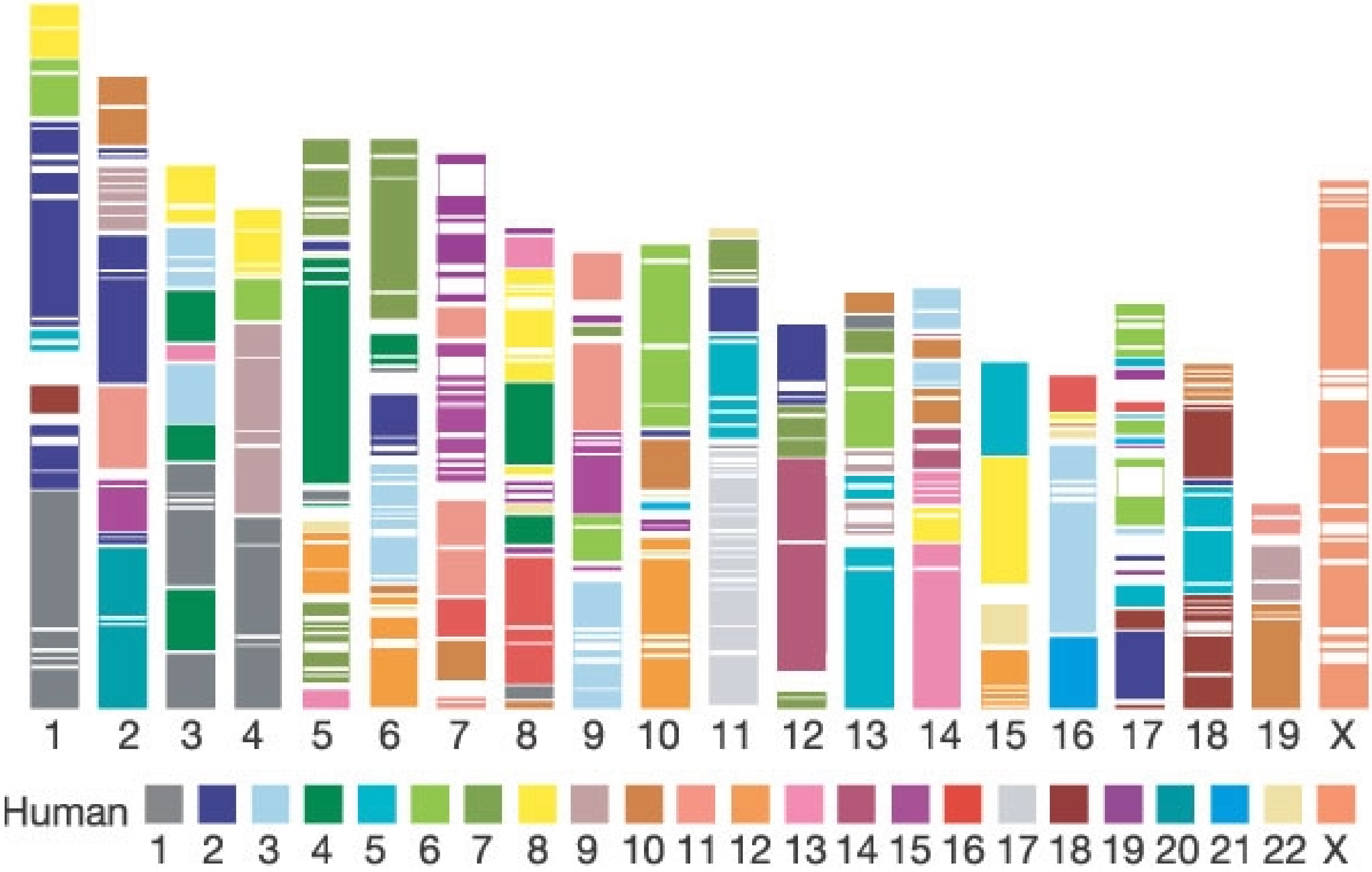
Genome Rearrangements CS426

Uri Keich

With some material taken from Ron Shamir's
online lecture notes

Types of rearrangements

- Unichromosomal genome:
  - Reversal (inversion)
- Multichromosomal genome:
  - Fusion
  - Fission
  - Translocation
  - These can be modeled as reversals



