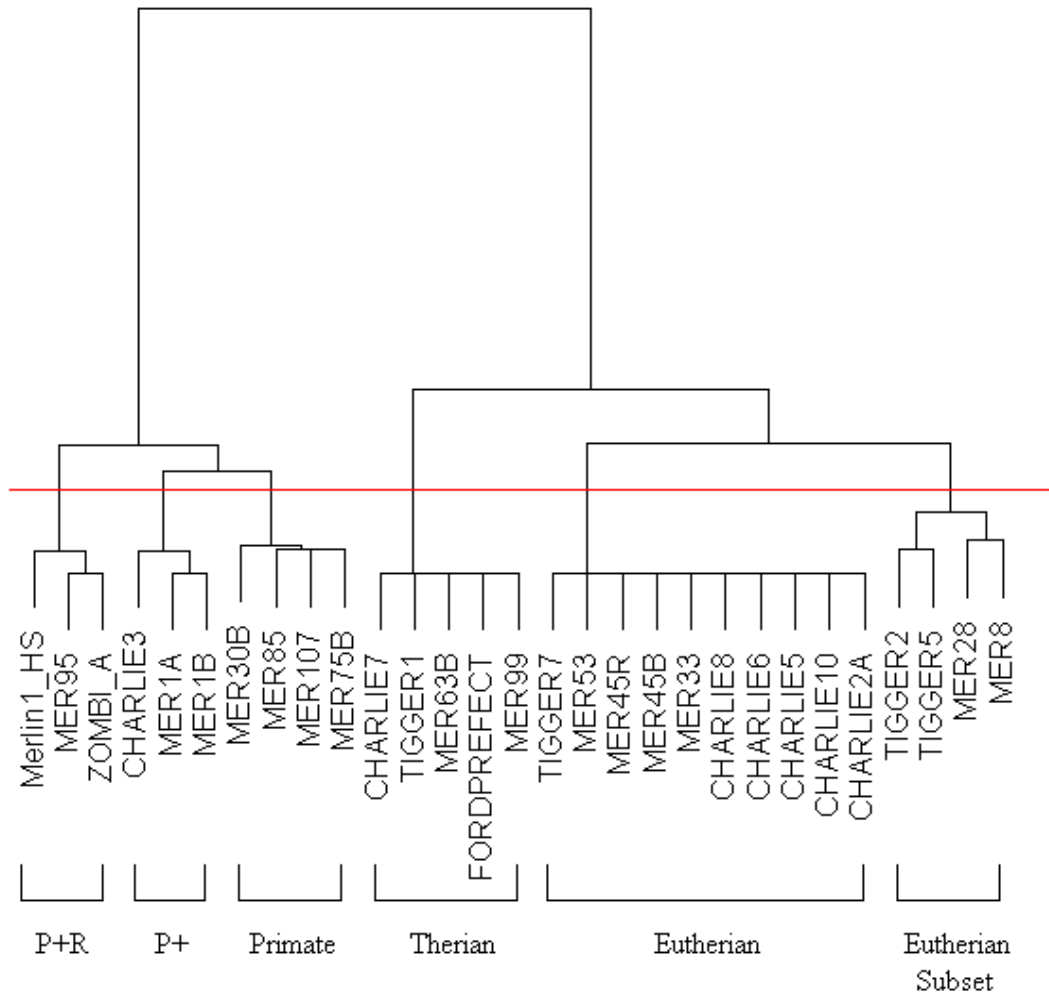
- ▶ Barbara McClintock – Nobel Prize 1983
- ▶ *Ac* and *Ds* elements effect colour of maize kernel
- ▶ Difficulty mapping elements - they changed their positions in different maize plants.
- ▶ *Ac* elements are 4563bp long.
- ▶ *Ds* elements are *Ac* elements that have undergone deletions



