Transforming human into mice

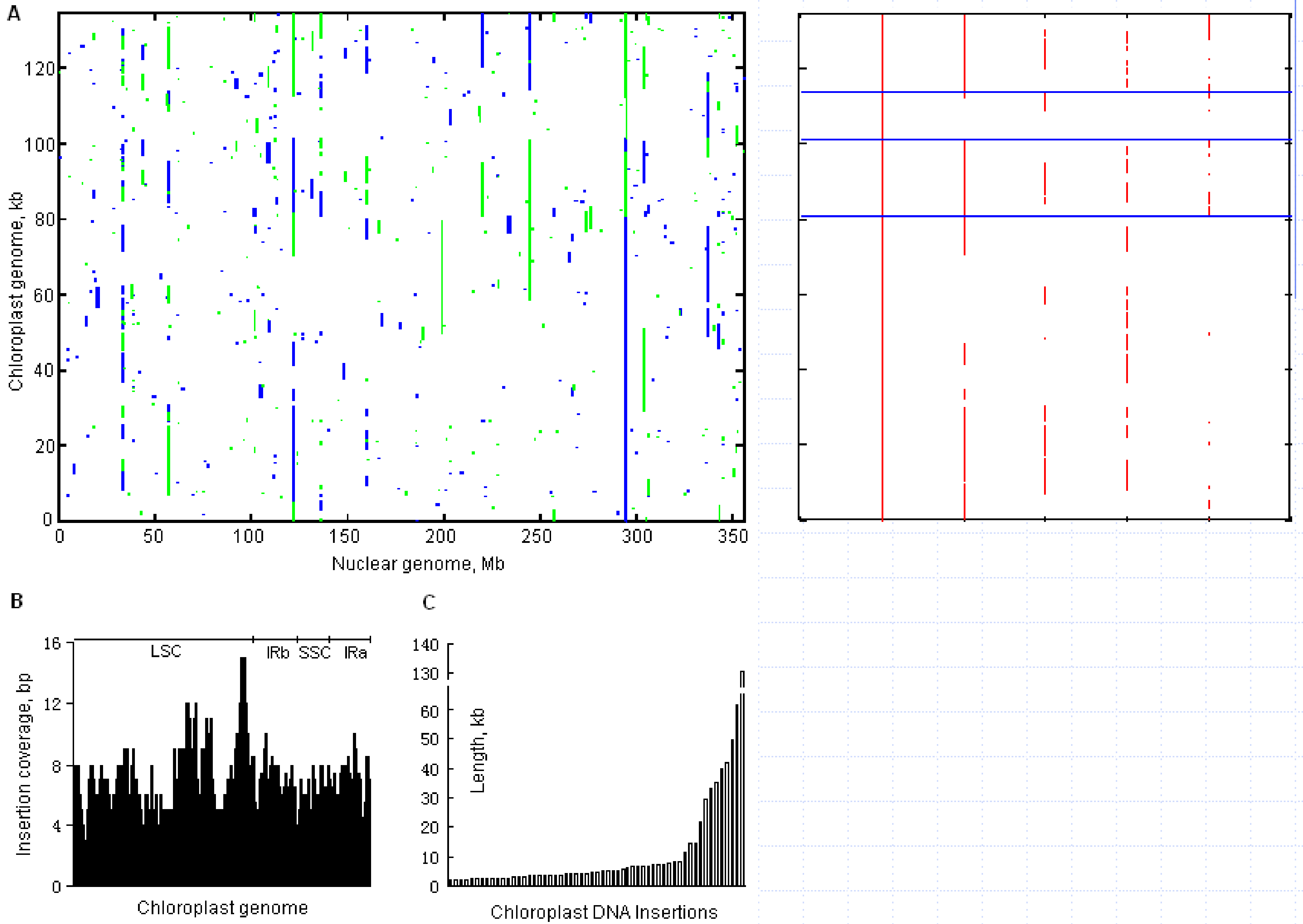
-------*Genome rearrangement*

1. A biological question
2. Operations and distances
3. The Hannenhalli-Pevzner Theory

Topic 5

1. A biological question

- Chloroplast DNA insertions in rice nuclear genome: we were confused at first

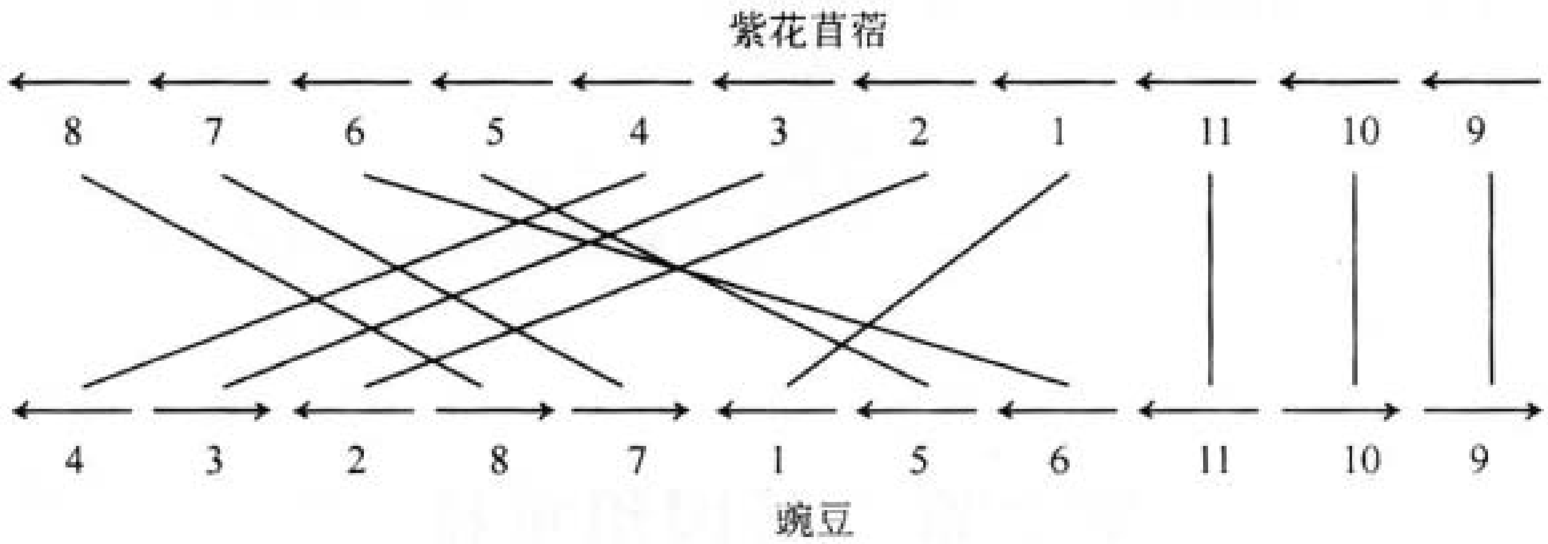

A
Chloroplast genome, kb
0
20
40
60
80
100
120
0
50
100
150
200
250
300
350
Nuclear genome, Mb
B
Insertion coverage, bp
16
12
8
4
0
LSC
IRb
SSC
IRa
Chloroplast genome
C
Length, kb
140
130
60
50
40
30
20
10
0
Chloroplast DNA Insertions