Waterston et al. Nature 2002

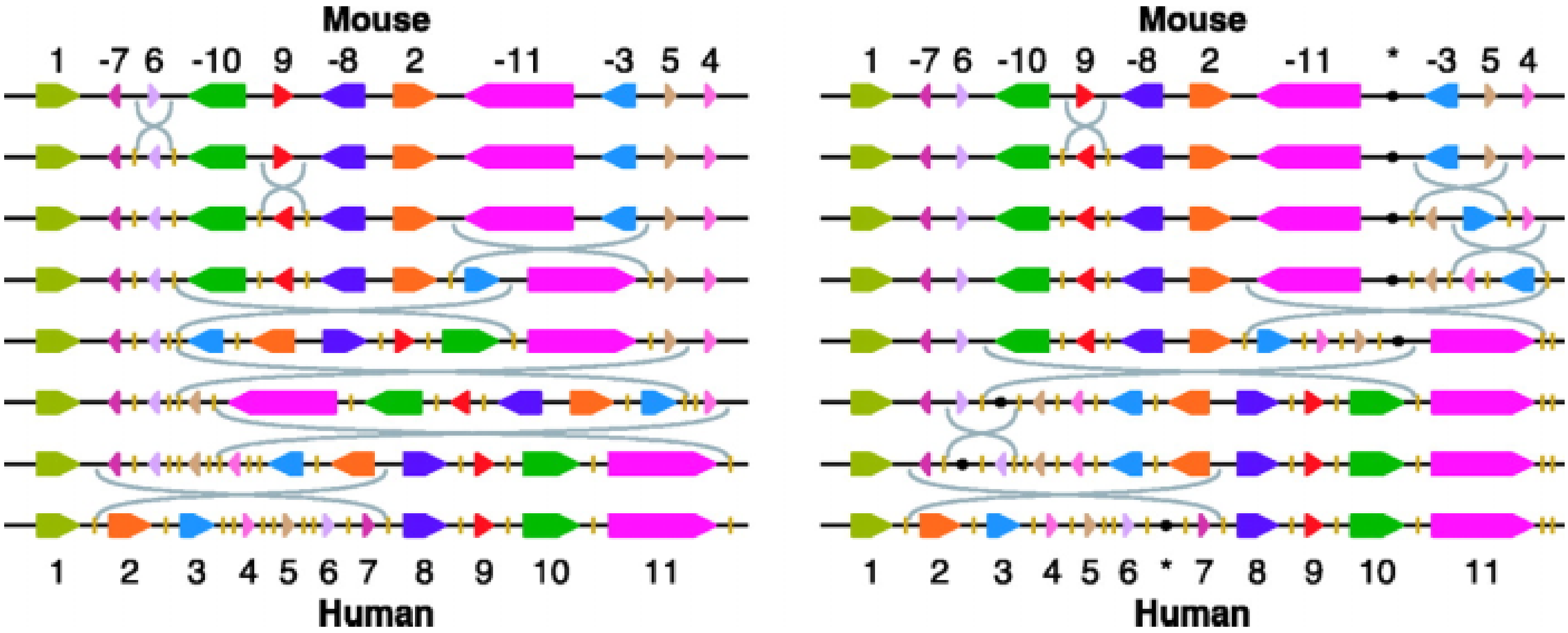
Studying evolution

- Number of rearrangement events yields an estimate of the evolutionary distance between the species
- Avoids the ``too close’’ and ``too far’’ problems associated with alignments
- Looking at the whole genome avoids the potential problems of tree per gene

Synteny blocks

- Large unshuffled blocks of genes are called *synteny blocks*
- The relative order of the synteny blocks is represented by a permutation $\pi = \pi_1\pi_2\ldots\pi_n$
  - Human X: 1 2 3 4 5 6 7 8 9 10 11
  - Mouse X: 1 7 6 10 9 8 2 11 3 5 4

Human and mouse X



Tesler & Pevzner 2003

Reversal distance

- A reversal $\rho(i,j)$ reverses the order of $\pi_i \ldots \pi_j$: $\pi\,\rho(i,j) = \pi_1 \pi_2 \ldots \pi_{i-1}\, \pi_j \ldots \pi_i\, \pi_{j+1} \ldots \pi_n$
- Reversal distance $d(\pi)$= minimal number of reversals that will take π to the identity.