Introduction to Transposons



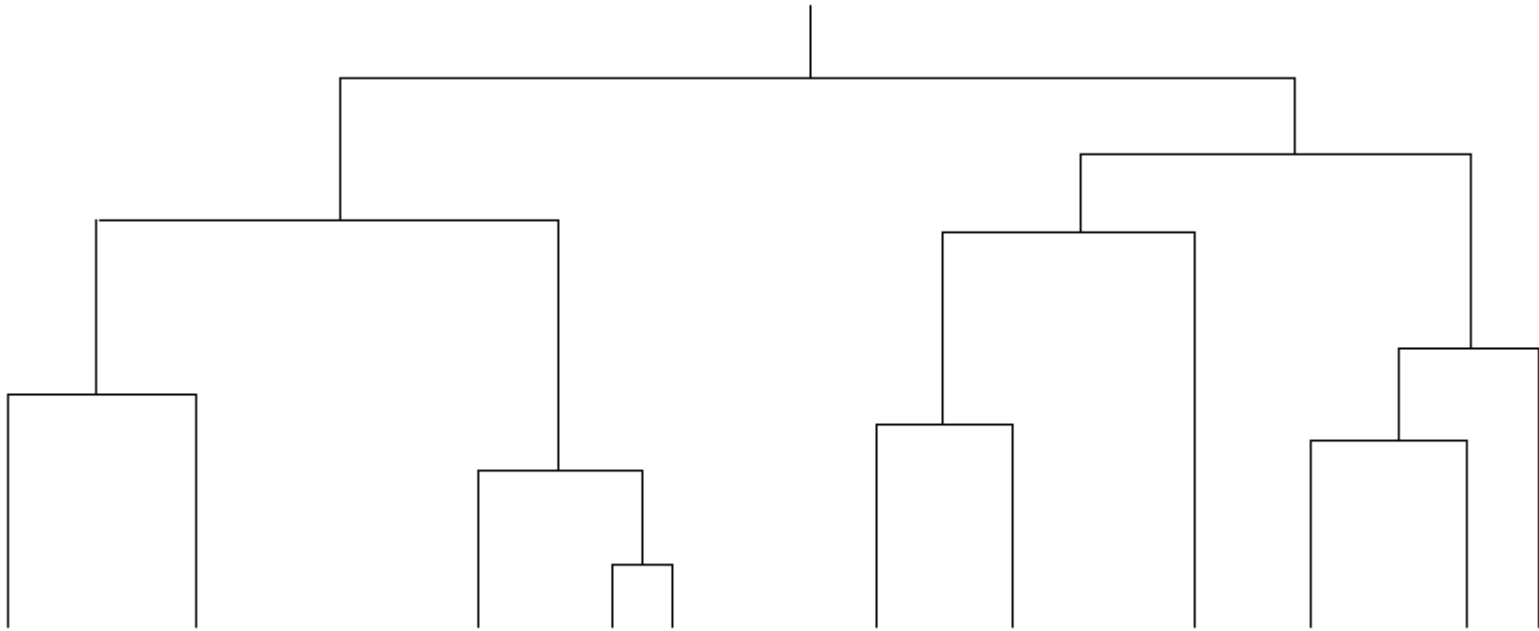
Families of Transposons



-
1. Introduction to Transposons
 2. **Analysing Class II transposons as a population**
 3. Analysing the GC environment of Alu elements



Transposons as a population

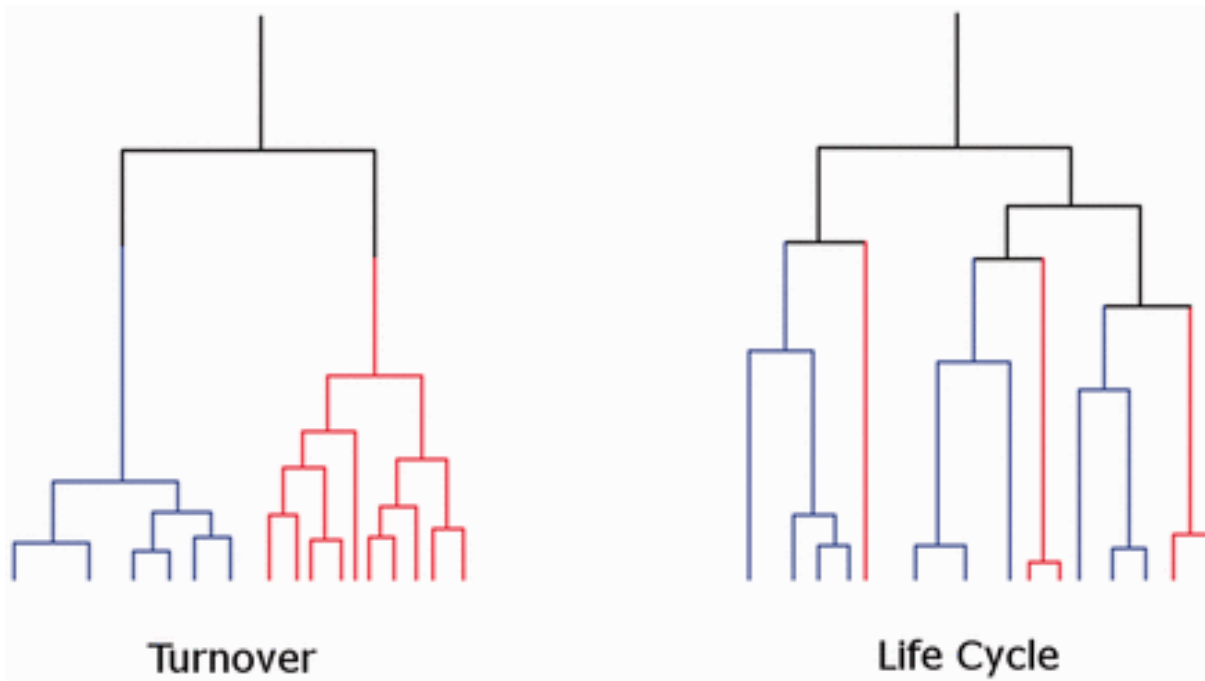


Birth, Proliferation & Death

- ▶ When did each family first appear?
- ▶ Are elements quickly inactivated and removed from the genome, current elements being relatively young? (Turnover)
- ▶ Or do they persist for 100s millions of years (Lifecycle)



