Is the arrangement of the genes in ancestral chromosomes linear?

Eric Tannier, INRIA, LBBE, Lyon, France

Joint work with
S. Bérard, C. Gallien, in Montpellier
B. Boussau, G. Szollosi, V. Daubin, in Lyon

MCEB, June 2012

PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES

Volume 5 FEBRUARY 15, 1919 Number 2

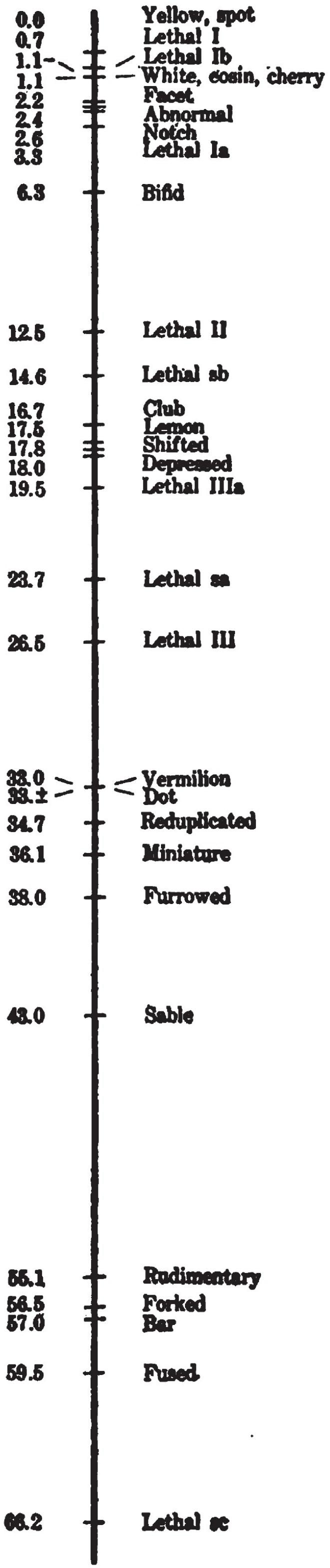
IS THE ARRANGEMENT OF THE GENES IN THE CHROMOSOME LINEAR?

By W. E. Castle

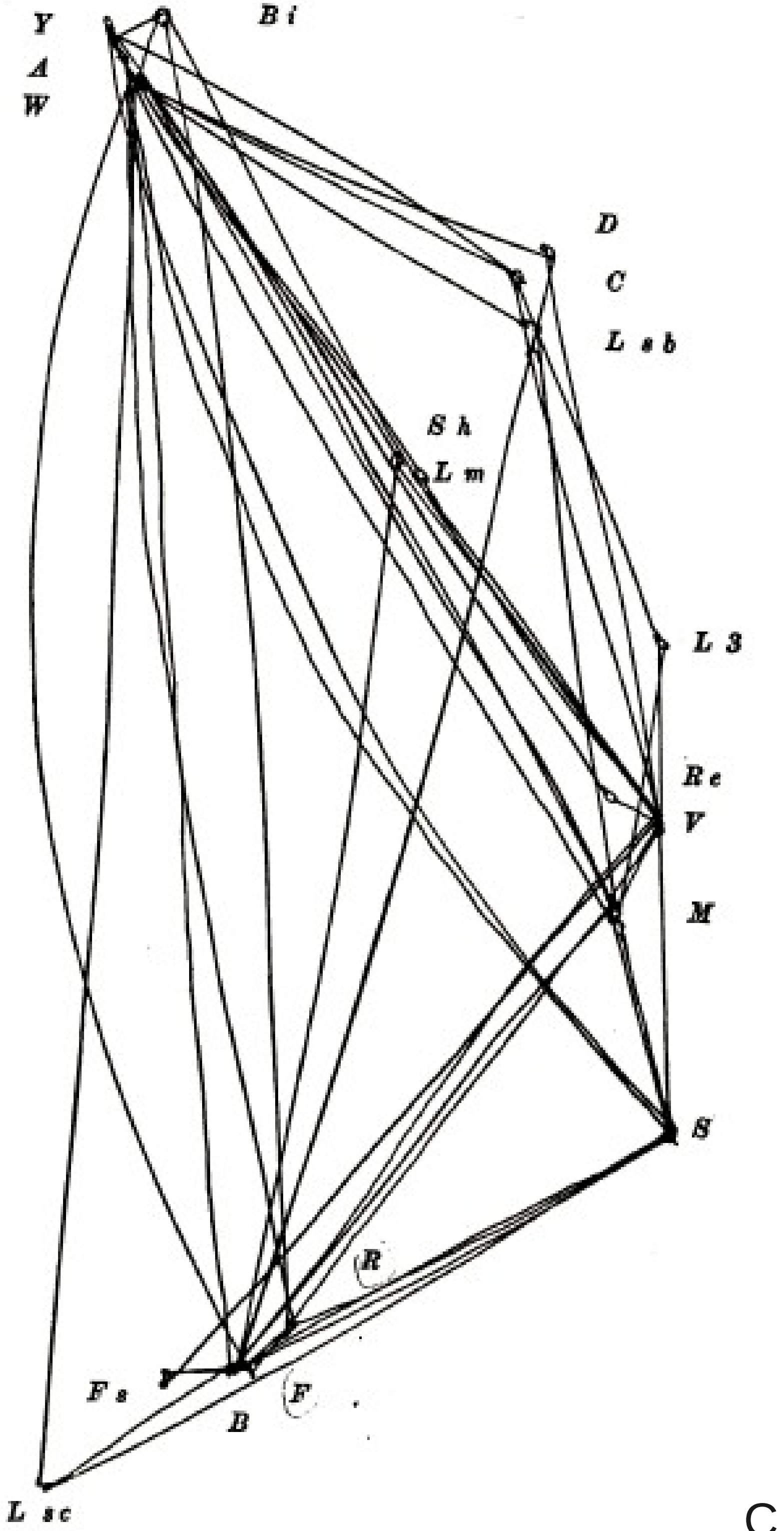
Bussey Institution, Harvard University

Read before the Academy, November 18, 1918

Every biologist is familiar with the remarkable discoveries of Morgan and his associates concerning the germ-cells of Drosophila. One of the most important of these discoveries is concerned with the phenomenon of linked inheritance. This kind of inheritance, while entirely conformable with Mendel's law, forms a very distinct and important class of cases whose existence has been brought to light since the rediscovery of the general law in 1900. Under the general law it is found that characters which behave as distinct units in heredity assort quite independently of each other. Thus if parents are crossed one of which possesses two characters, A and B, while the other lacks them, then the offspring of this cross will transmit A and B sometimes associated in the same gamete, sometimes in different gametes, the two events being under the laws of chance equally probable.



