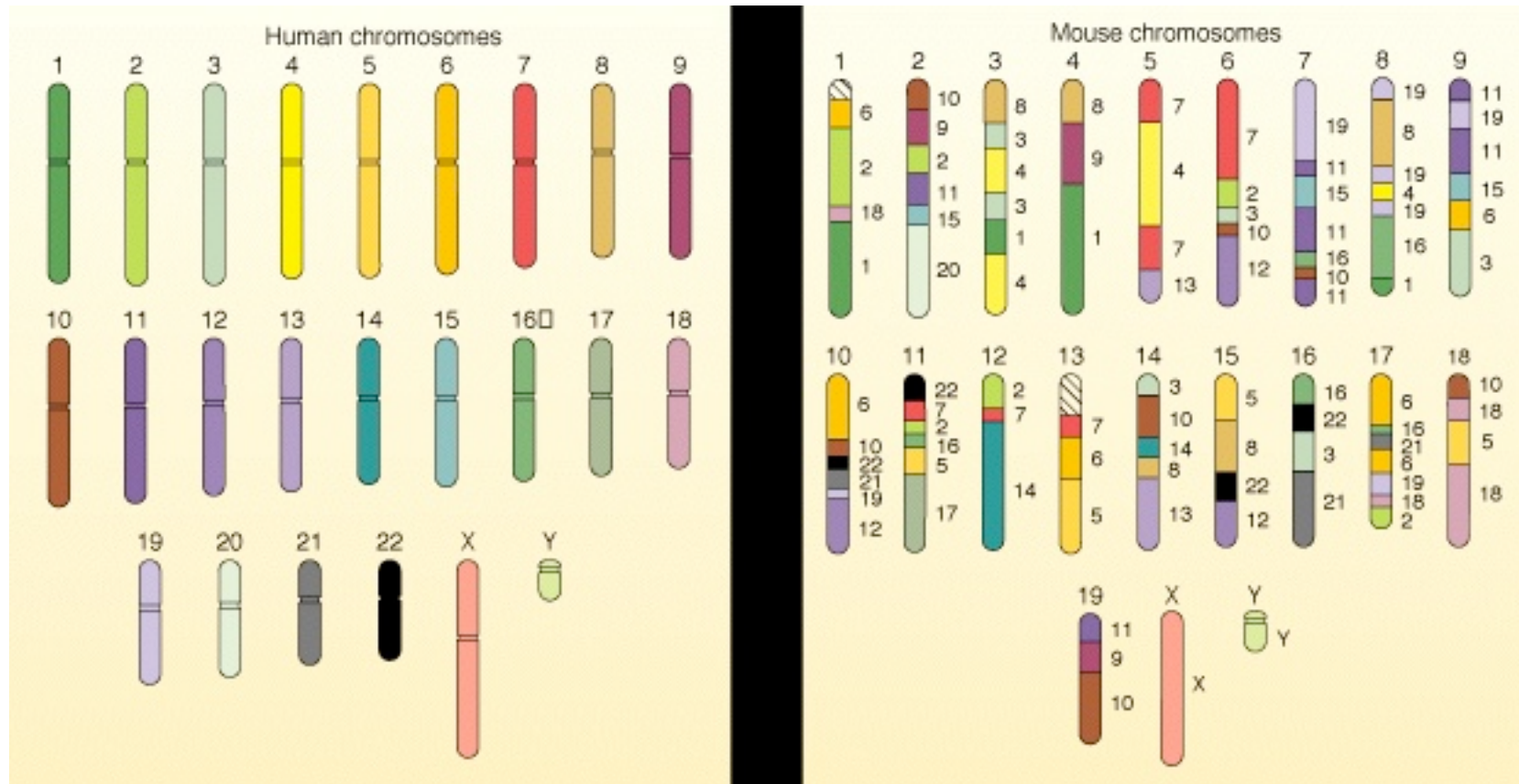


Intro



<http://fig.cox.miami.edu/Faculty/Dana/synteny.jpg>

Whole-genome alignment

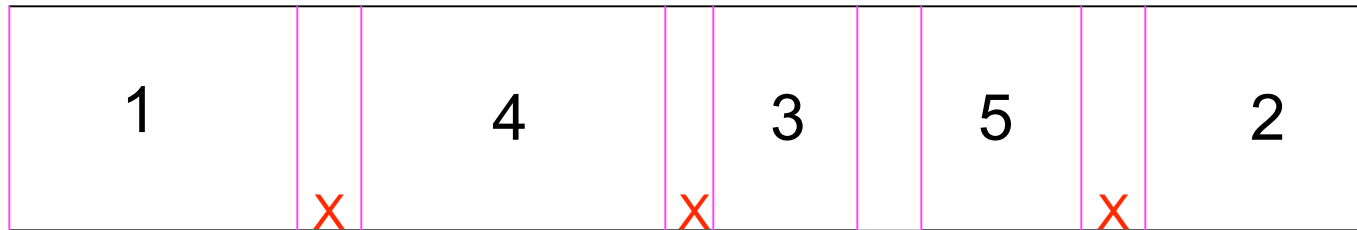
- Advanced data structures can also be used to efficiently speed up genomic alignments of closely-related organisms.
- Ideas behind the MUMmer 1.0 tool:
 - Find maximal unique matches (MUMs)
 - Find the best arrangement of MUMs
 - Fill in the gaps

MUM rationale

- Long unique matches have a high probability of being included in the final genomic alignment.
- We need to set the minimum MUM length high-enough, however, to avoid random noise.

Overview

Genome A