Proliferation possibilities

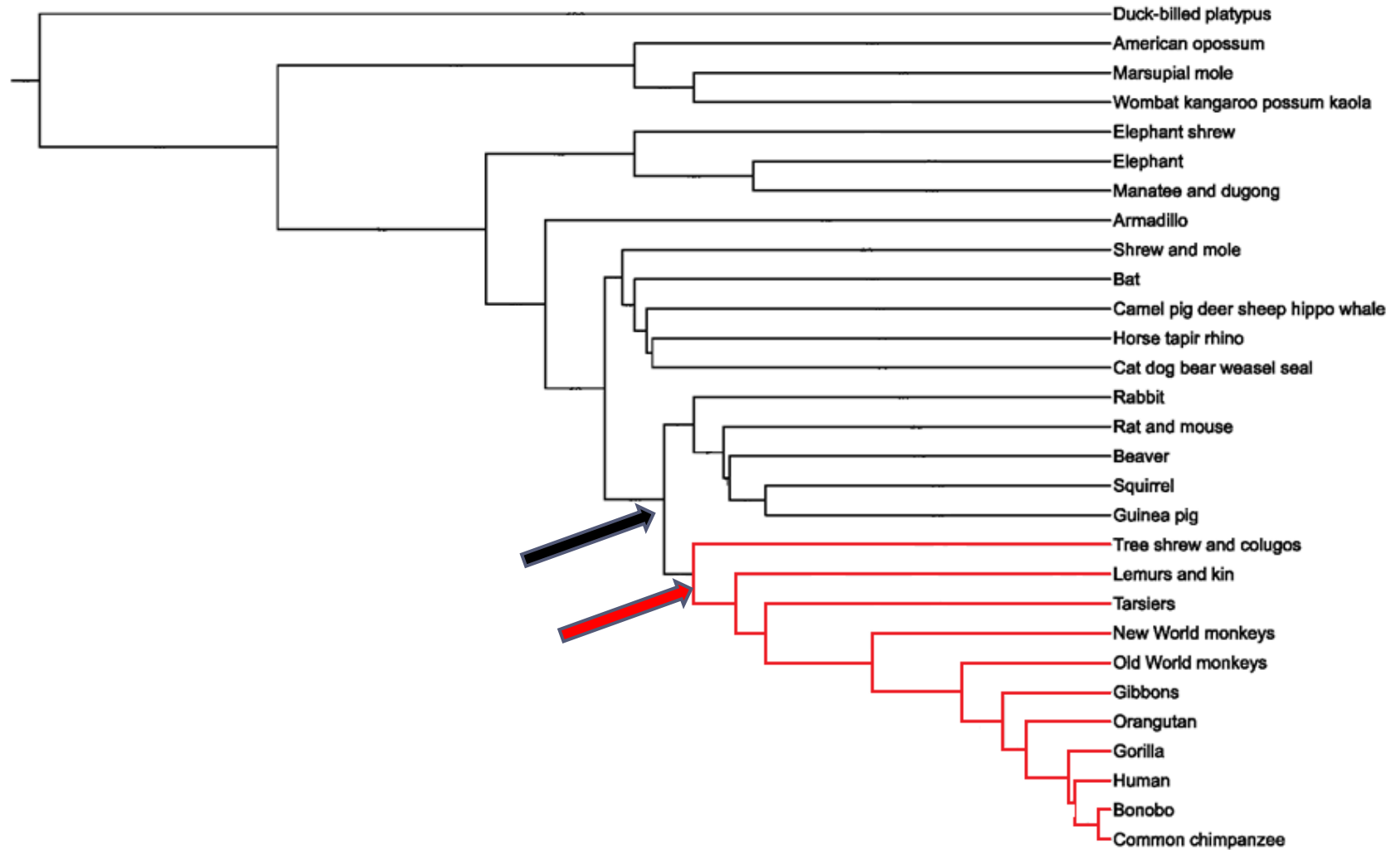


Birth, Proliferation & Death

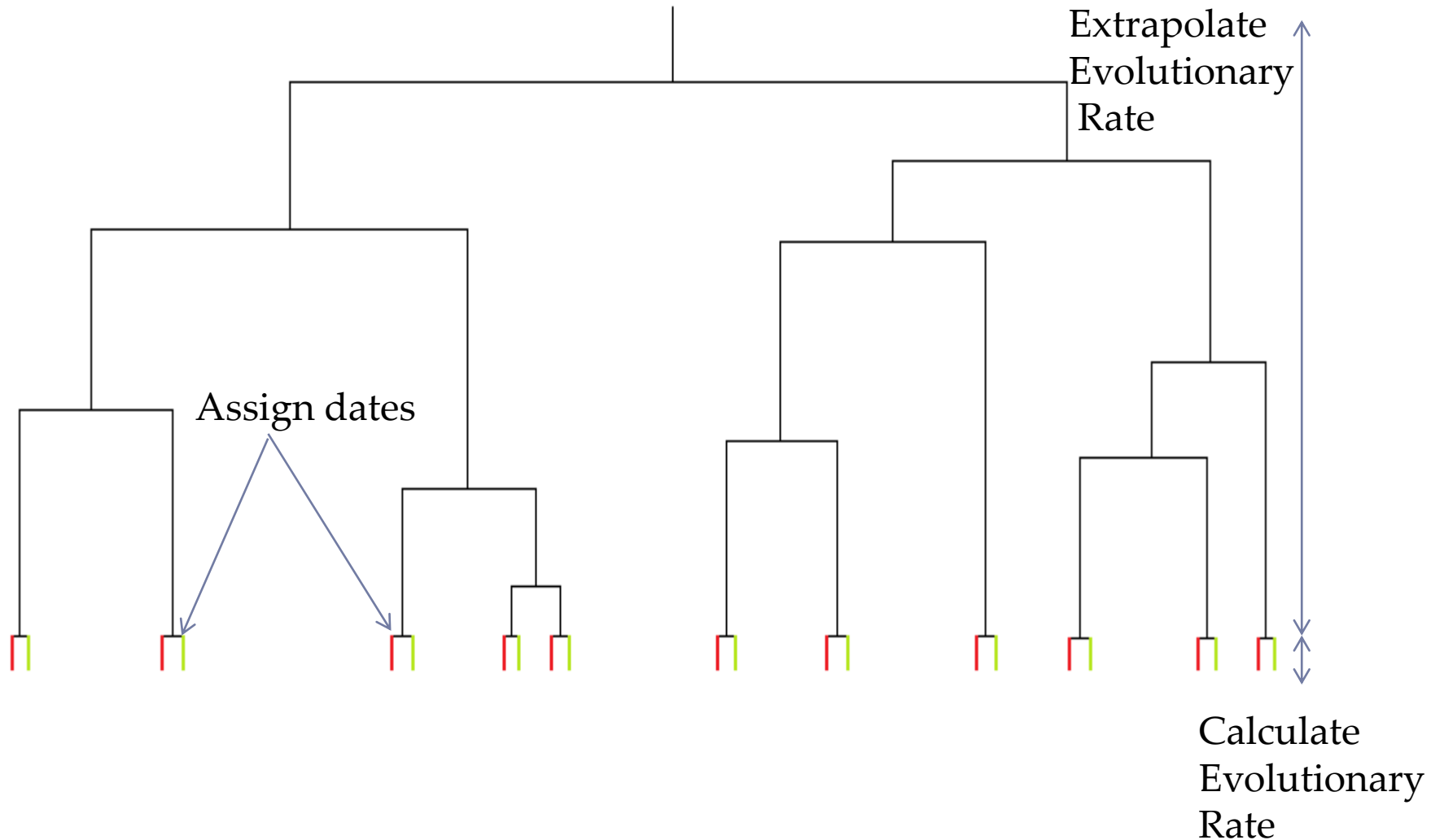
- ▶ When did each family first appear?
- ▶ Are elements quickly inactivated and removed from the genome, current elements being relatively young? (Turnover)
- ▶ Or do they persist for millions of years (Lifecycle)



Presence in Modern Organisms



Molecular Dating



Birth, Proliferation & Death