Bridges, Morgan, 1916



Castle, 1919

THE SPATIAL RELATIONS OF GENES

By A. H. Sturtevant, C. B. Bridges, and T. H. Morgan
Columbia University and Carnegie Institution of Washington

Communicated April 11, 1919

ARE GENES LINEAR OR NON-LINEAR IN ARRANGEMENT?

By W. E. Castle
Bussey Institution, Harvard University

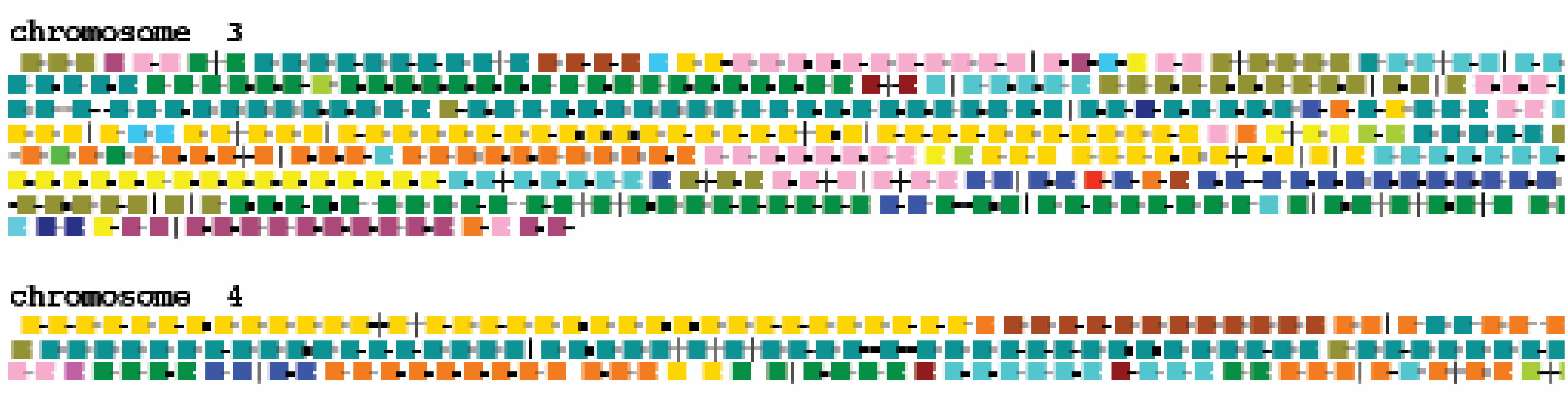
Communicated, August 13, 1919

THE EVIDENCE FOR THE LINEAR ORDER OF THE GENES

By T. H. Morgan, A. H. Sturtevant and C. B. Bridge
Department of Zoölogy, Columbia University

Communicated February 25, 1920

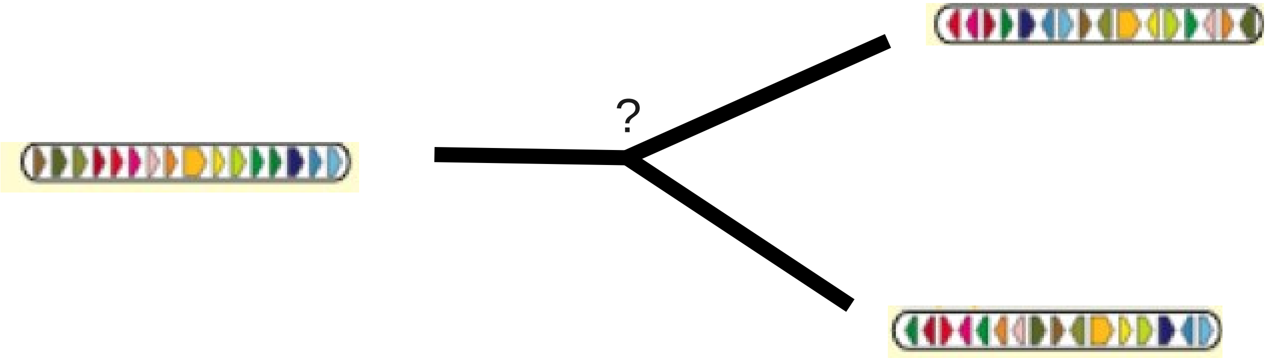
Despite Castle's dictum that we "have failed in two different attempts to establish the linear theory in the case of the three genes yellow, white and bifid," we are bold enough to maintain that the data furnished, and still furnish, the proof called for. We wish to call attention to the fact that in his last paper Castle ignores our proof of the linear order that is furnished by building up the whole chromosome (or even large sections of it) by "distances" so short that no double cross-over classes appear.



Today we tend to believe that chromosomes are and used to be (more or less) linear (or circular) arrangements of genes.

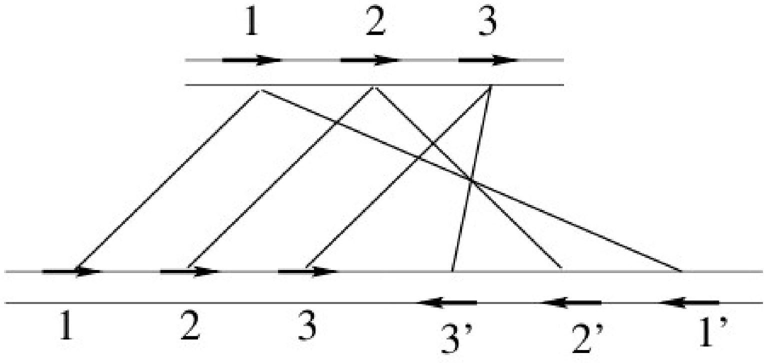
What would we gain by forgetting it?