Genome A'

We have 5 matches that can not be extended to left or right

We have 4 gaps to fill between these matches

Finding MUMs

- MUMs are by definition unique in both sequences
 - Requires building a generalized suffix tree for both genomes
 - Internal nodes w/ only two leaves, one from each input, are unique and not right-extensible
 - Check for left-extensibility, then go!

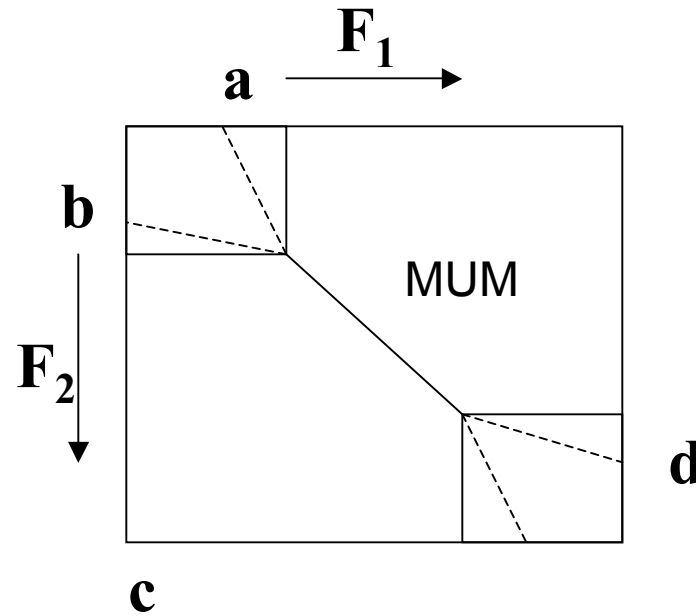
Early MUMmer Algorithms

- Arranged MUMs relative to one genome using Longest Increasing Subsequence (LIS) algorithm
- Filled in small gaps using dynamic programming
 - Space inefficient for large gaps