- ▶ When did each family first appear?
- ▶ Are elements quickly inactivated and removed from the genome, current elements having a recent common ancestor element? (Turnover)
- ▶ Or do they persist for 100s millions of years (Lifecycle)



Which Proliferation process?

Turnover

Phylogeny 'origin' is later than expected

Different lineages predict different origin dates

Life Cycle

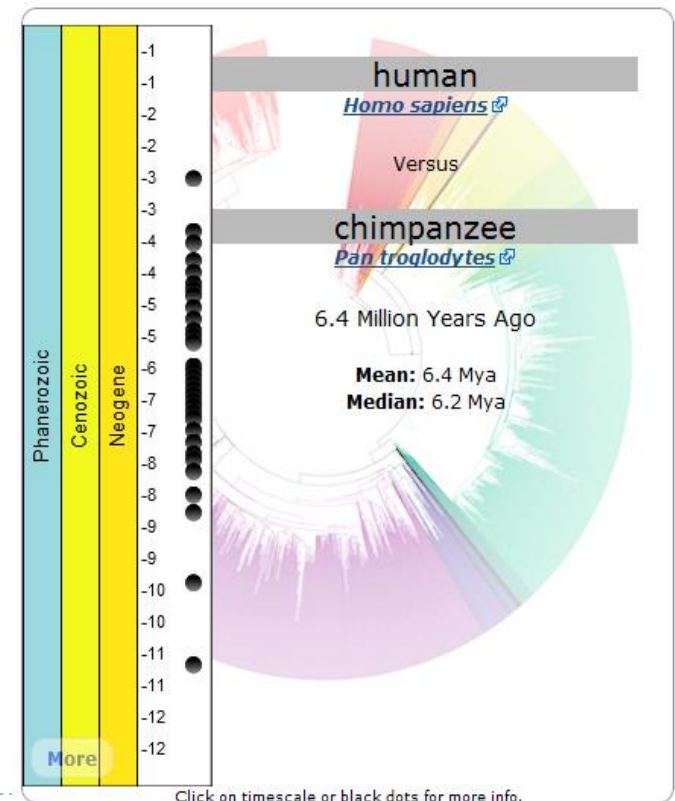
Phylogeny 'origin' is at similar date to expected