breakpoints

- A breakpoint in π is a pair $|\pi_i - \pi_{i+1}| > 1$
- Let b(π) denote the number of breakpoints of π
- Claim 1.

$$b(\pi)/2 \le \lceil b(\pi)/2 \rceil \le d(\pi) \le n-1$$

Find the reversal distance

- NP hard (Caprara 99)
- Approximation algorithms
- A *strip* is a maximal subsequence without breakpoints
- A strip can be increasing or decreasing (1)
- Convenient: add 0 and n+1

A 4-approximation algorithm

- Claim 2. If π contains a decreasing strip then there is a reversal that decreases b(π)
- Proof:
  - Look for K, the smallest k>0 that is part of a decreasing strip
  - Consider K-1: increasing strip
- If there is a decreasing strip reduce b(π) otherwise reverse an increasing strip
- Number of steps is bounded by 2b(π)

A 2-approximation algorithm

- Sankoff & Kececioglu 95
- Claim 3. If there exists a decreasing strip and every reversal of such a strip results in "decrease-less" π then there exists a reversal that removes 2 breakpoints.
- Proof. Consider π_i the smallest and π_j the largest element in a decreasing strip

Breakpoint graph

- G(π) is a graph on n+2 vertices where π_i and π_{i+1} are joined by a black edge while i and i+1 are joined by a gray edge
- The graph is balanced: the number of black edges equals the number of gray edges incident to each vertex
- There exists a cycle decomposition of G(π) into edge disjoint alternating cycles
- Let c(π) be the maximal number of edge disjoint alternating cycles

Reversal distance and c(π)

- Claim 4 (Bafna & Pevzner 93). For every permutation π and reversal ρ, $c(\pi\,\rho) - c(\pi) \leq 1$
- Corollary: $d(\pi) \geq n+1 - c(\pi)$
- Pevzner 00: equality holds “for most biological examples”
- Finding c(π) is NP-hard though (Caprara 99)

Signed permutations

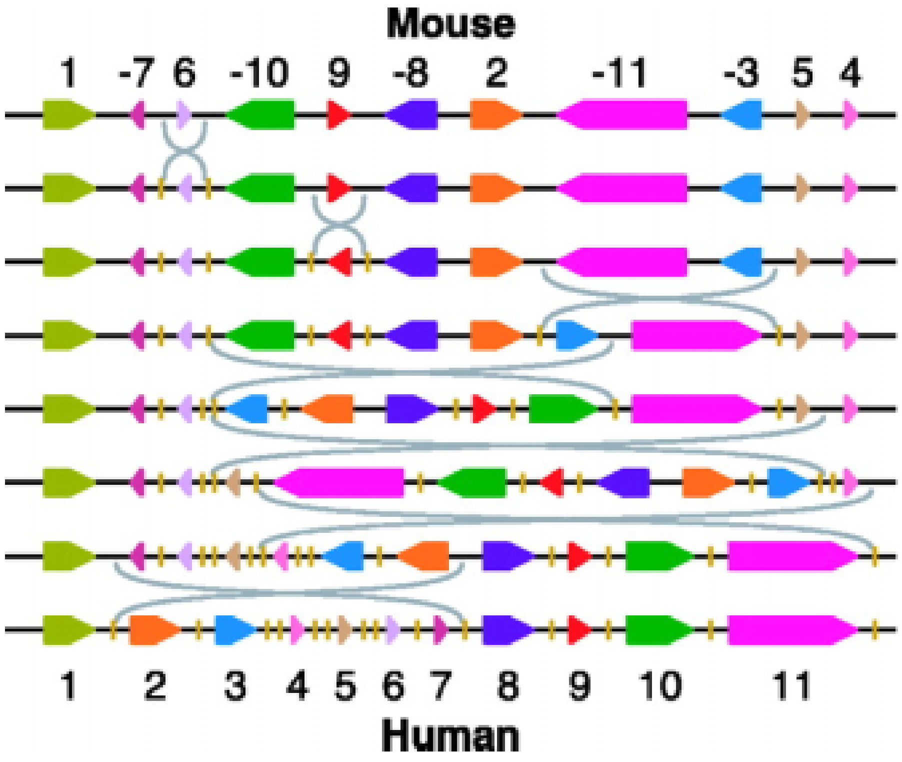
- Genes actually have a direction
- A reversal of part of the chromosome will reverse the direction of the genes
- Signed permutation: permutation+signs
- A reversal $\rho(i,j)$ is defined as