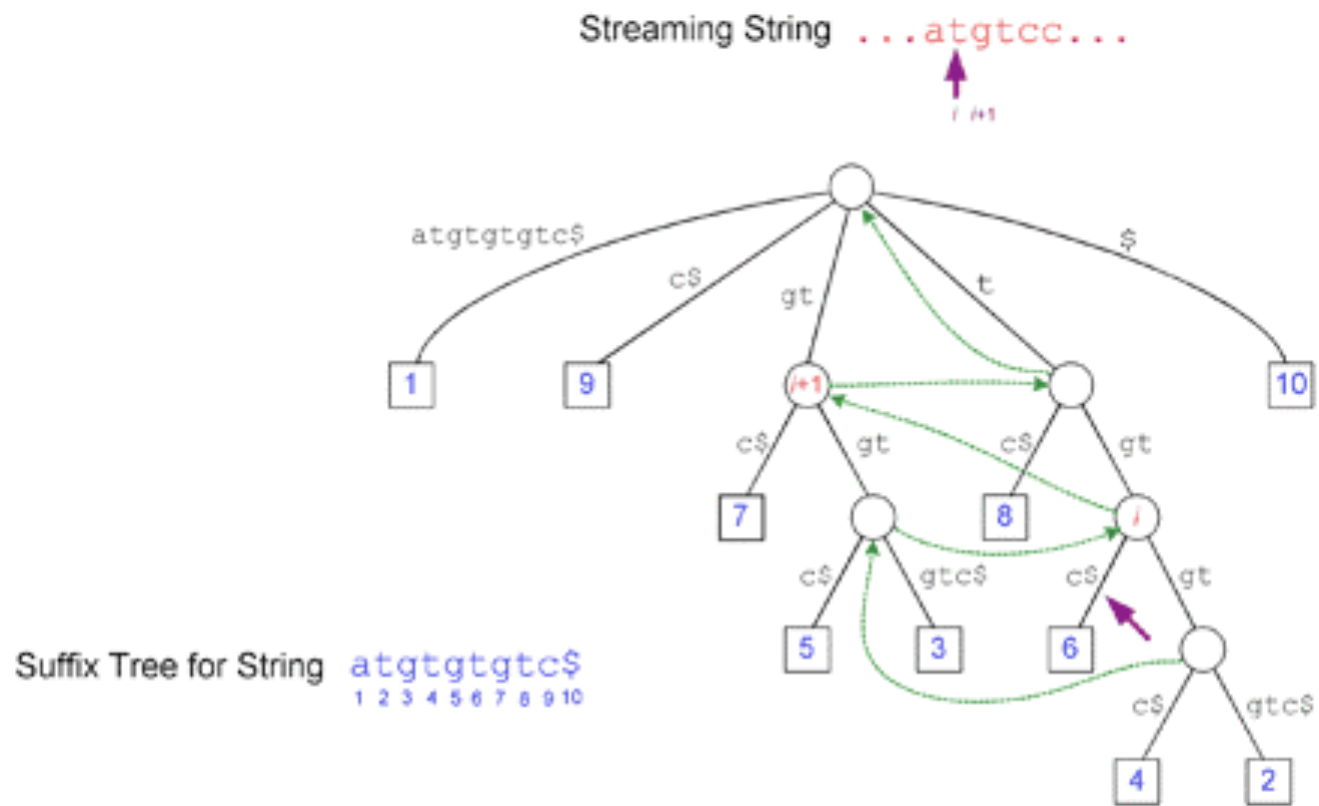
MUMmer 2.0

- Improved space implementation of suffix tree using a few tricks (17 bytes/base)
- Introduced banded dynamic programming and advanced clustering to tackle larger gaps
- Used suffix tree “streaming” of multiple queries against a reference

Banded Dynamic Programming



- **Compute only lower and upper rectangles based on desired percent similarity**



Suffix links are in green

From Delcher et al., 2002, Nucleic Acids Res30(11):2478-83

Applications

- Comparing different genome assemblies
- Locating genome duplications and conserved segments
- Gene finding through comparative genomics
- Analyzing pathogenic bacteria against their harmless close relatives

MUMmer 2

- Three times faster
- One-third memory usage
- Support protein sequence and multiple sequences.
- Entire human chromosomes
- Can align **millions** of nucleotides in **a few minutes** on a **desktop computer**.