If we model genomes by linear arrangements (**permutations, strings**), and try to model their evolution, negative computational results arise for the comparison of three genomes



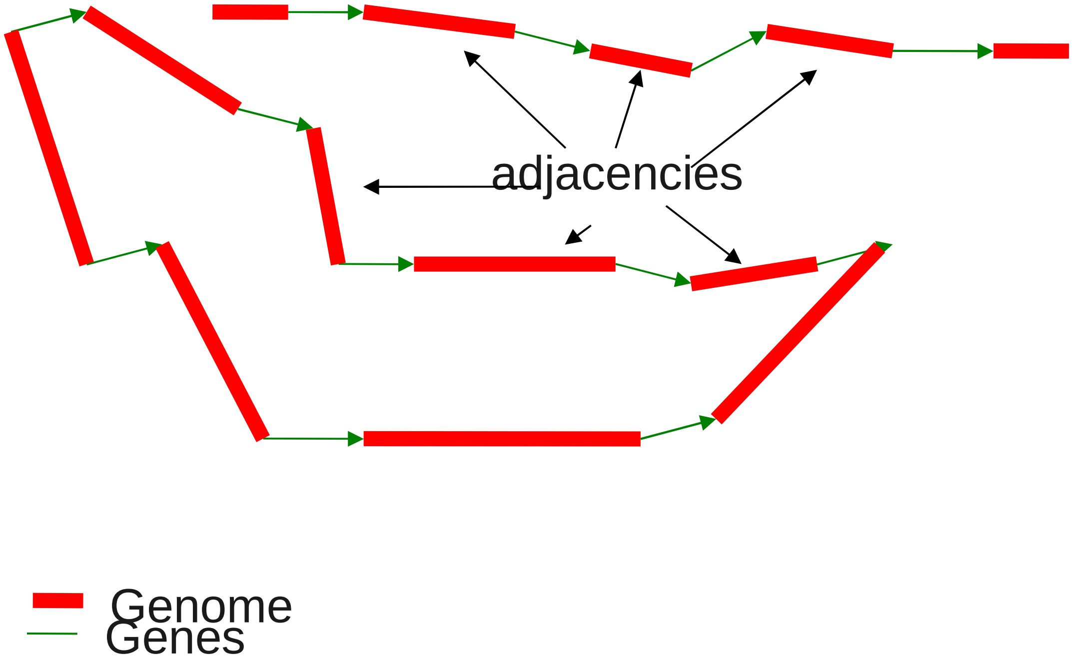
Bryant D: The complexity of the breakpoint median problem, 1998.
Caprara A: The reversal median problem, 2003.

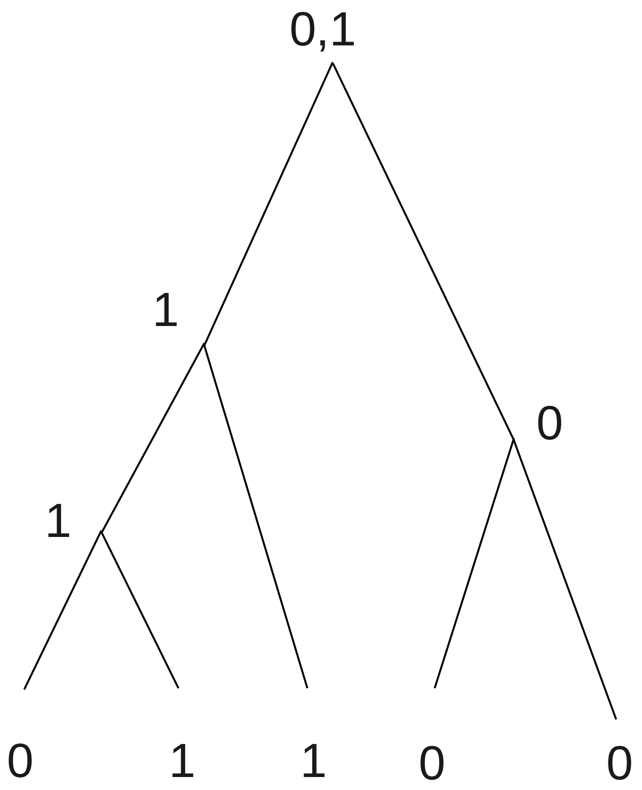
Or even for the comparison of two genomes if they have unequal gene contents due to duplication and loss



Blin G, et al: Comparing Genomes with Duplications: a Computational Complexity Point of View, 2007.

If we relax the definition of a genome to a collection of adjacencies between genes (chromosomes are linear or circular arrangements), some problems are tractable for genomes with equal gene content.