Different lineages predict similar origin dates



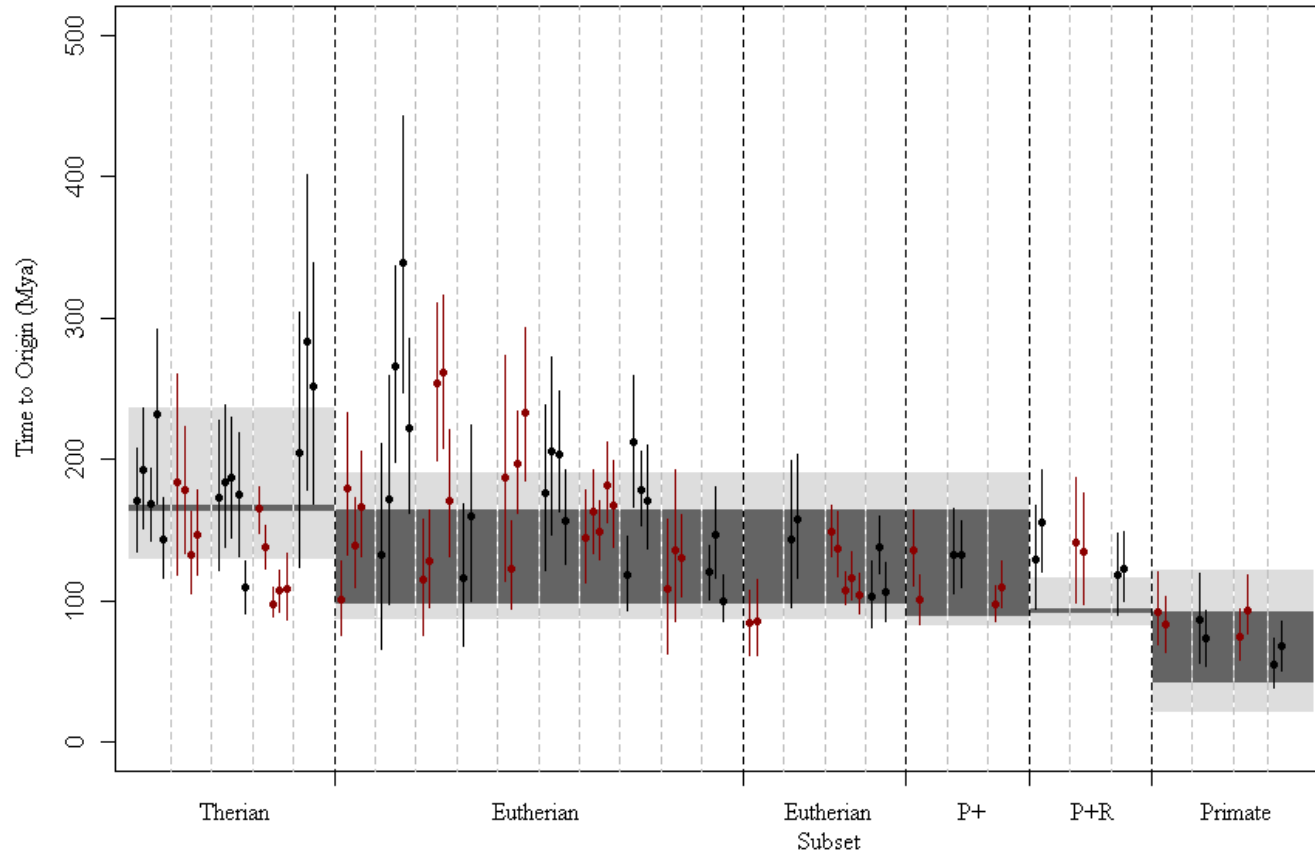
Species Comparisons

- ▶ Human – Chimp -> 6.4 My
- ▶ Human – Orangutan -> 15.7 My
- ▶ Dog – Panda -> 45 My
- ▶ Dog – Cat -> 55.7 My
- ▶ Cow – Pig -> 64.5 My



Age of Class II families Results

(Hellen & Brookfield 2013, Molecular Biology & Evolution)



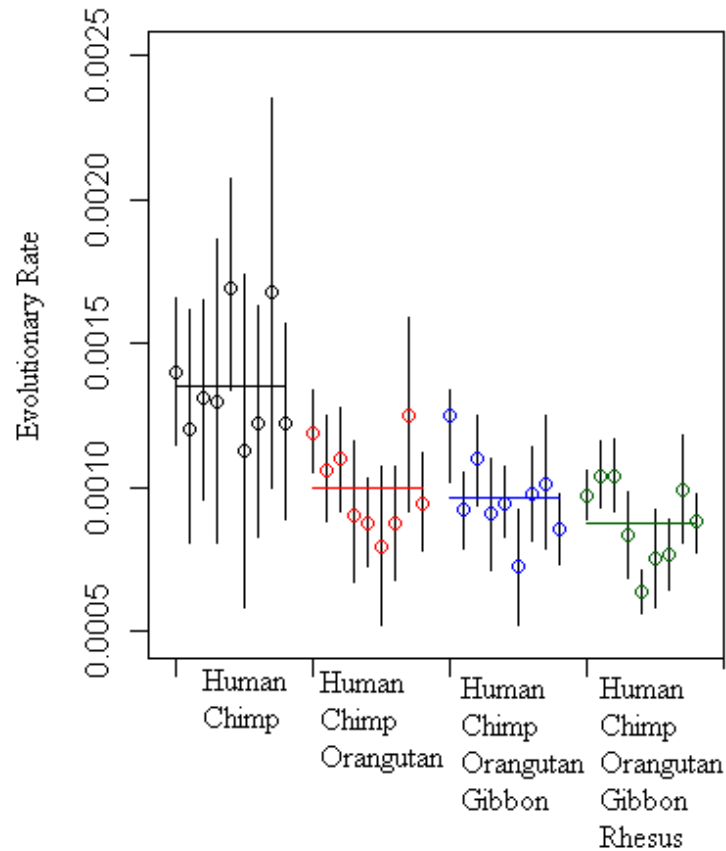
Life Cycle Proliferation Process

- ▶ Most Analyses traced back to at least expected time
- ▶ Primate, Carnivore & Artodactyl analyses agree
- ▶ Issues with large error on origin predictions – due to large error on evolutionary rate
- ▶ Questions about why some prediction are earlier than expected