26 genes   Reversal Distance: 10

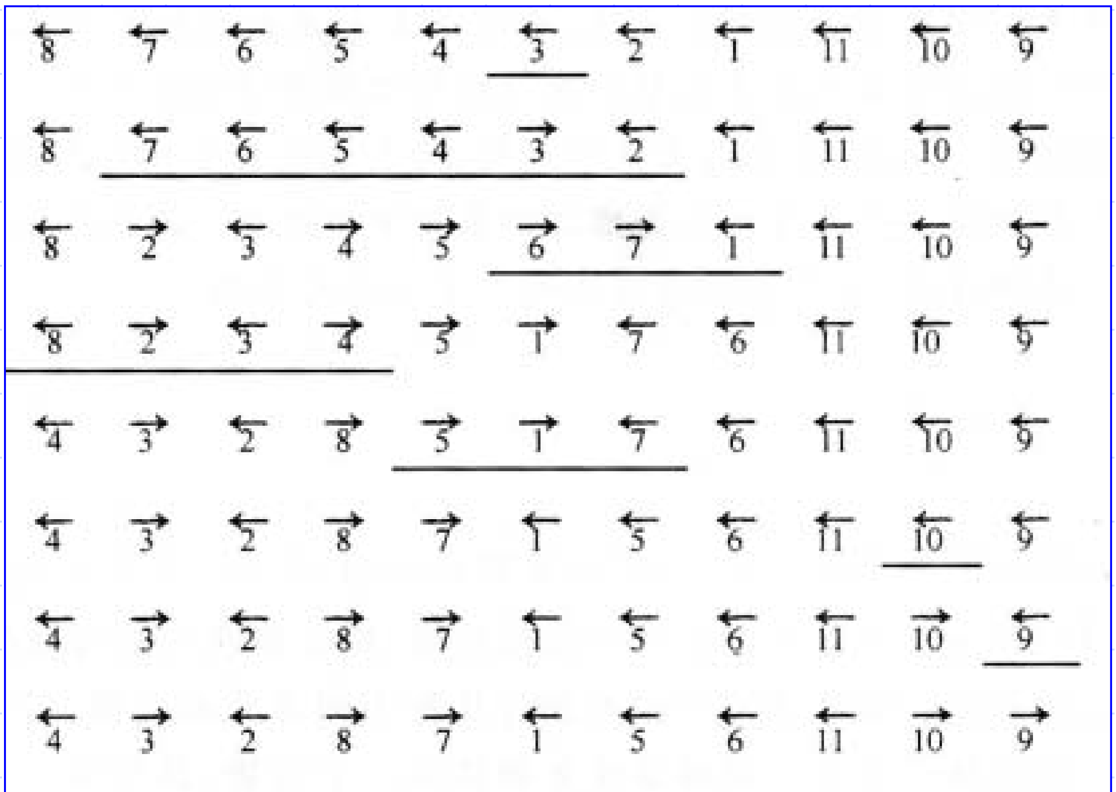
One optimal reversal scenario

Step	Description																										
0	(Source)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
1	Reversal	1	2	3	4	-5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
2	Reversal	1	2	3	4	-5	-6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
3	Reversal	1	2	3	4	-5	-6	7	8	9	10	11	-12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
4	Reversal	1	2	3	4	-5	-6	7	8	9	10	11	-12	-13	14	15	16	17	18	19	20	21	22	23	24	25	26
5	Reversal	1	2	3	4	-5	-6	7	8	-11	-10	-9	-12	-13	14	15	16	17	18	19	20	21	22	23	24	25	26
6	Reversal	1	2	3	4	-5	-6	7	-16	-15	-14	13	12	9	10	11	-8	17	18	19	20	21	22	23	24	25	26
7	Reversal	1	2	3	4	14	15	16	-7	6	5	13	12	9	10	11	-8	17	18	19	20	21	22	23	24	25	26
8	Reversal	1	2	3	4	14	15	16	-7	6	5	13	12	-19	-18	-17	8	-11	-10	-9	20	21	22	23	24	25	26
9	Reversal	1	2	3	4	14	15	16	-7	6	5	-26	-25	-24	-23	-22	-21	-20	9	10	11	-8	17	18	19	-12	-13
10	Reversal (Destination)	1	2	3	4	14	15	16	20	21	22	23	24	25	26	-5	-6	7	9	10	11	-8	17	18	19	-12	-13

Evolutionary significance of inversions in legume chloroplast DNAs. *Palmer et al, Current Genetics, 1988, 14:65-74*