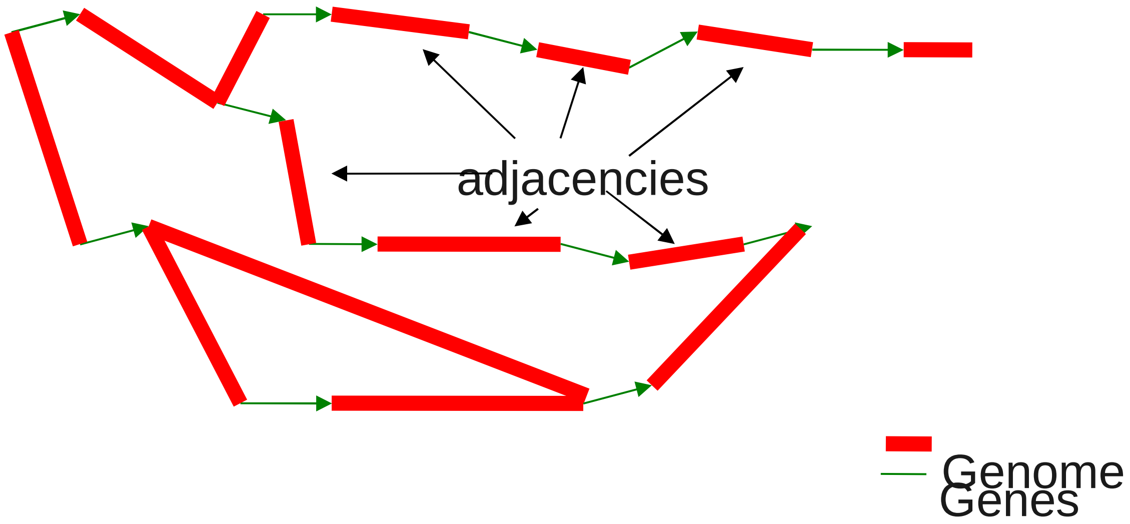




Feijao, Meidanis, IEEE/TCBB 2011

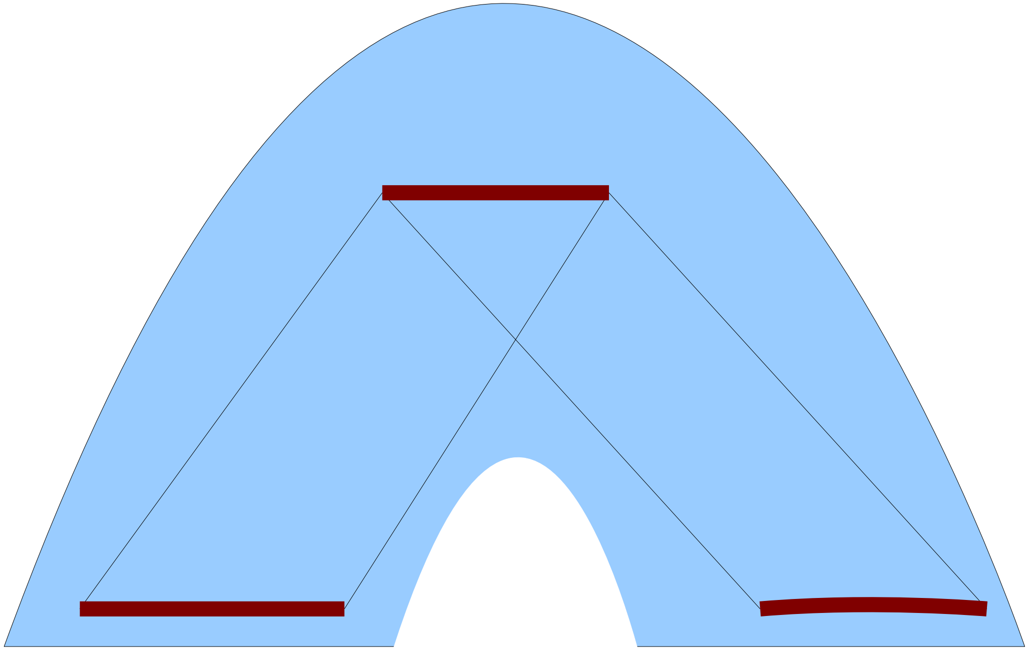
If we relax again: a genome is a collection of relations between genes (and a gene may be in relation with an arbitrary number of other genes)

Advantage: it can be used with any type of relation (synteny, regulation, interaction, metabolism)



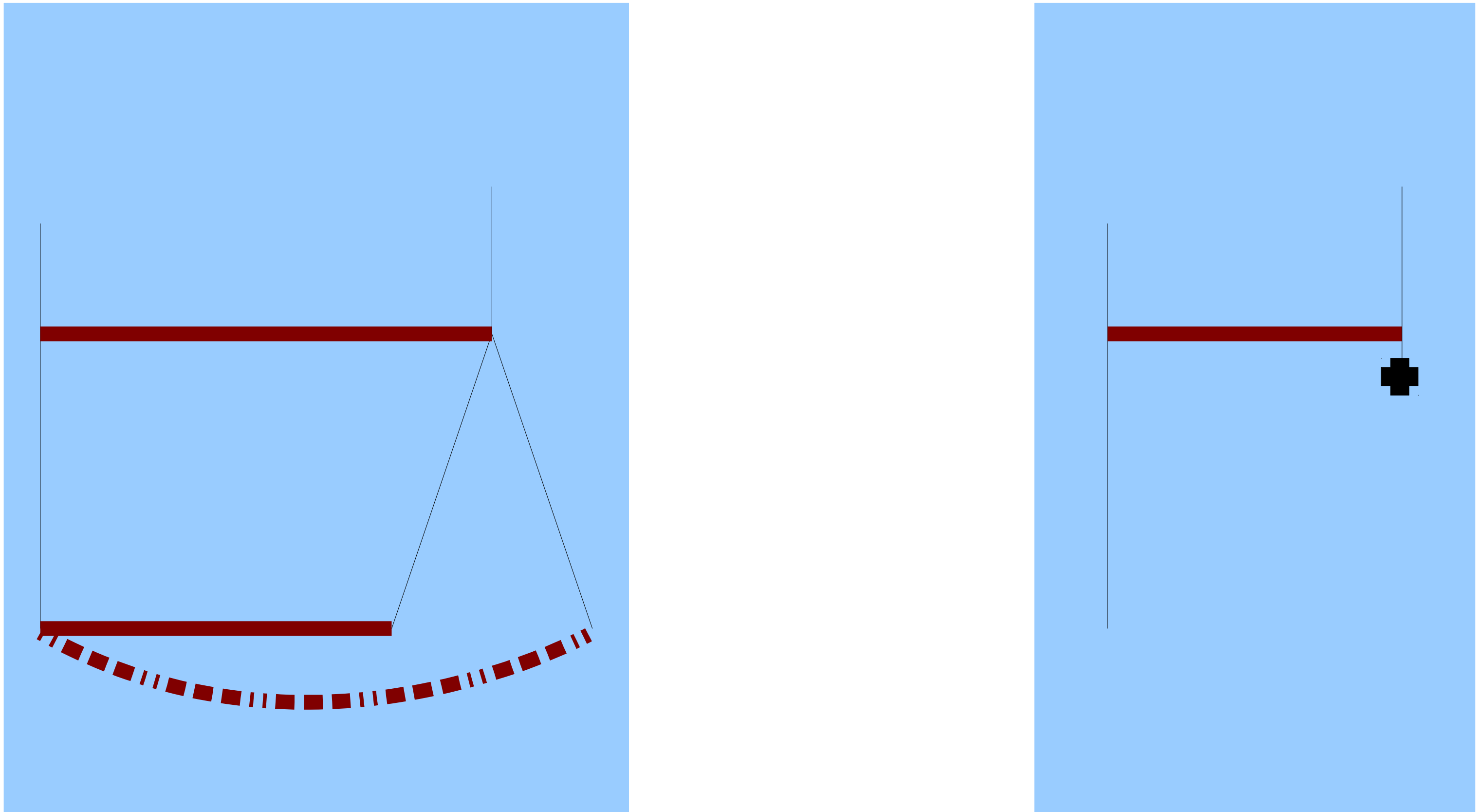
The history of a relation

A relation between two genes can be transmitted with the genes, and in consequence can evolve by
- spéciation