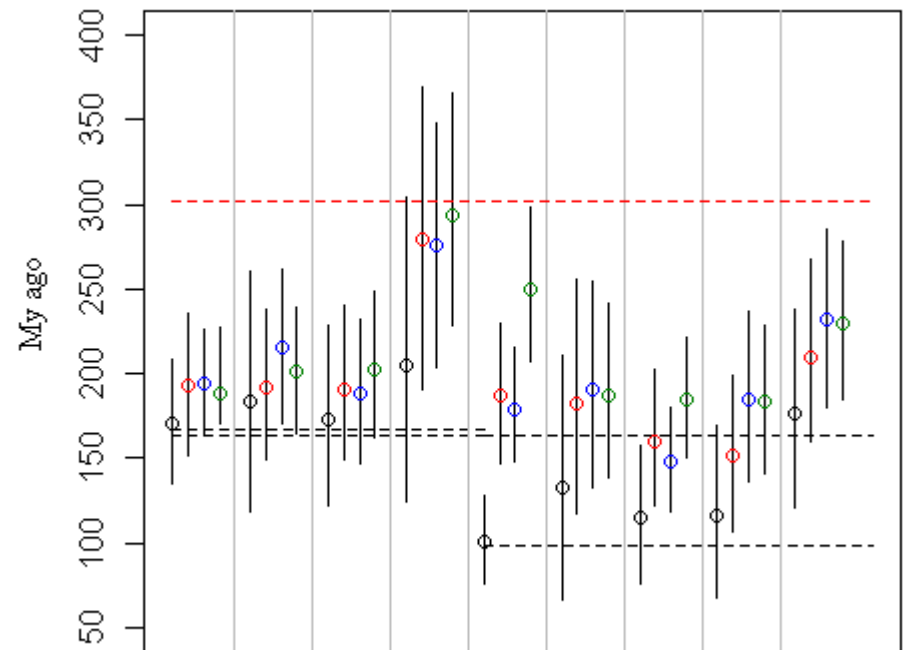


Using Multiple Homologs

Evolutionary Rate



Origin Date



-
1. Introduction to Transposons
 2. Analysing Class II transposons as a population
 3. **Analysing the GC environment of Alu elements**



Alus

- ▶ Human Genome sequencing paper contained section on Transposable elements inc. Alus (Lander *et al.*, 2001)
- ▶ Young Alu families (Human specific) were found to be underrepresented in GC rich regions compared to older families
- ▶ GC-rich areas are also likely to be gene-rich areas



IHGSC Hypotheses

1. There is a higher rate of random loss of Alus in GC-poor regions
2. Negative selection is acting against Alu elements in GC-poor areas
 - ▶ Unlikely as LINE1 are shown to accumulate in GC poor areas
3. Positive selection is acting in favour of Alu elements in GC-rich areas
 - ▶ Assume that Alus confer Darwinian fitness