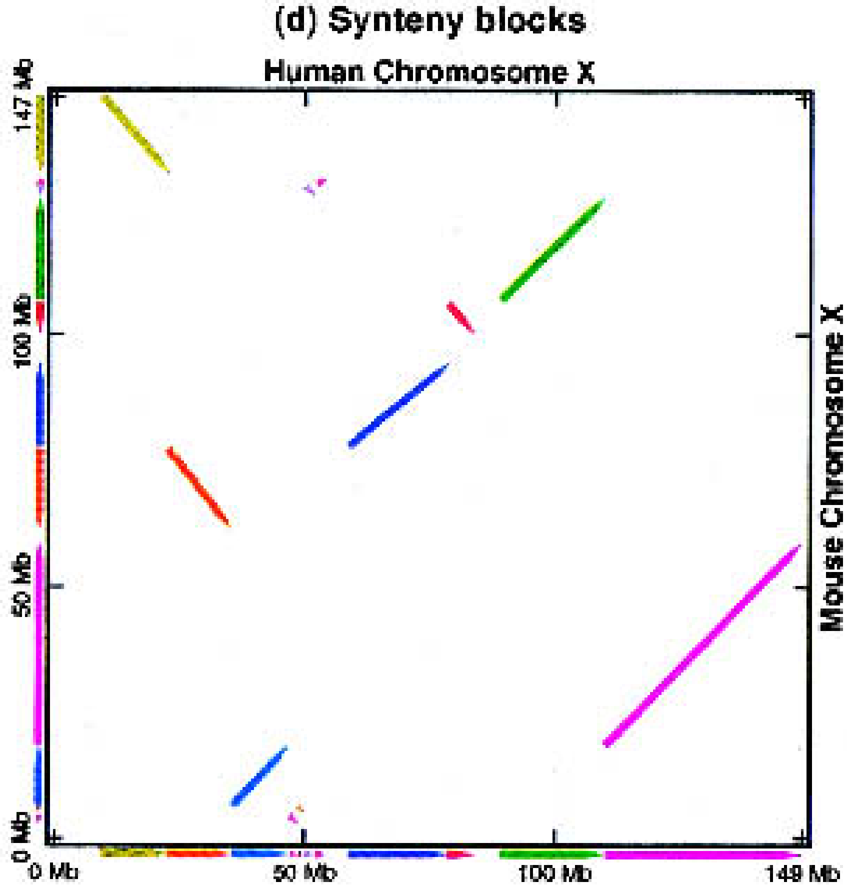
Another solution

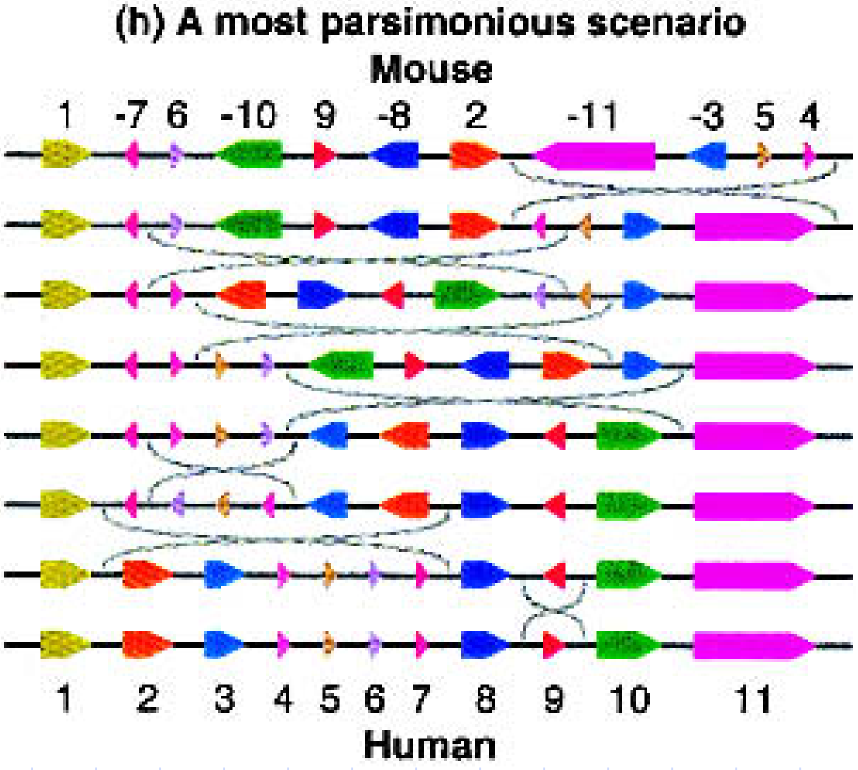


An important conclusion

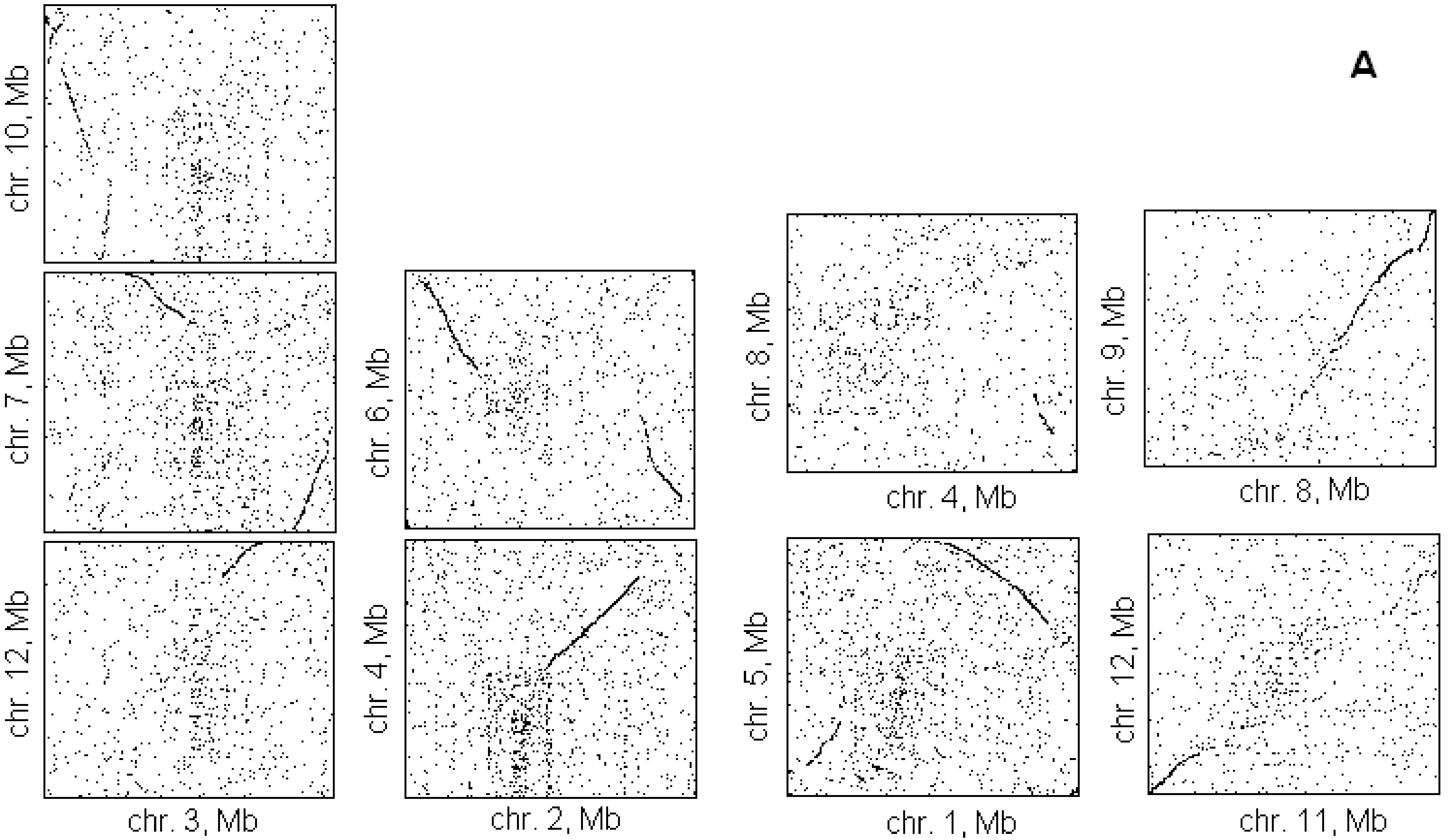
- Evolutionary significance of inversions in legume chloroplast DNAs. *Palmer et al, Current Genetics, 1988, 14:65-74*
- Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Palmer, J.D. and Herbon, L.A. 1988. J. Mol. Evol.* ***27:*** *87-97.*

Genome rearrangements in mammalian evolution: lessons from human and mouse genomes. *Pevzner and Tesler, Genome Research, 2003, 13:37-45*





Rice *(Zhang et al. in press)*



Q: A most parsimonious rearrangement scenario

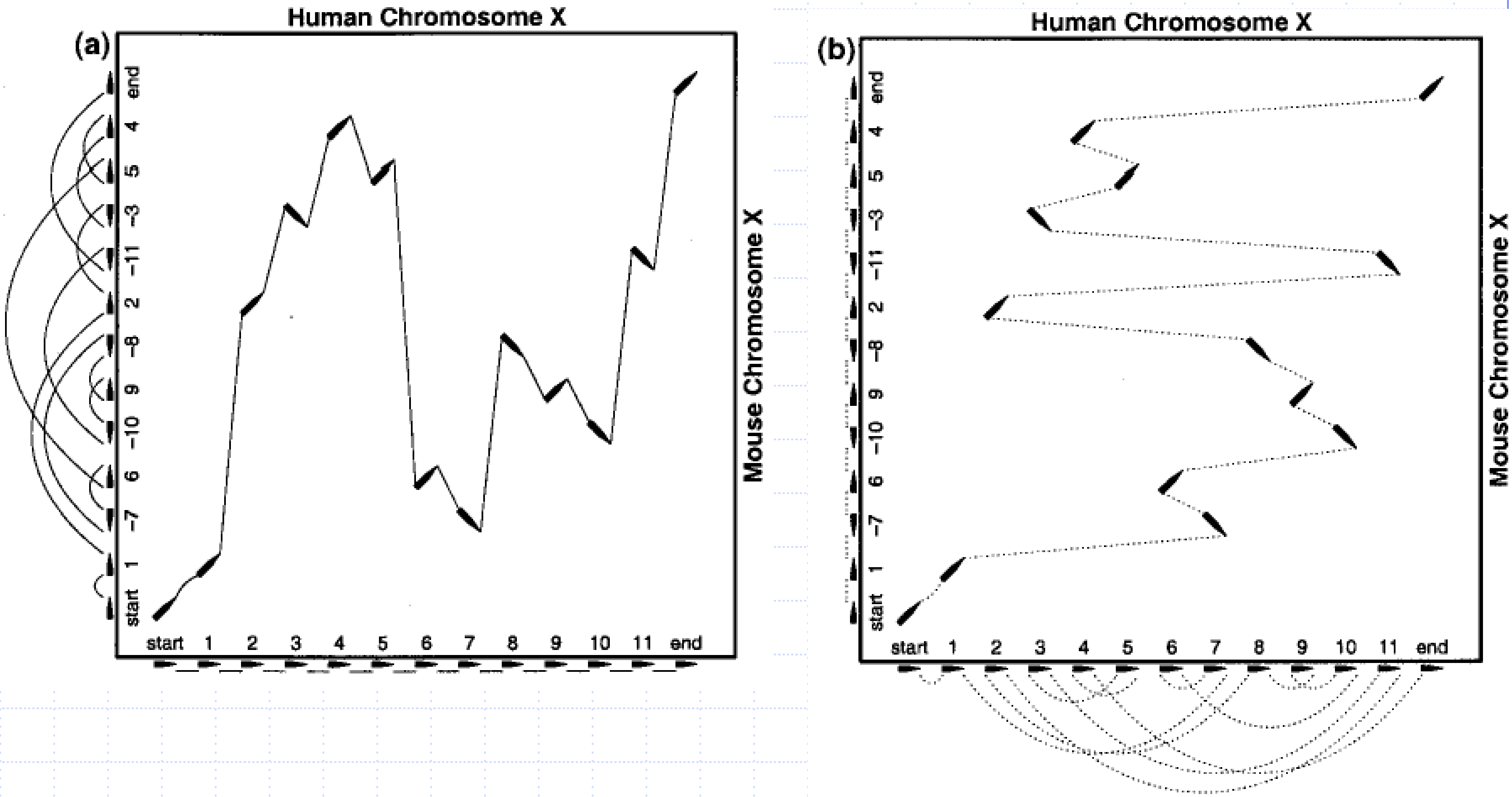
- Parsimony assumption: We believe nature can always find the shortest way