The history of a relation

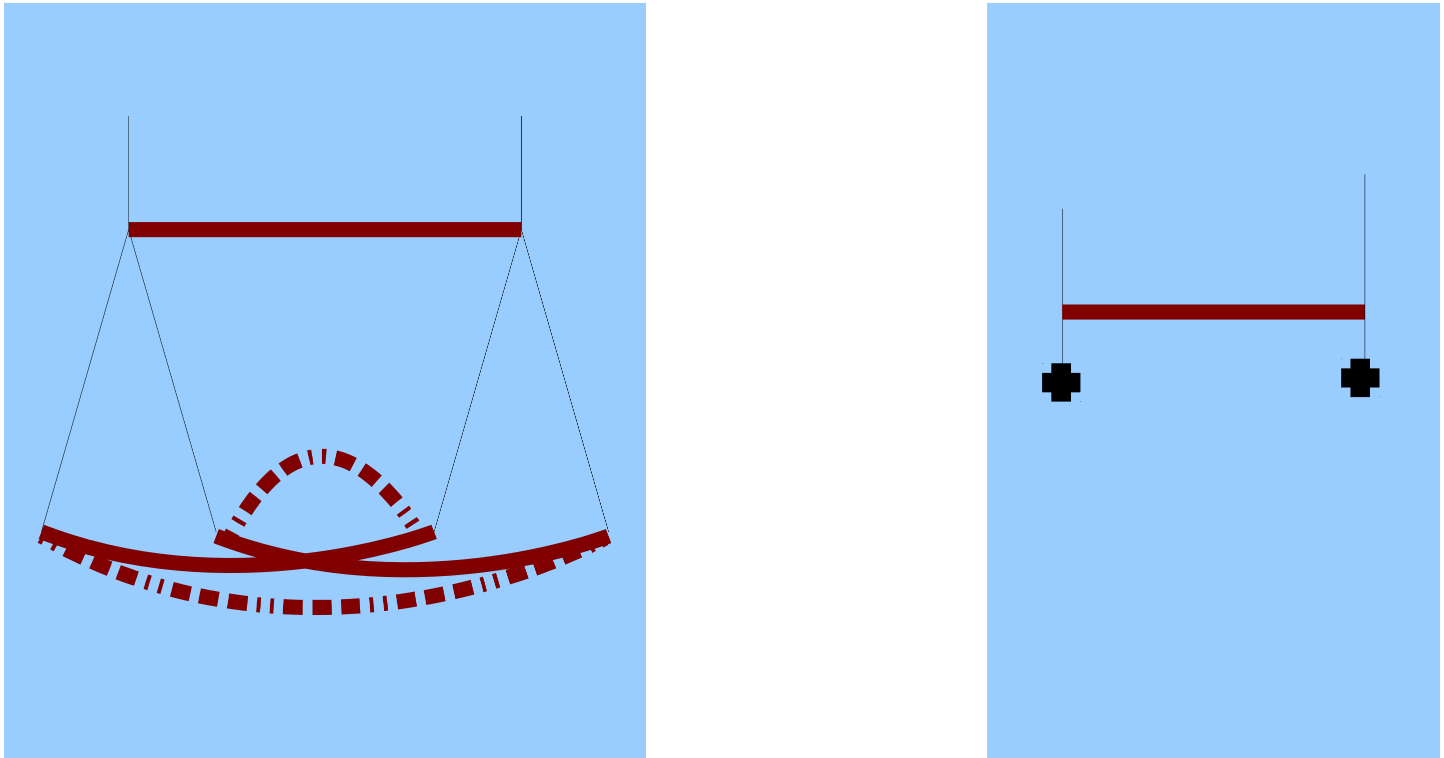
A relation between two genes can be transmitted with the genes, and in consequence can evolve by
- spéciation
- duplication, loss of a gene



The history of a relation

A relation between two genes can be transmitted with the genes, and in consequence can evolve by

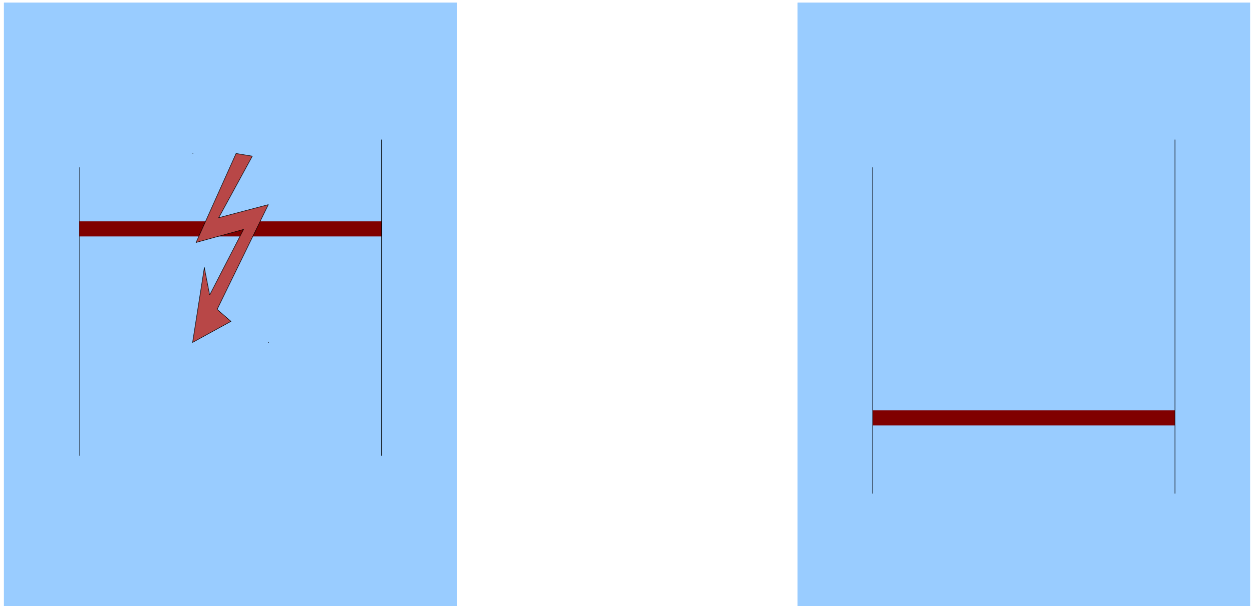
- spéciation
- duplication, loss of a gene
- duplication, loss of the two genes