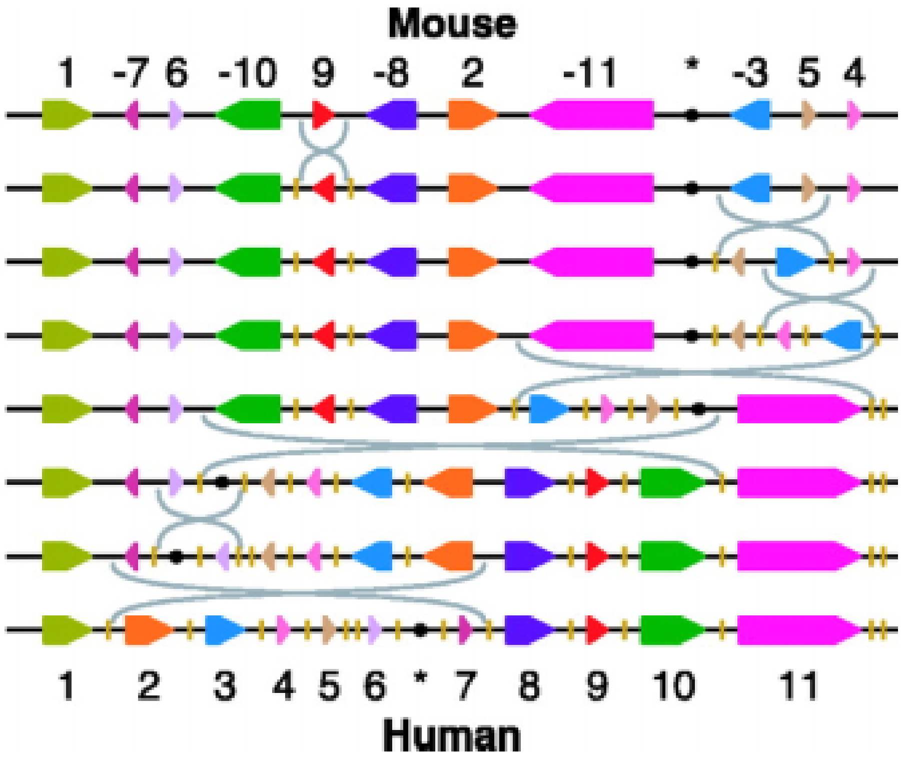
$$\pi\,\rho(i,j) = \pi_1, \pi_2 \ldots, \pi_{i-1}, -\pi_j \ldots, -\pi_i, \pi_{j+1} \ldots \pi_n$$

The (signed) reversal distance

- $d(\pi)$= minimal number of (signed) reversals that will take π to the identity
- Problems: find $d(\pi)$ and minimal rearrangement scenarios
- Clearly $d(\pi) \leq n + (n-1) = 2n-1$

Human and mouse X





Tesler & Pevzner 2003

Surprise!

- Polynomial - Hannenhalli & Pevzner 95
- Key example: π =(3 4 1 2) signed distance is 3 but unsigned is 2 (cutting strips)
- For signed permutations there exists an optimal scenario which does *not* cut strips
- For unsigned permutations we might have to cut strips
- Not longer than 2 though (Hannenhalli & Pevzner 96)

Signed permutations to unsigned

- Transform π a signed permutation on n elements to π', a signed one on 2n+2
- $\pi_i>0$ is replaced by $2\pi_i-1$, $2\pi_i$
- $\pi_i<0$ is replaced by $2\pi_i$, $2\pi_i-1$
- π = -1, 3, 5, 4, 6, -2 is replaced by
- π' = 0, 2 1, 5 6, 9 10, 7 8, 11 12, 4 3, 13
- A reversal $\rho(i,j)$ is legal if *i* is odd and *j* is even