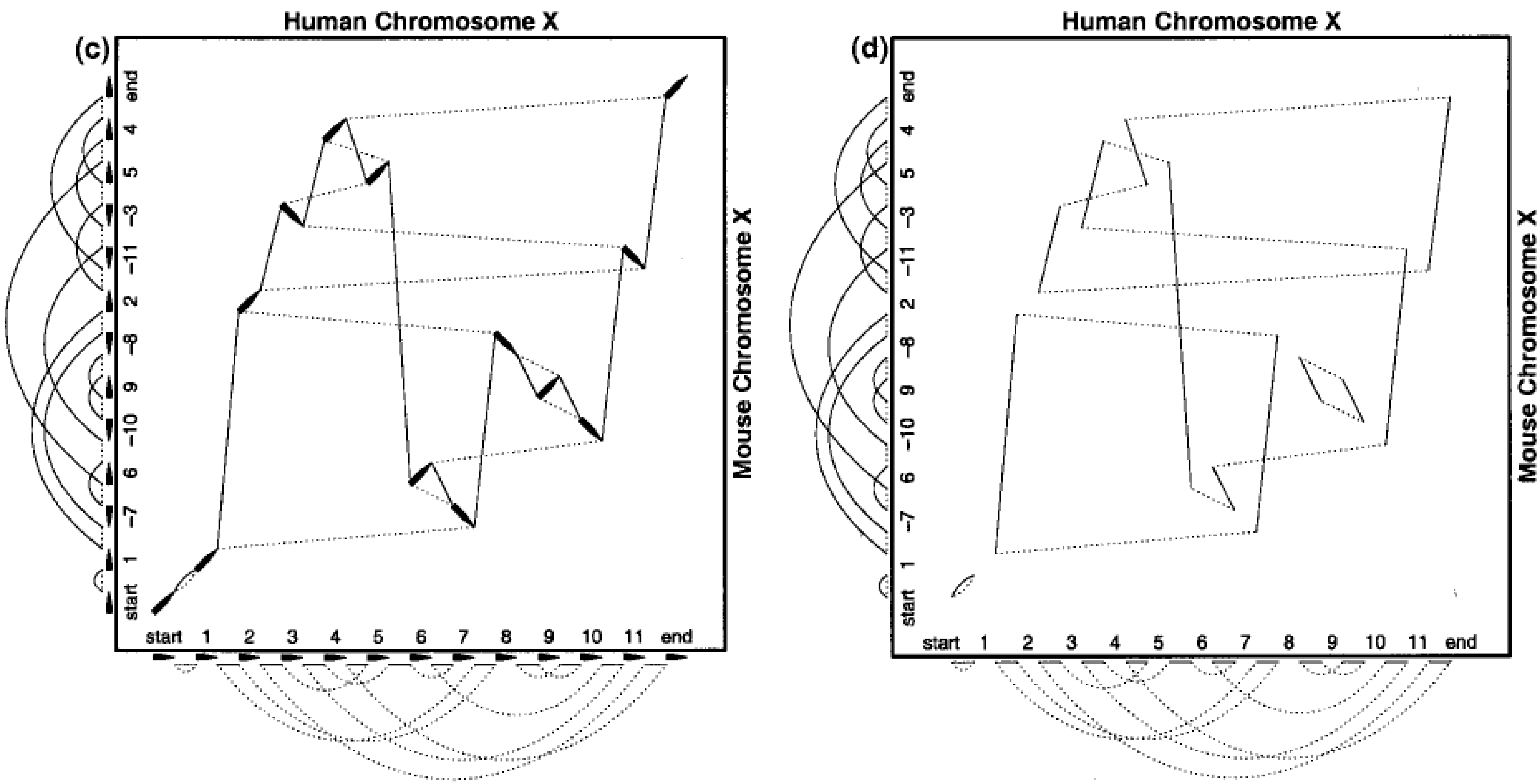
2. Operations and distances

- A brief history: genome rearrangements in molecular evolution (*Dobzhansky and Sturtevant, 1938*); The shortest rearrangment scenario for unichromosomal genome (*Palmer et al, 1988*); Notion of a breakpoint (disruption of gene order) (*Watterson et al, 1982; Nadeau and Taylor, 1984; Sturtevant and Dobzhansky, 1936*); Breakpoint graph (*Bafna and Pevzner, 1993, 1996*); Tansforming one genome into another---a polynomial algorithm (*Hannenhalli and Pevzner, 1995*)

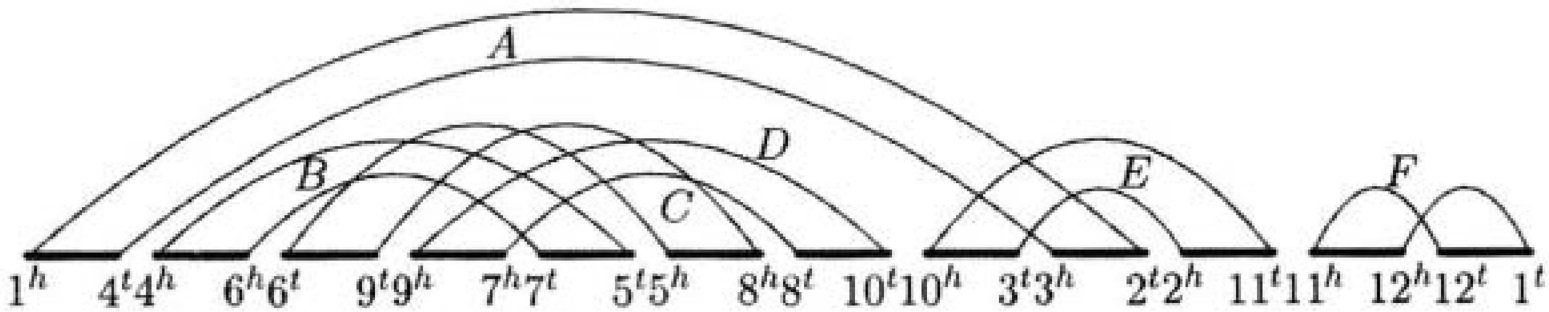
- Edit distance
  - Reversal distance
  - Transposition distance
  - Translocation distance
  - genomic distance: reversal, translocation, fusion and fission for multichromosomal genomes
- Breakpoint

Construction of the breakpoint graph from synteny blocks *(Pevzner and Tesler, 2003)*





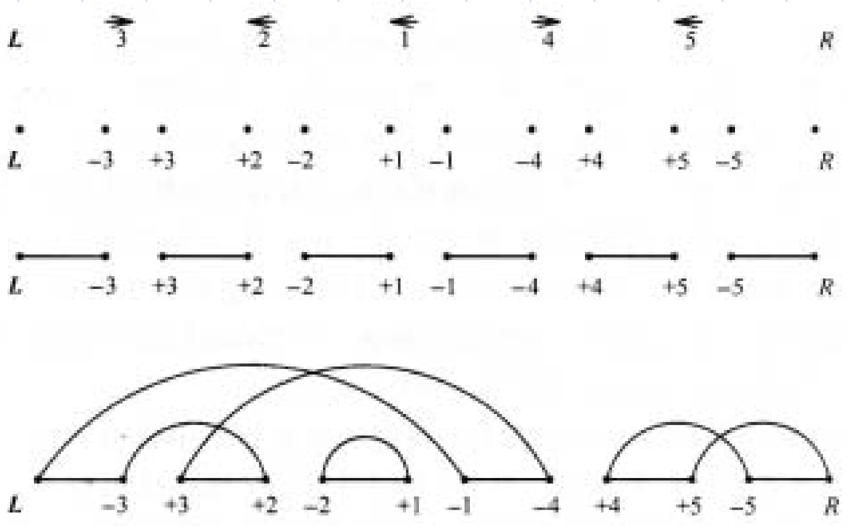
(c)
Human Chromosome X
Mouse Chromosome X
start
1
2
3
4
5
6
7
8
9
10
11
end
-7
6
-10
9
-8
2
-11
-3
5
4
(d)
Human Chromosome X
Mouse Chromosome X