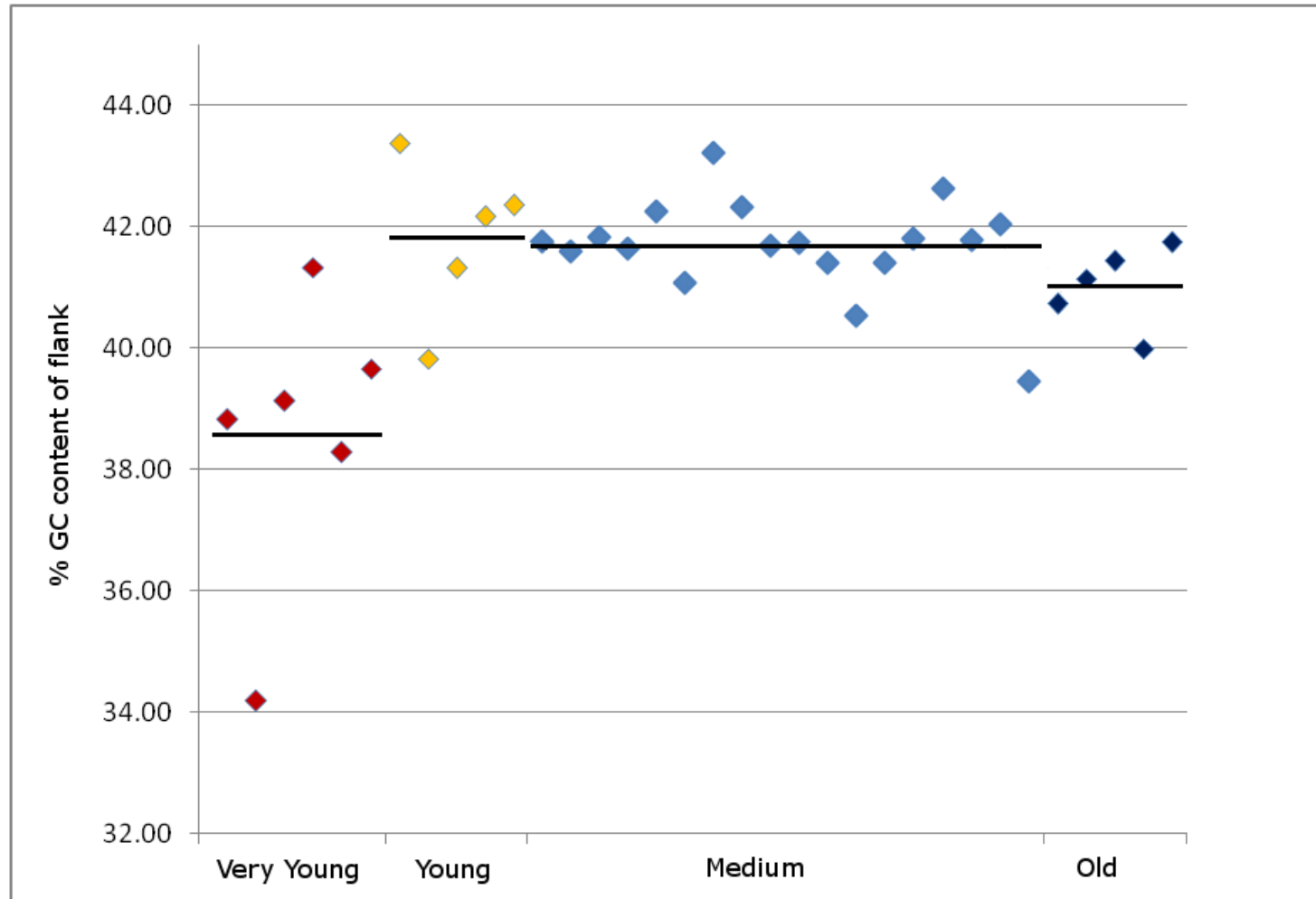


Response to findings

- ▶ Once elements have been in the genome for five million years they will be fixed in the population, and natural selection cannot, in principle, increase their abundance. - Brookfield 2001
- ▶ Alu elements are not preferentially degraded in GC-poor regions – Belle et al. 2005
- ▶ Various other studies attempting to explain GC discrepancy.



Young Alu families in GC-poor regions



Analysis by dating each element

- ▶ Previous analyses rely on dating elements by subfamilies
- ▶ BUT subfamilies are around for Millions of years.
- ▶ Old element from 'young subfamily' may be older than Young element from 'Old family'
- ▶ Date each element by presence in other Primate species