The history of a relation

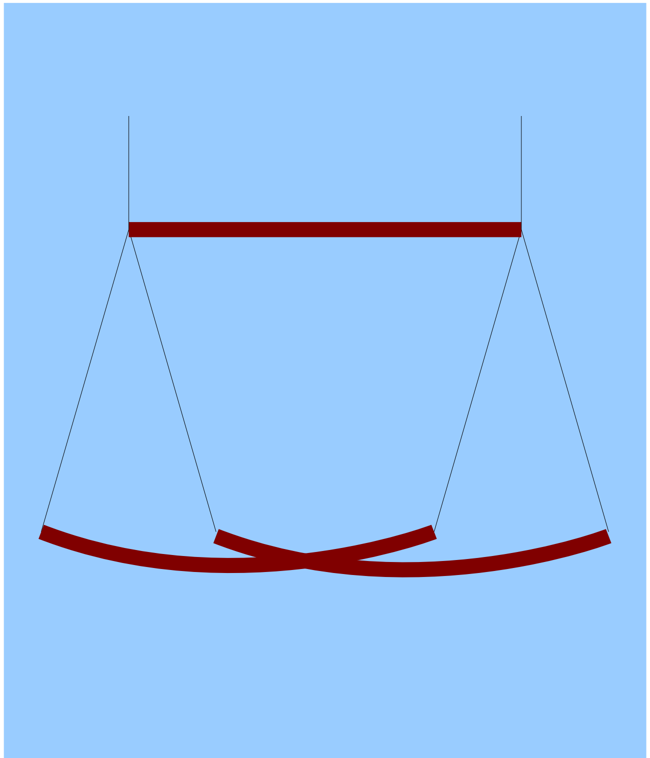
A relation between two genes can be transmitted with the genes, and in consequence can evolve by

- spéciation
- duplication, loss of a gene
- duplication, loss of the two genes
- gain or loss of the relation (rearrangement)



Given a species tree, a set of reconciled gene trees (with duplications and losses) and the organisation of the extant chromosomes, we can compute ancestral genomes (ancestral set of relations between genes) minimizing the number of gains and breakages of adjacencies in a few minutes.

For each couple of gene tree nodes (from one or two gene trees)
C0 : cost of a history in which genes are not in relation
C1 : cost of a history in which genes are in relation