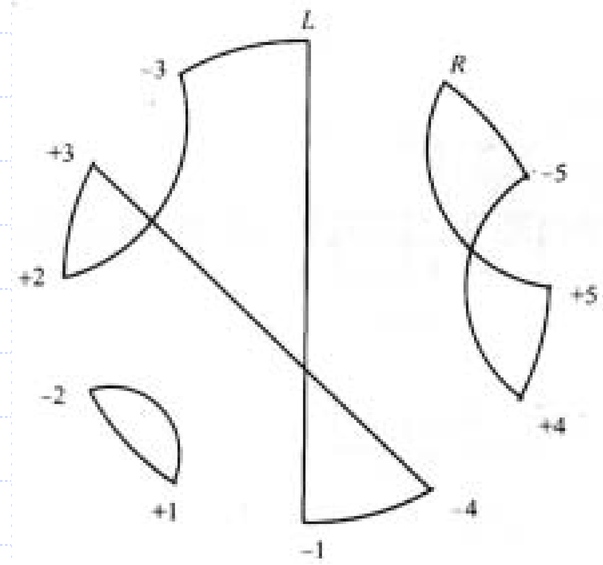


A
B
C
D
E
F
1^h 4^t4^h 6^h6^t 9^t9^h 7^h7^t 5^t5^h 8^h8^t 10^t10^h 3^t3^h 2^t2^h 11^t11^h 12^h12^t 1^t

Reality-desire diagram

- (Setubal and Meidanis, 1997)
- (Sankoff and El-Mabrouk, 2002)





An alternative description for the parsimony question

- Number of cycles (*c*) is maximized when *H1=H2*, in which case each cycle has one black edge and one gray edge
- How to maximize a RD diagram using the least operations?

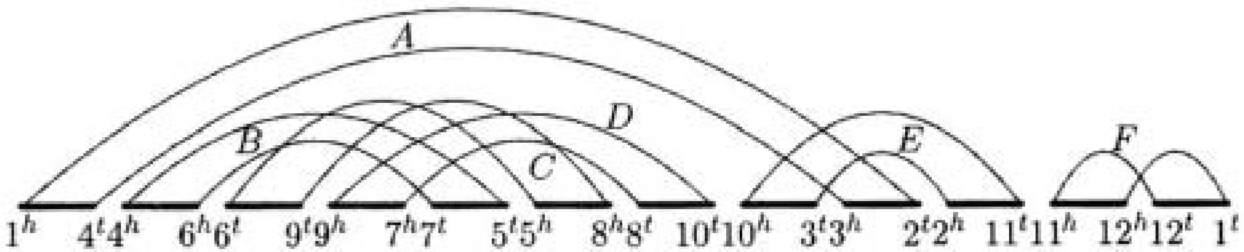
3. The Hannenhalli-Pevzner Theory

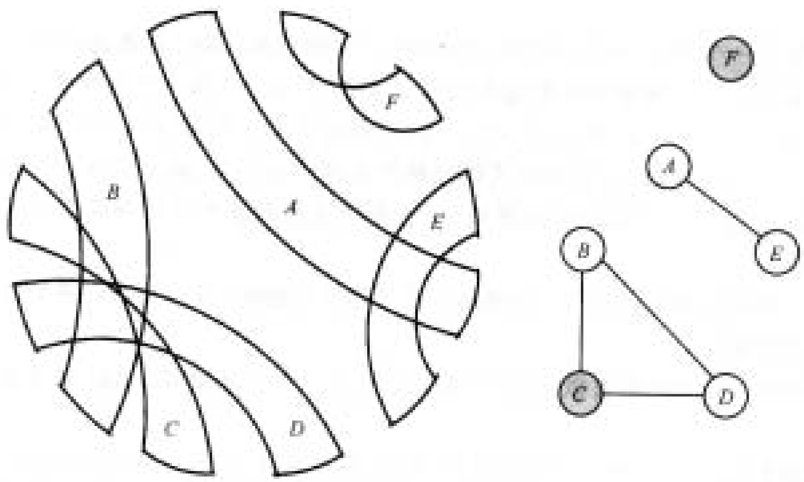
- Polynomial-time algorithms for calculating the edit distance between two genomes
- The Hannenhalli-Pevzner formula:

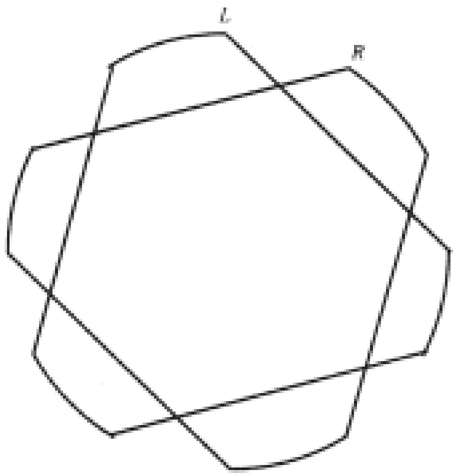
$$d(H_1, H_2) = n(\mathcal{G}) - c(\mathcal{G}) + m(\mathcal{G}) + f(\mathcal{G})$$

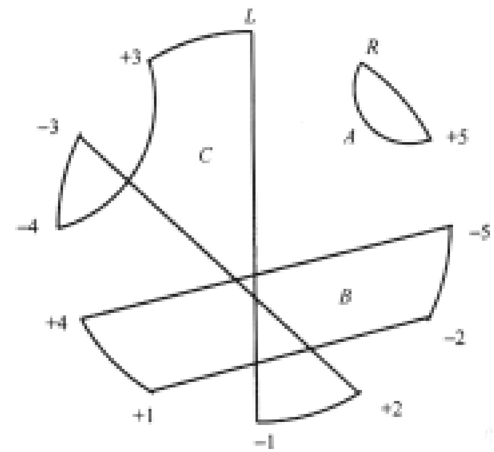
where $d(H_1, H_2)$ is the minimum number of rearrangement operations (reversals and/or translocations), $n(\mathcal{G})$ is the number of black edges of $\mathcal{G}$, $c(\mathcal{G})$ is the number of cycles, $m(\mathcal{G})$ is the number of bad components, and $f(\mathcal{G})$ is a correction of size 0, 1, or 2 depending on the set of bad components.

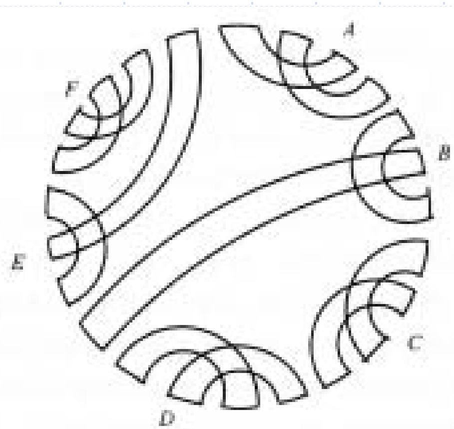
- Size of a cycle: the number of black edges it contains, for example, size 1.
- A rearrangement operation: proper, improper or bad. Difference between the number of cycles before and after applying the operation: 1, 0, -1
- A component: good or bad. A component is termed good if can be transformed to a set of cycles of size 1 by a series of proper operations, and bad otherwise.