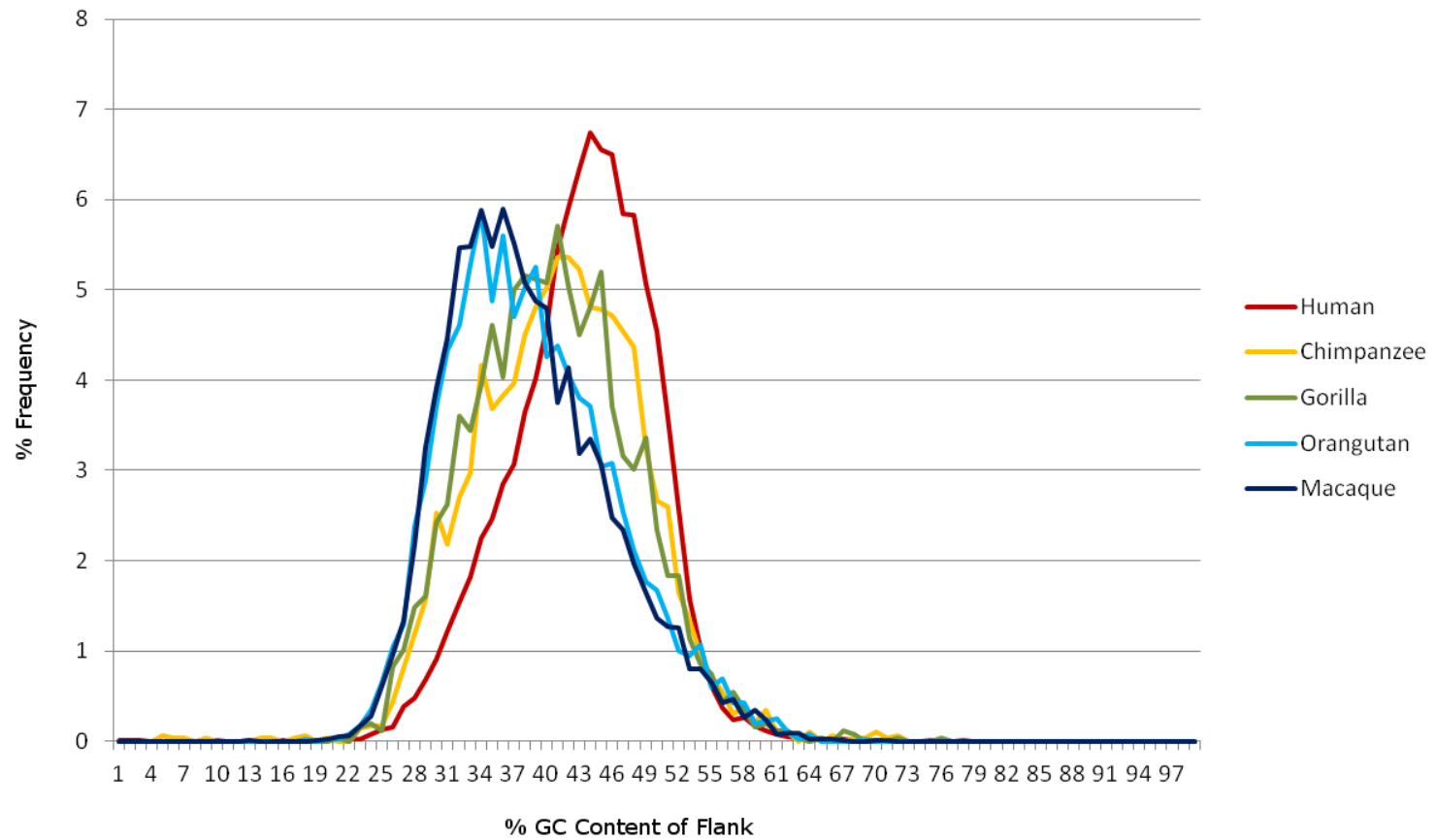


Methods

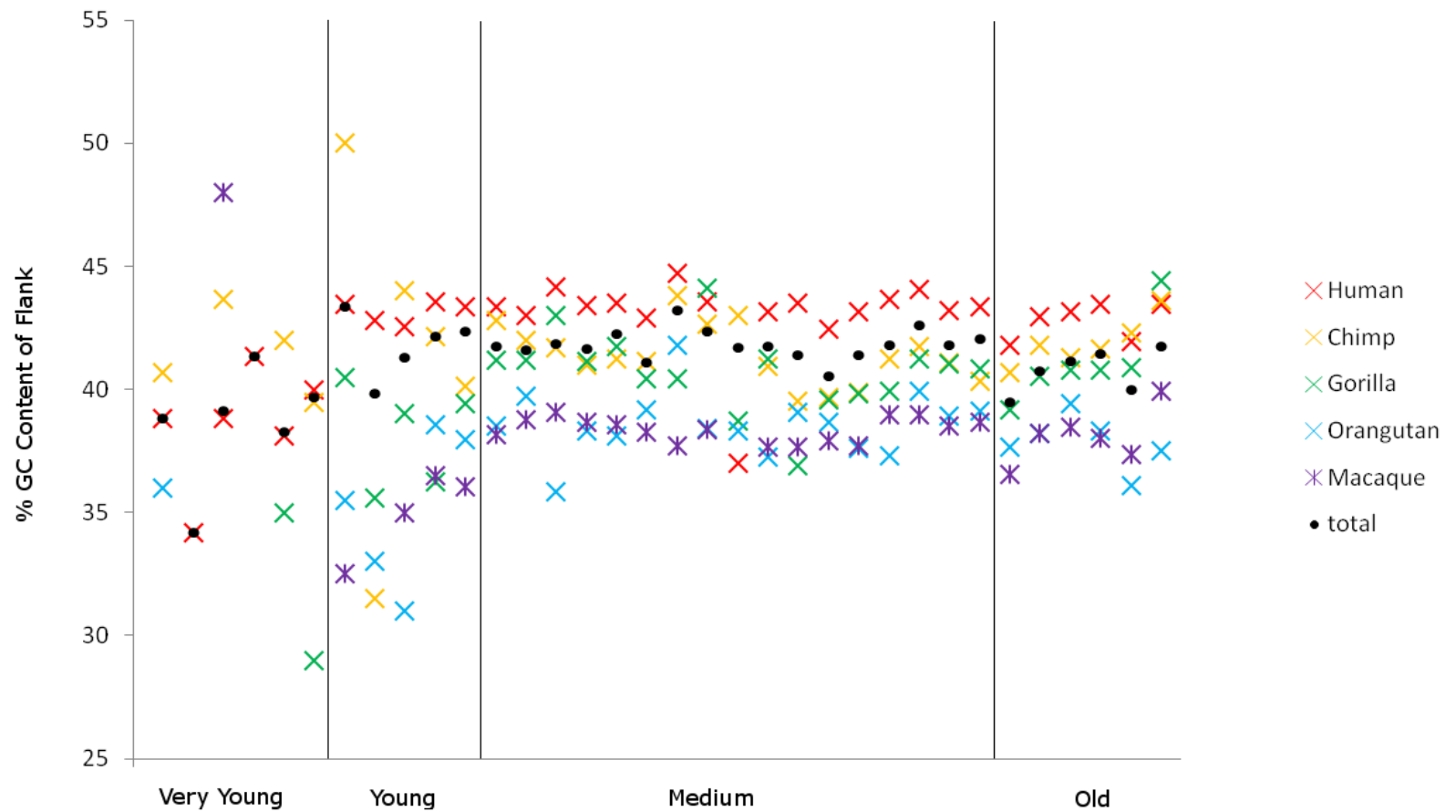
- ▶ 290,715 elements downloaded from UCSC (5 chromosomes)
- ▶ Reciprocal BLAST used to determine homologs in other Primates
 - Chimpanzee (CGSC 2.1.3/panTro3)
 - Gorilla (gorGor3.1/gorGor3)
 - Orangutan (WUGSC 2.0.2/ponAbe2)
 - Macaque (MGSC Merged 1.0/rheMac2)
- ▶ Only elements with clear evolutionary history (no missing elements in expected species) included.
- ▶ 48,039 elements



Alu Results



Alu Results



Alu Results

- ▶ Elements are preferentially lost from HIGH GC regions
 - ▶ Makes more sense as these are the areas with high gene content
- ▶ It is possible that there is an insertion bias in the younger Alu families (suggested by Belle *et al*, 2005)
 - ▶ No consensus sequence based proof
 - ▶ Future work on individual element sequences



Acknowledgements

- Prof John Brookfield
- Lisa French