In the case of two speciation nodes

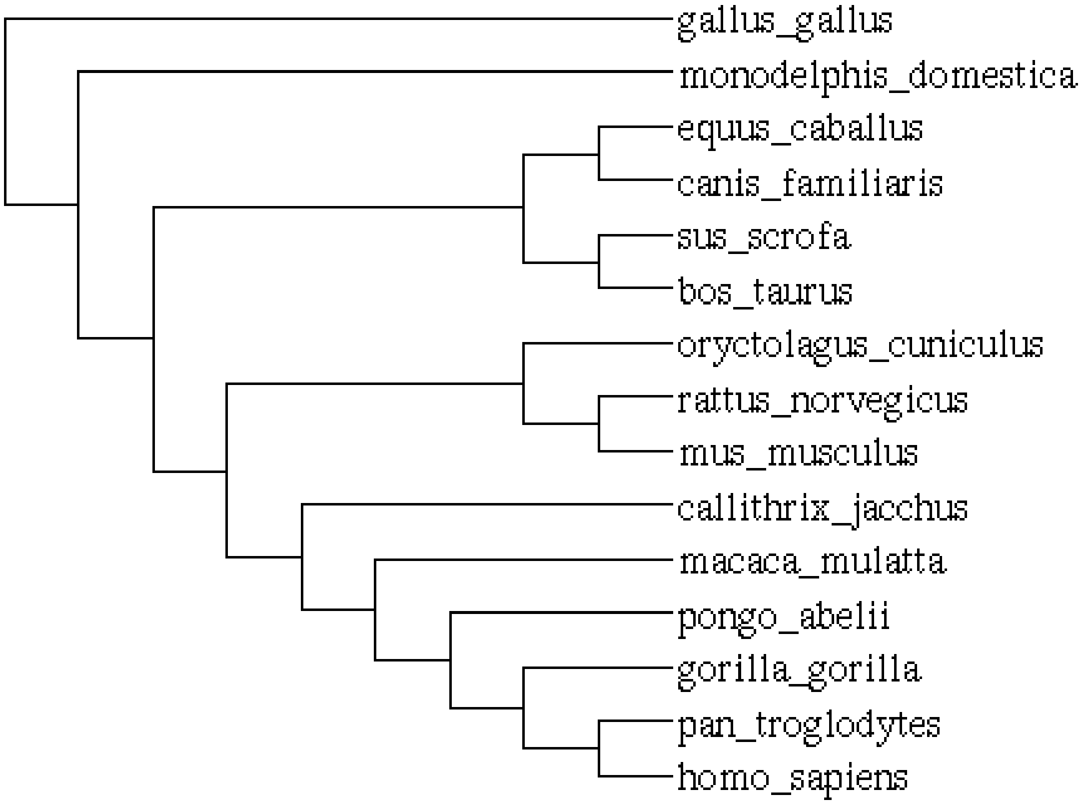
$$c_1(v^1,v^2) = \min$$

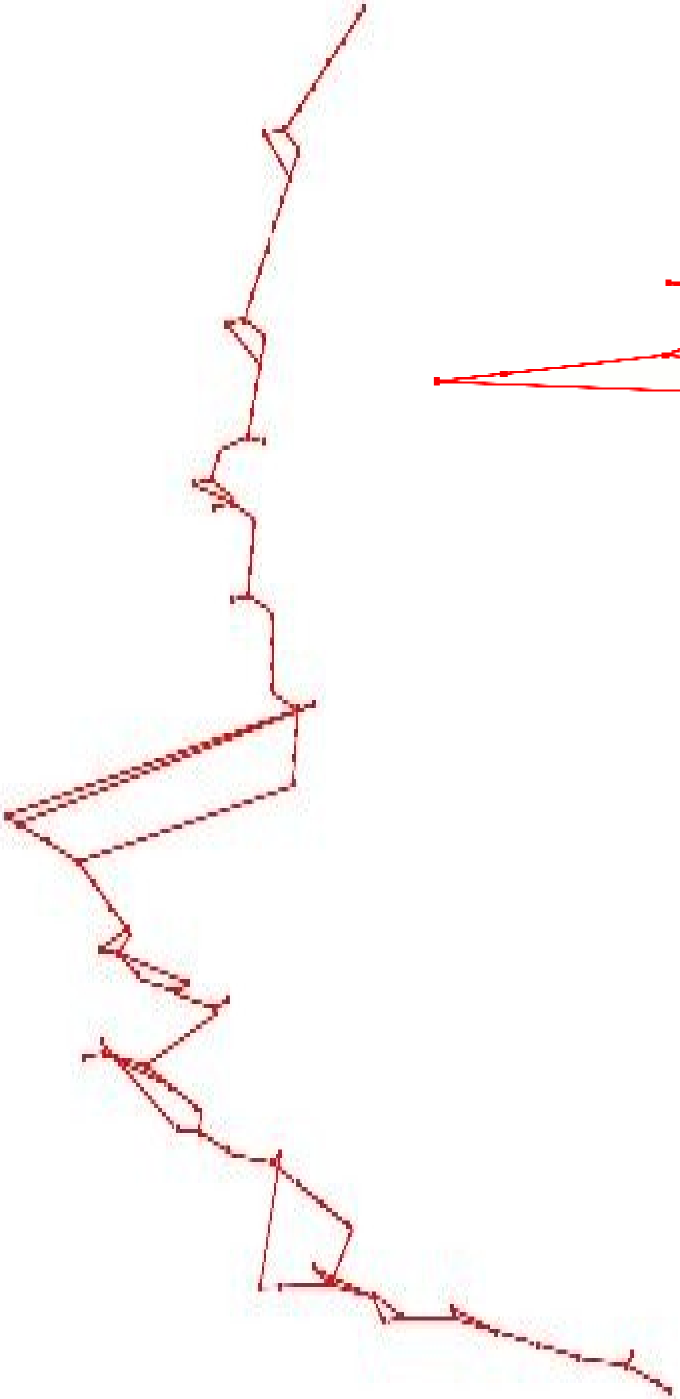
$$\begin{cases} c_1(ca(v^1),ca(v^2)) & + & c_1(cb(v^1),cb(v^2)) & & \\ c_1(ca(v^1),ca(v^2)) & + & c_0(cb(v^1),cb(v^2)) & + & C(Break) \\ c_0(ca(v^1),ca(v^2)) & + & c_1(cb(v^1),cb(v^2)) & + & C(Break) \\ c_0(ca(v^1),ca(v^2)) & + & c_0(cb(v^1),cb(v^2)) & + & 2*C(Break) \end{cases}$$