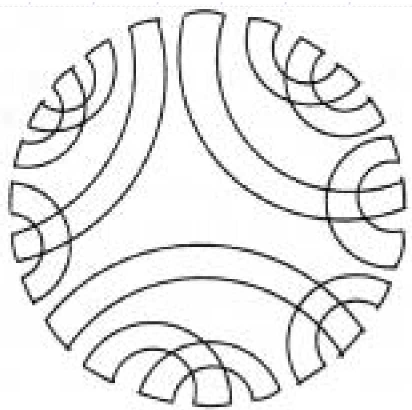

A
B
D
C
E
F
1^h 4^t4^h 6^h6^t 9^t9^h 7^h7^t 5^t5^h 8^h8^t 10^t10^h 3^t3^h 2^t2^h 11^t11^h 12^h12^t 1^t


F
B
A
E
C
D

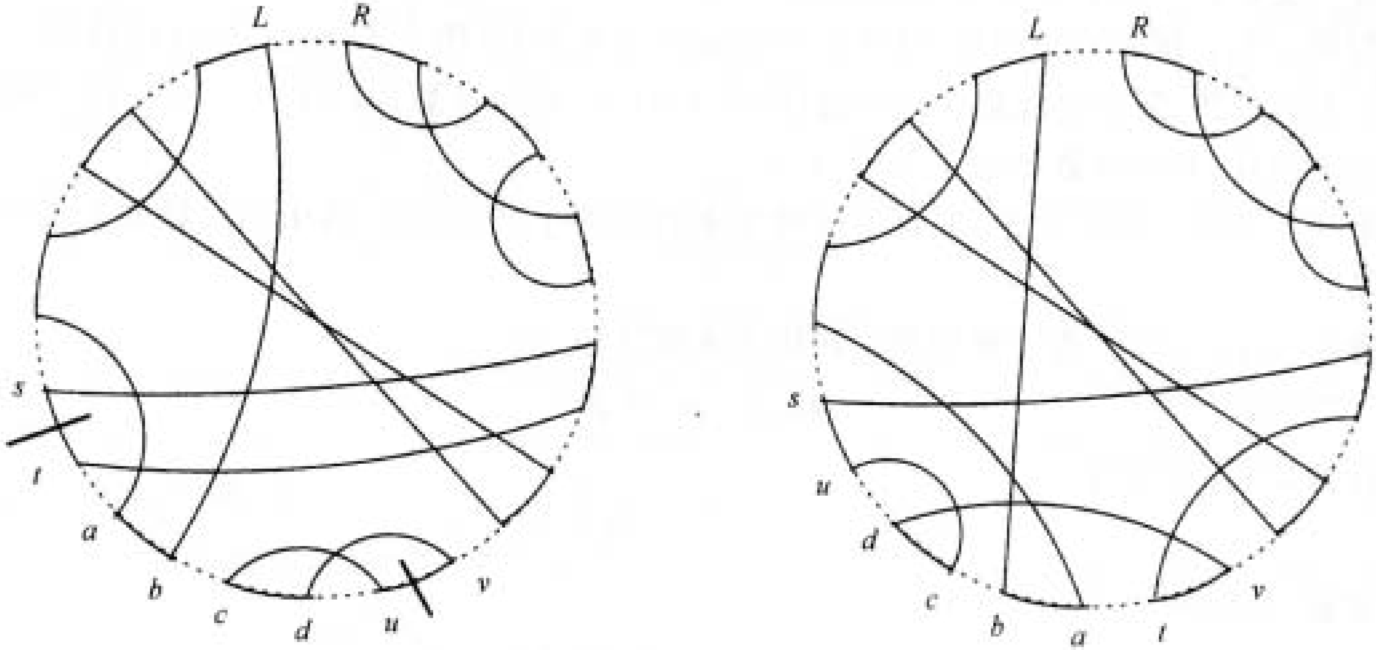

L
R


L
R
+3
−3
C
A
+5
−4
−5
B
+4
−2
+1
+2
−1


A
F
B
E
C
D



Example



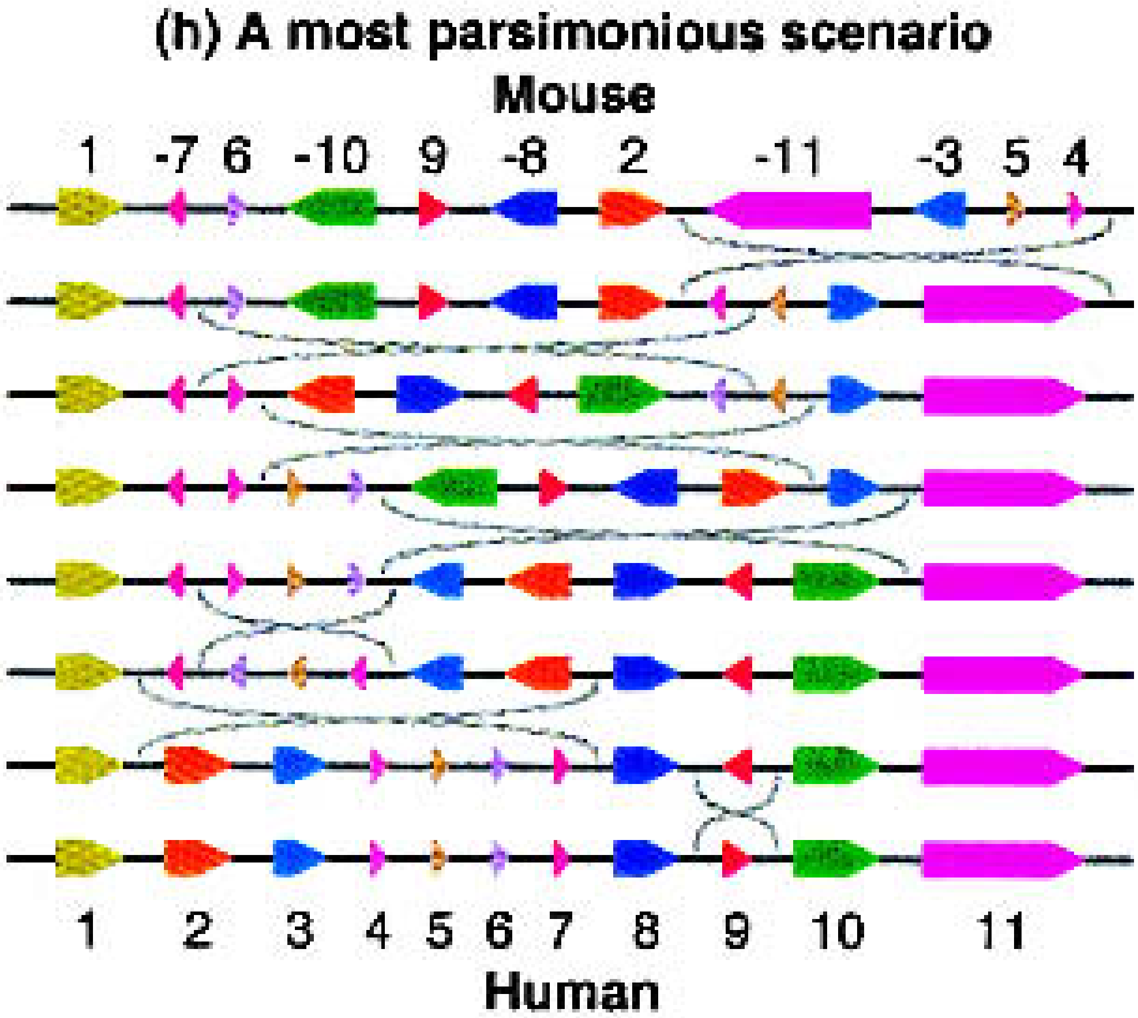
How many steps do they have to go?

