The 3D Organization of Chromatin Explains Evolutionary Fragile Genomic Regions by

Camille Berthelot, Matthieu Muffato,
Judith Abecassis and Hugues Roest Crollius

Speaker: Ilia Minkin

24th April 2015

Two types of genome alterations

1. Small point mutations:

ACTTG
AGT–G

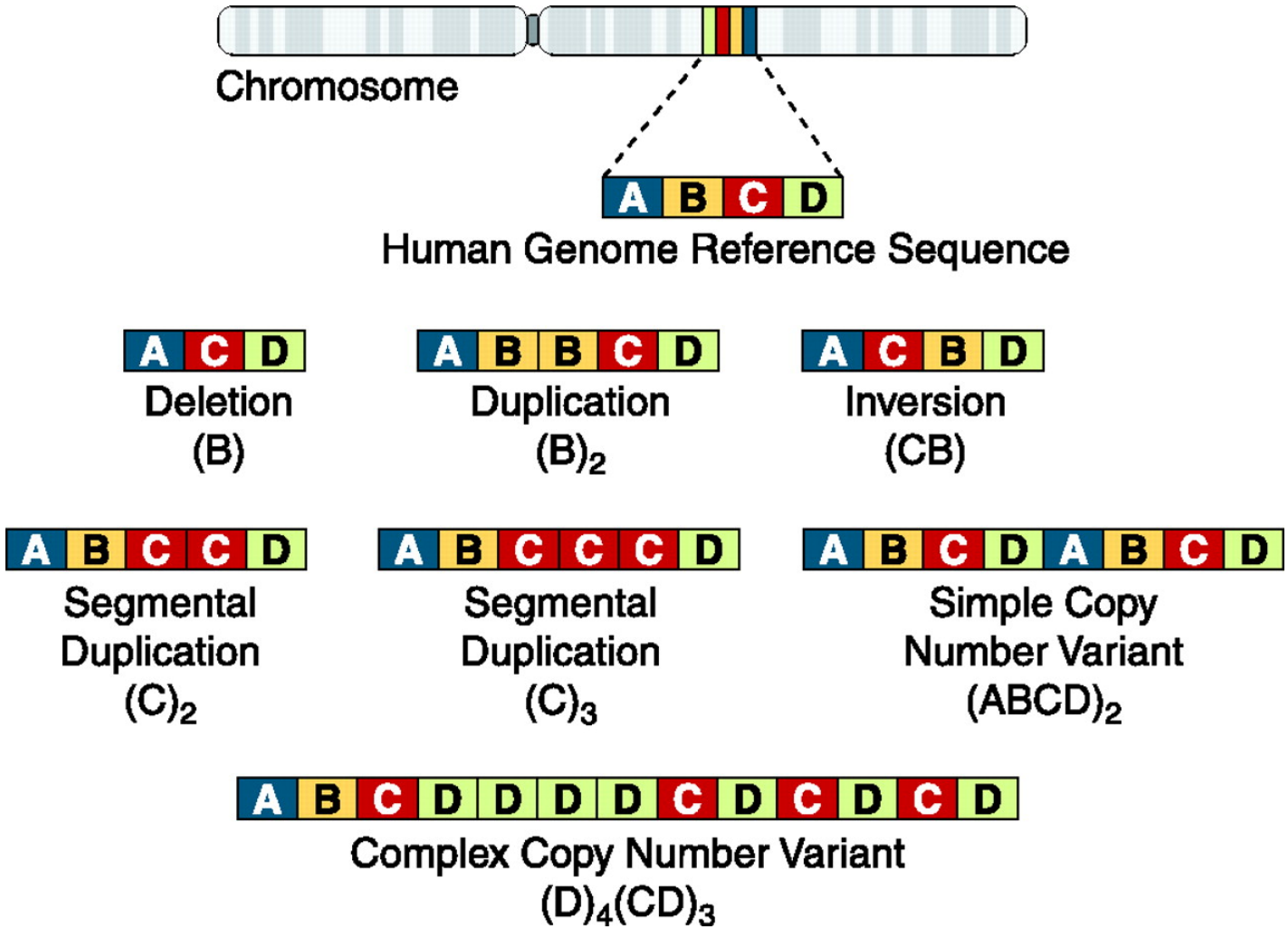
Two types of genome alterations

1. Small point mutations:

ACTTG
AGT–G

2. Large rearrangements:
   - Inversions
   - Transpositions
   - Fusions
   - ...

Genome Rearrangements



Source: *Dierssen et al, 2009*

Motivation

Rearrangements:

- Are a major driving force in evolution
- Play large role in diseases (e.g. cancer)

Known causes:

- Non-homologous end joining
- Non-allelic homologous recombination
- Replication fork stalling
- ...

The Big Question

Are rearrangements more likely to happen in one parts of a genome than the others?

The Big Question

Are rearrangements more likely to happen in one parts of a genome than the others?

Two hypothesis:

1. Uniform distribution of breakpoints
2. Breakpoints get reused

The Big Question

Are rearrangements more likely to happen in one parts of a genome than the others?

Two hypothesis:

1. Uniform distribution of breakpoints
2. Breakpoints get reused

Why are some regions of a genome more "fragile" than the others?

Methodology

Assume that genes are "unbreakable"

Then, how does intergene length affects rearrangement rate?