Breakpoint graph (2)

- Given a signed π construct G(π’) as before
- Remove internal black and gray edges (every other edge)
- π’ = 0, 2 1, 5 6, 9 10, 7 8, 11 12, 4 3, 13

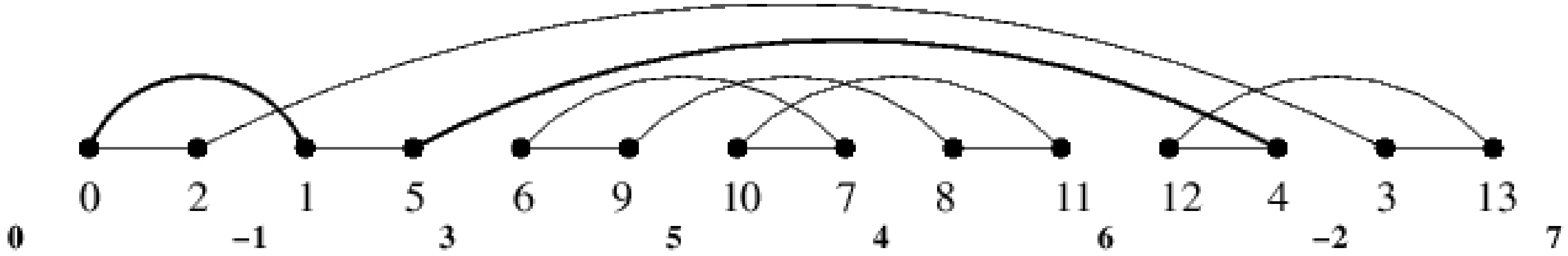


Fig. 1. The Breakpoint Graph of $\pi = (0\ -1\ 3\ 5\ 4\ 6\ -2\ 7)$

d(π) and cycle decomposition

- Every vertex in our graph is incident on exactly one black and one gray edge
- Alternating cycle decomposition
- $c(\pi)$ is the maximal number of cycles
- Claim 5: $c(\pi\rho) - c(\pi) \leq 1$
- Corollary: $d(\pi) \geq n+1- c(\pi)$
- Proof (of Claim 5): case inspection

Oriented pairs

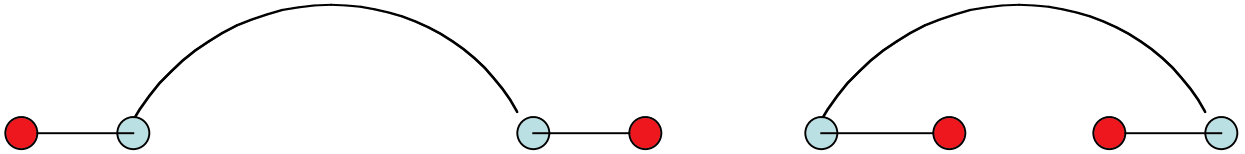
- A pair of consecutive integers (π_i, π_j) with opposite signs is an *oriented pair (*Bergeron 01)
- In π=3 1 6 5 -2 4 7: (1,-2) and (3,-2)
- Oriented pairs induce a naturally associated *oriented reversals* which create strips
- For example, the pair (1,-2) induces

  $3\ 1\ \underline{6\ 5\ \text{-}2}\ 4\ 7$ --> 3 1 2 -5 -6 4 7
- More generally, either $\rho(i,j-1)$ or $\rho(i+1,j)$

Oriented pairs and the bp graph

- Oriented pairs (≡ oriented gray edge)

- Unoriented pairs (≡ unoriented gray edge)



Oriented pairs and c(π)

- Claim 6: for an oriented reversal ρ,
$$c(\pi\rho) - c(\pi) = 1$$
- Proof: HW
- Corollary: If a sequence of *d* oriented reversals takes us from π to *id* then d (π) = *d*

Elementary reversal algorithm

- The *score* of an oriented reversal ρ is the number of oriented pairs in $\pi\ \rho$
- Bergeron (01): As long as π has an oriented pair choose the reversal with maximal score
- Stops at a positive permutation ω
- If ω = *id* then we are done, otherwise need to remove “hurdles” but ***not*** to backtrack