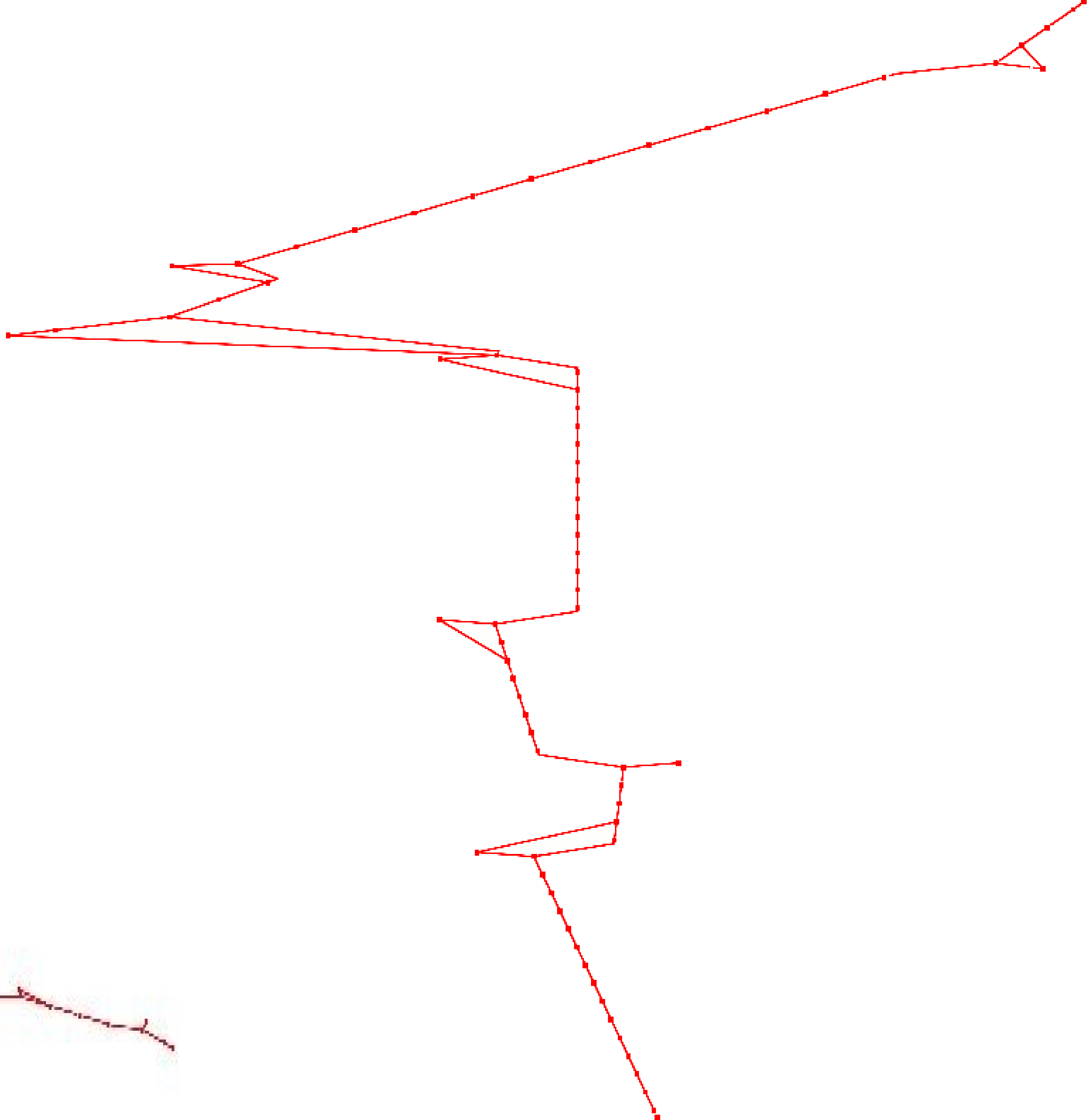
$$c_0(v^1,v^2) = \min$$

$$\begin{cases} c_0(ca(v^1),ca(v^2)) & + & c_0(cb(v^1),cb(v^2)) & & \\ c_1(ca(v^1),ca(v^2)) & + & c_0(cb(v^1),cb(v^2)) & + & C(Gain) \\ c_0(ca(v^1),ca(v^2)) & + & c_1(cb(v^1),cb(v^2)) & + & C(Gain) \\ c_1(ca(v^1),ca(v^2)) & + & c_1(cb(v^1),cb(v^2)) & + & 2*C(Gain) \end{cases}$$

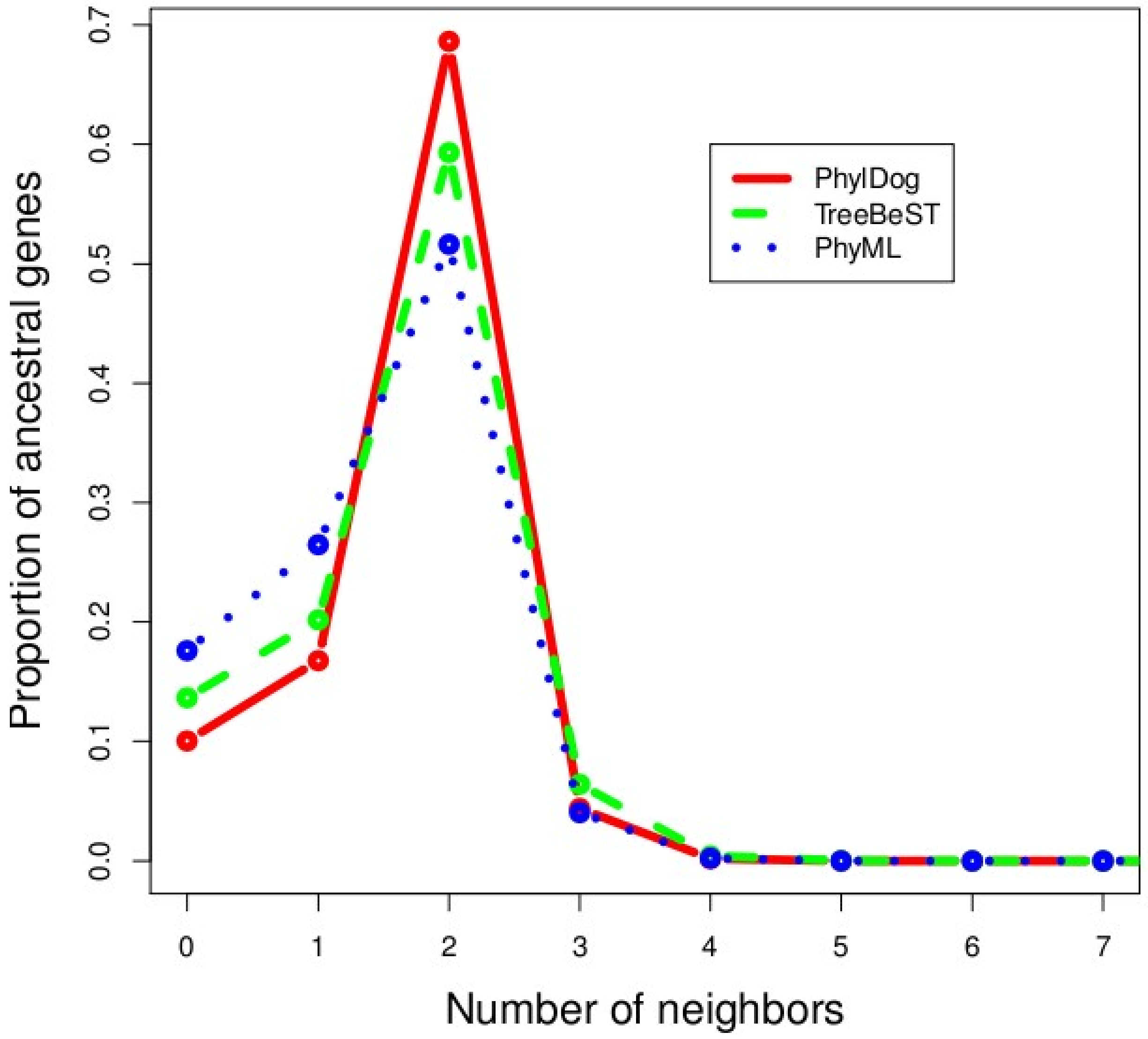
Is the arrangement of the genes in ancestral mammalian chromosomes linear?







Non linear ancestral chromosomes measure the quality of gene trees



Additional material :

a paper at ECCB 2012 (to appear in Bioinformatics)

a software called DeCo written by Sèverine

a Ph-D position on models for co-evolution

Eric.Tannier@inria.fr