If breakpoint density is uniform, number of breakpoints should increase in proportion to intergene length — Poisson distribution

Methodology

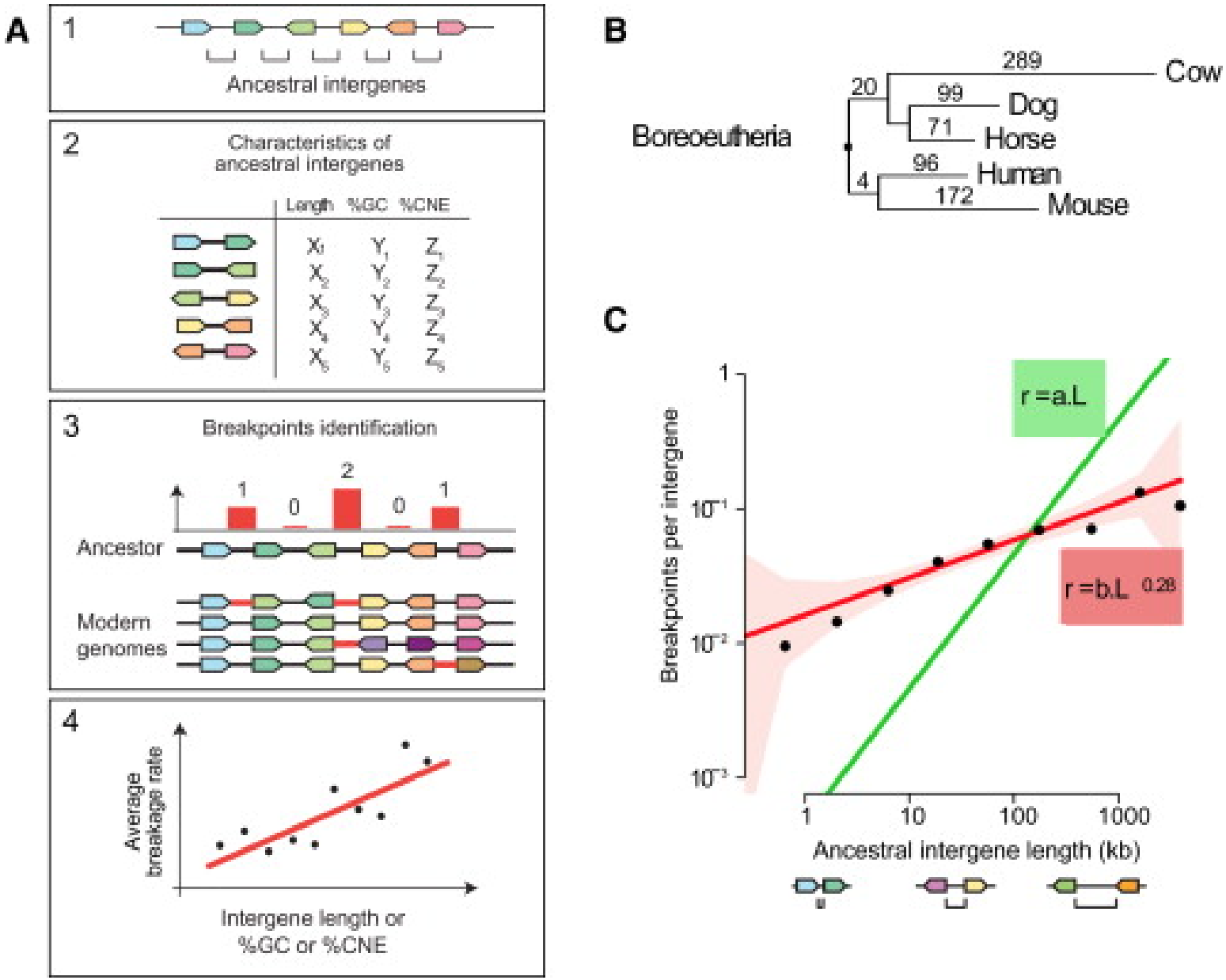
Boreoeutheria: the last common ancestor of primates, rodents, and laurasiatherians

Null hypothesis: breakpoints are distributed uniformly

Stages of the study:

- Reconstruct gene order of Boreoeutheria
- Identify breakpoints w.r.t human, mouse, dog, cow and horse
- Do Poisson regression of "breakage rate"
- Expect linear law if null hypothesis is true

Breakpoint Identification



How to Explain the Equation?

$$r = 2.410^{-3} \times L^{0.38}$$

93% of variation in breakpoint occurrence is explained by intergene length

Maybe GC content is the real cause?

Is GC Content The Explanation?

No.

Is GC Content The Explanation?

No.

Added GC content in regression – non-significant coefficient

Are CNEs The Explanation?

CNEs – conservative non-coding elements.
Located in the intergenes, may affect rearrangement rate. Do they?

Are CNEs The Explanation?

CNEs – conservative non-coding elements.
Located in the intergenes, may affect rearrangement rate. Do they?

No.

Are CNEs The Explanation?

CNEs – conservative non-coding elements.
Located in the intergenes, may affect rearrangement rate. Do they?

No.

Added CNE rate in regression – improved explanation rate only by 3%

Inversions within Intergenes

OK, maybe some breakpoints are more likely than the others.

We work with gene markers — see only rearrangements disrupting their order.

What if there are many missing rearrangements within intergenes?

We can try to simulate rearrangements and see what happens

Inversions within Intergenes

Rearrangements have been shown to occur between regions in close 3D proximity in the nucleus

Contact probability is a good proxy for rearrangement probability

Simulate and sample breakpoint pairs, choose detectable ones

Even if we restrict to detectable breakpoints only, simulation confirms the random breakpoint hypothesis

Open Chromatin is the Culprit

Stick with the simulation – restrict rearrangements to only **open chromatin** regions

Voilà – simulation coincides with the model!
It implies that chromatin state and proximity of genes may explain fragility of some genomic regions

Conclusion

It seems that chromatin state and proximity of genes may explain fragility of some genomic regions

Thank you!