Algorithms in Comparative Genomics (392189)

Dr. Daniel Dörr

Winter 2018/19

1. Introduction

Literature:

Pevzner, P., & Tesler, G. (2003). *Genome rearrangements in mammalian evolution: lessons from human and mouse genomes.* Genome Research, Vol. 13(1), pp. 37–45.

Pevzner, P. (2000). *Chapter 10: Genome Rearrangements.* In: Pevzner, P. (ed.) Computational Molecular Biology: An Algorithmic Approach, MIT Press, pp. 175-181.

1.1. What are and what causes genome rearrangements?

- DNA strand breaks are common, but usually repaired by DNA-ligases.
- If two chromosomal strands are spatially close and double strand breaks occur simultaneously, errors in re-ligating the correct strands with each other can occur; these are called *genome rearrangements*.
- Mutagens: chemicals, UV-light, etc..
- Topoisomerases induce double strand breaks when changing chromatid states between supercoiled and uncoiled regions.

Genome rearrangements are studied by observing the succession of genomic markers of two genomes from different species, as exemplified in Figure 1.1.

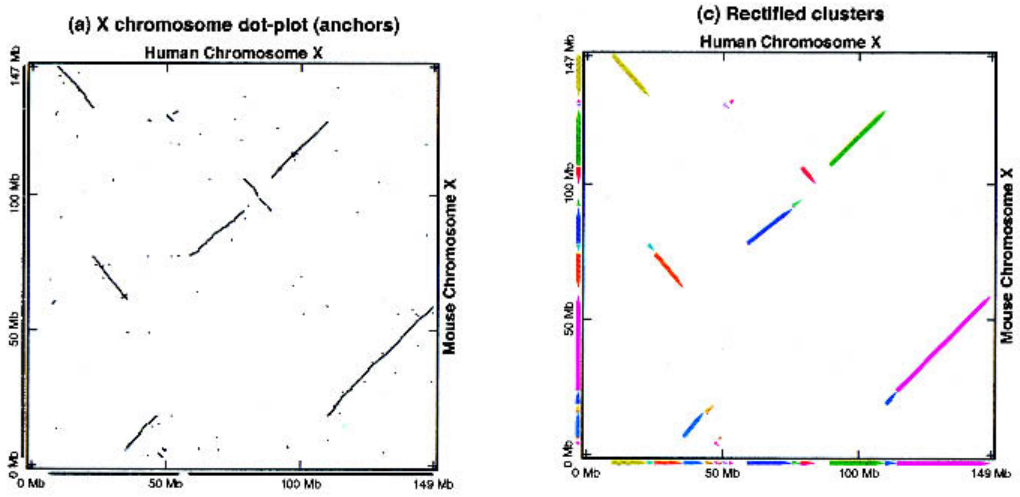
There are many types of genome rearrangements: reversal, transposition, translocation, block interchange, fusion, fission, circularization, linearization (see Figure 1.2).

Note: We consider rearrangement events to be undirected!

1.2. Unichromosomal genome model

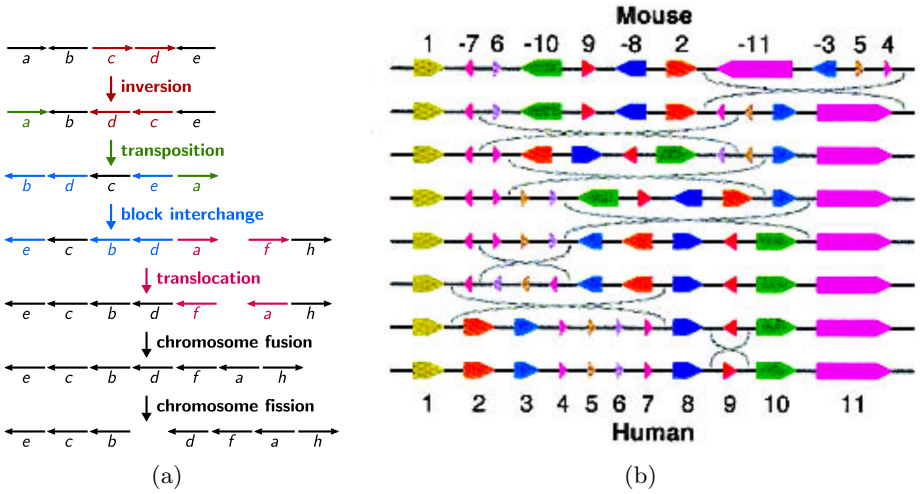
A *chromosome* is a DNA molecule composed of antiparallel strands that can be read in either of the two possible directions. A *gene* is associated with an interval on a DNA strand hand has a *reading direction* (5'-to-3' or left-to-right, by convention).

In the next two chapters, algorithms for genome rearrangement are discussed that are based on a unichromosomal, linear genome model, formally defined as follows:



(a) (b)

Figure 1.1.: A Gene order comparison of human and mouse chromosome X, source: Pevzner and Tesler (2003).



(a) (b)

Figure 1.2.: (a) Examples of genome rearrangements; (b) Rearrangement scenario between chromosome X of human and mouse, source: Pevzner and Tesler (2003).

Definition 1. *A* signed permutation *is a permutation on the set* $\{1, \ldots, n\}$ *in which each element has an orientation, indicated by a sign "+" or "−". To simplify, the "+" is usually omitted.*

A permutation of size n *representing a linear chromosome with* n *genes is bordered by*

0 *and* $n+1$.

Example 1. $\pi^1 = (0\ -2\ -1\ 4\ 3\ 5\ -8\ 6\ 7\ 9\ 10)$