26 genes   Reversal Distance: 10

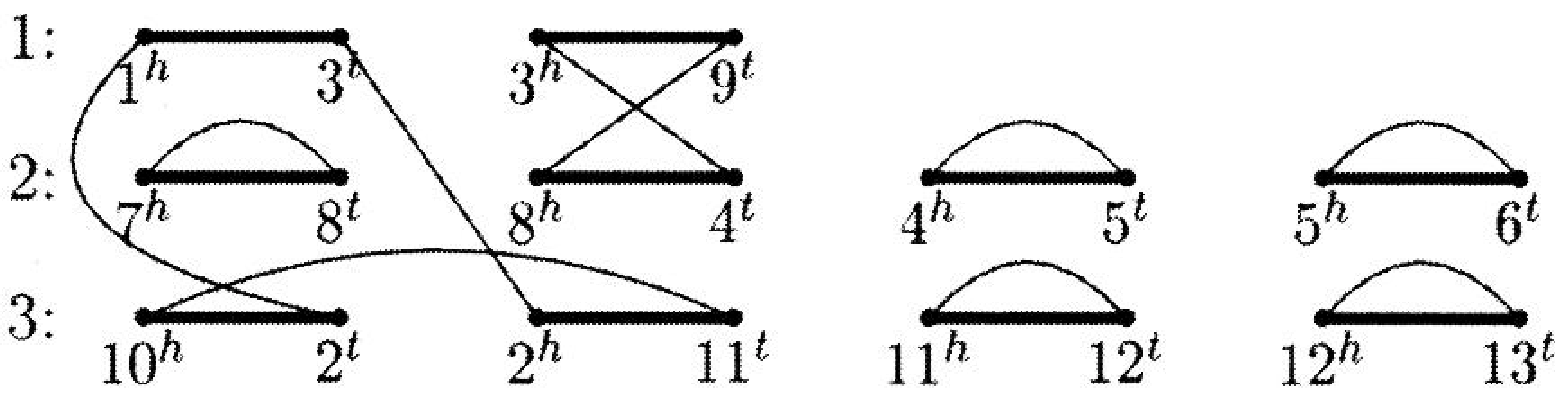
One optimal reversal scenario

Step	Description	
0	(Source)	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
1	Reversal	1 2 3 4 -5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
2	Reversal	1 2 3 4 -5 -6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
3	Reversal	1 2 3 4 -5 -6 7 8 9 10 11 -12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
4	Reversal	1 2 3 4 -5 -6 7 8 9 10 11 -12 -13 14 15 16 17 18 19 20 21 22 23 24 25 26
5	Reversal	1 2 3 4 -5 -6 7 8 -11 -10 -9 -12 -13 14 15 16 17 18 19 20 21 22 23 24 25 26
6	Reversal	1 2 3 4 -5 -6 7 -16 -15 -14 13 12 9 10 11 -8 17 18 19 20 21 22 23 24 25 26
7	Reversal	1 2 3 4 14 15 16 -7 6 5 13 12 9 10 11 -8 17 18 19 20 21 22 23 24 25 26
8	Reversal	1 2 3 4 14 15 16 -7 6 5 13 12 -19 -18 -17 8 -11 -10 -9 20 21 22 23 24 25 26
9	Reversal	1 2 3 4 14 15 16 -7 6 5 -26 -25 -24 -23 -22 -21 -20 9 10 11 -8 17 18 19 -12 -13
10	Reversal (Destination)	1 2 3 4 14 15 16 20 21 22 23 24 25 26 -5 -6 7 9 10 11 -8 17 18 19 -12 -13

Program *GRIMM* (Genome Rearrangements In Man and Mouse), after Brothers Grimm(1812)


(h) A most parsimonious scenario
Mouse
1 -7 6 -10 9 -8 2 -11 -3 5 4
1 2 3 4 5 6 7 8 9 10 11
Human

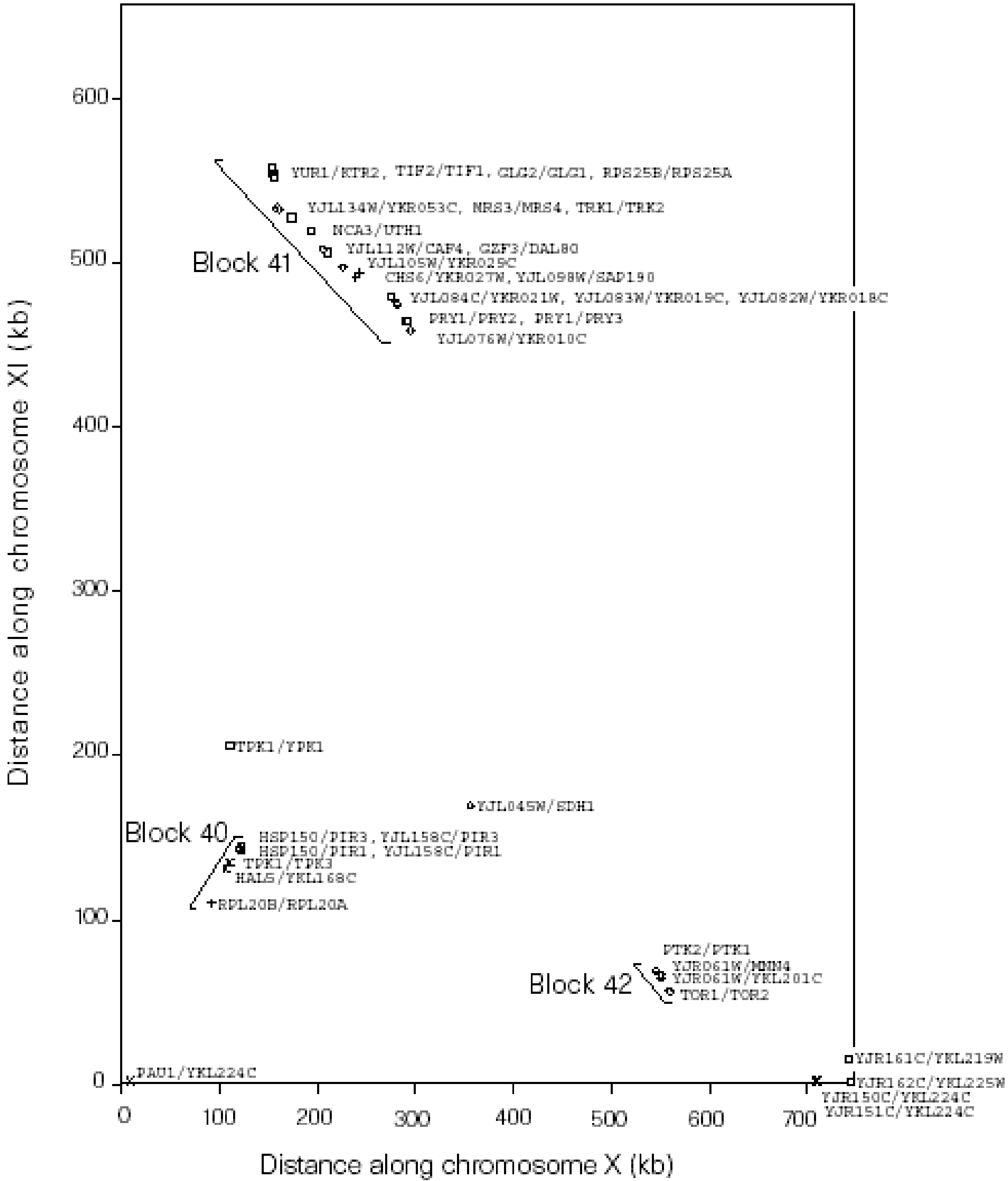
Genomes with three chromosomes



Genome doubling

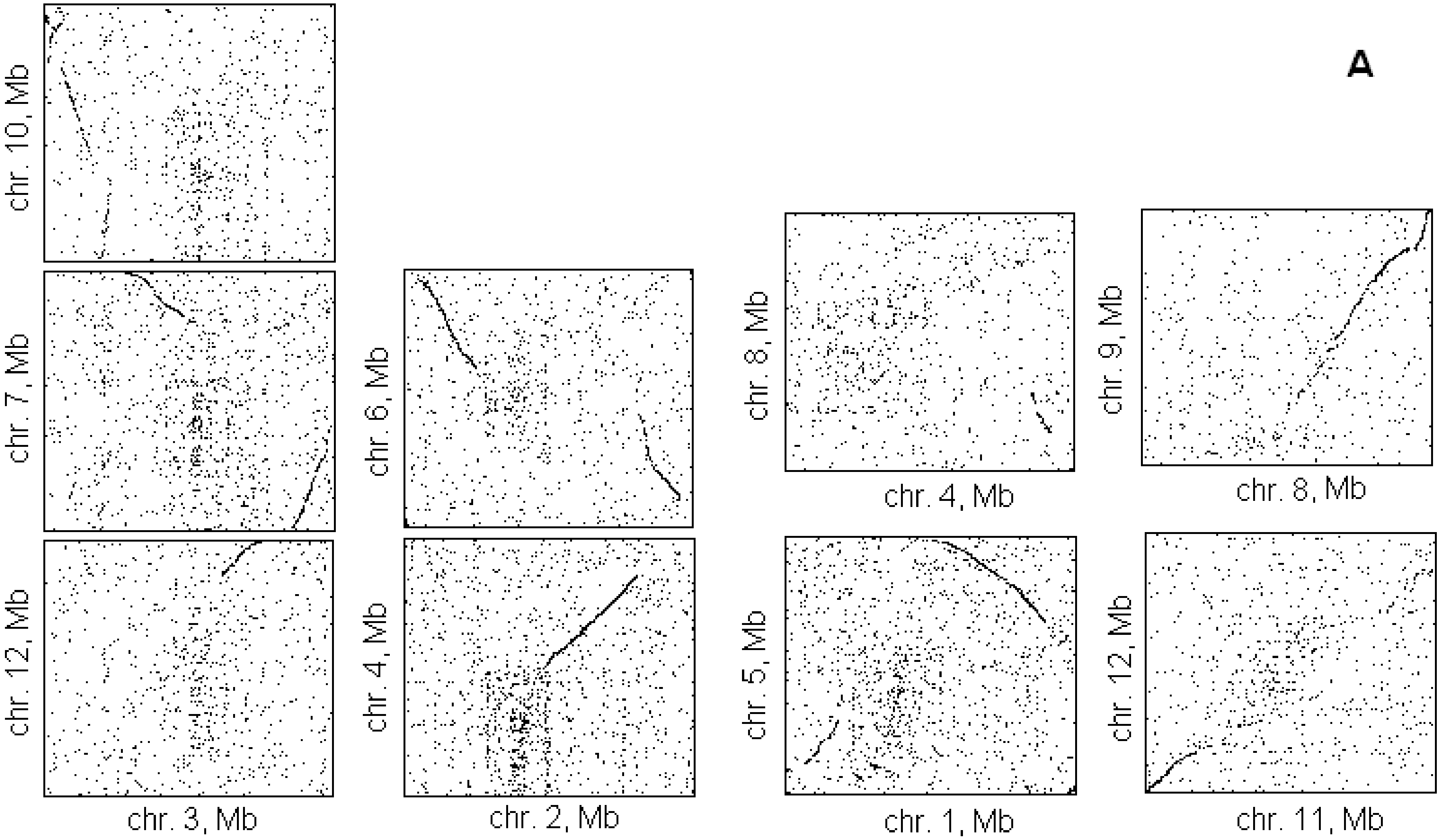
- How can we reconstruct some or most of the original gene order at the time of genome duplication or hybridization, based on traces conserved in the ordering of those duplicate genes still identifiable?
- Examples: yeast (Wolfe et al, 1997) and rice (Paterson et al, 2003, 2004)

◆Yeast

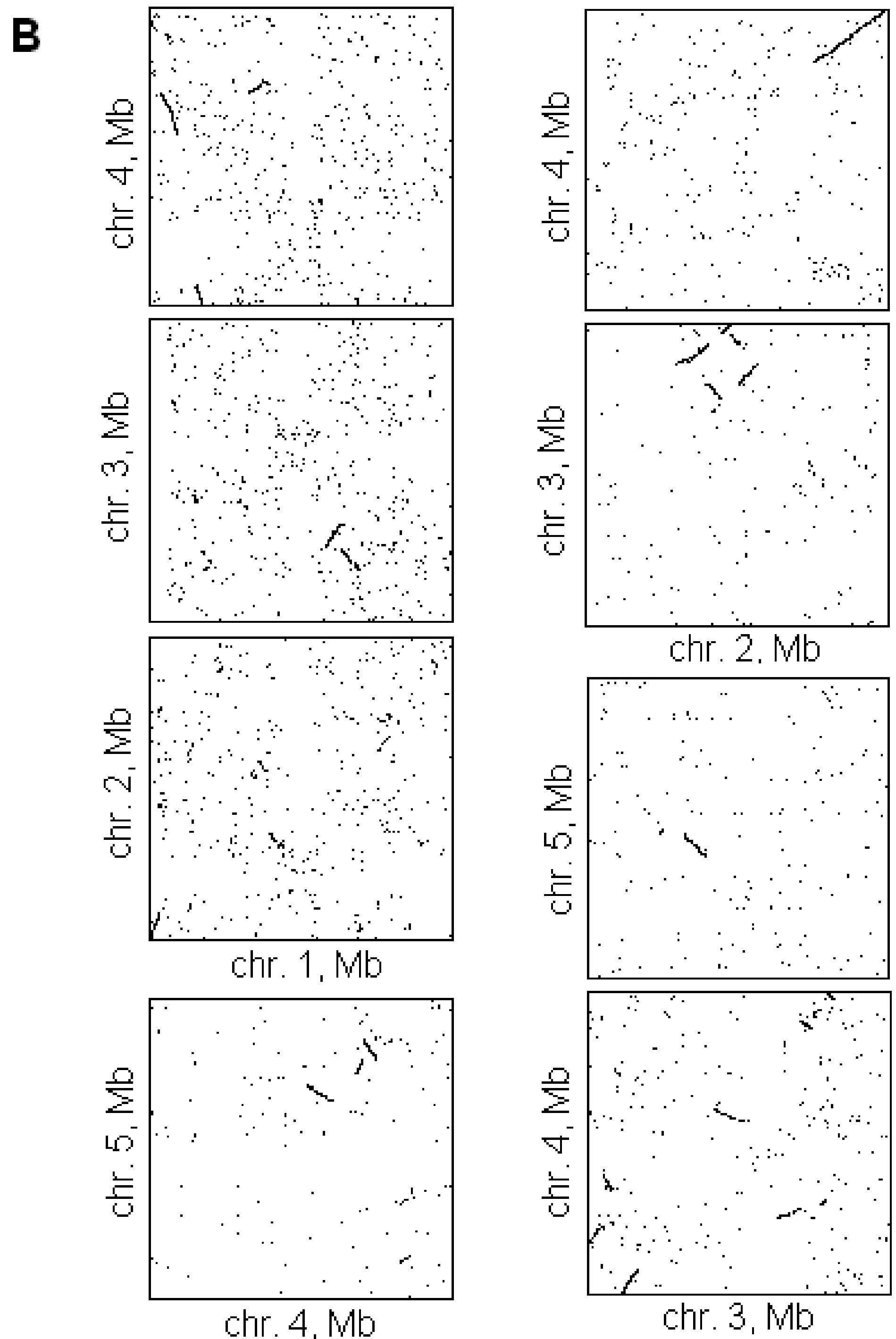


Block 41
YUR1/KTR2, TIF2/TIF1, GLG2/GLG1, RPS25B/RPS25A
YJL134W/YKR053C, MRS3/MRS4, TRK1/TRK2
NCA3/UTH1
YJL112W/CAF4, GZF3/DAL80
YJL105W/YKR029C
CHS6/YKR027W, YJL098W/SAP190
YJL084C/YKR021W, YJL083W/YKR019C, YJL082W/YKR018C
PRY1/PRY2, PRY1/PRY3
YJL076W/YKR010C
TPK1/YPK1
YJL045W/SDH1
Block 40
HSP150/PIR3, YJL158C/PIR3
HSP150/PIR1, YJL158C/PIR1
TPK1/TPK3
HAL5/YKL168C
RPL20B/RPL20A
PTK2/PTK1
YJR061W/MNN4
YJR061W/YKL201C
Block 42
TOR1/TOR2
YJR161C/YKL219W
PAU1/YKL224C
YJR162C/YKL225W
YJR150C/YKL224C
YJR151C/YKL224C
Distance along chromosome XI (kb)
Distance along chromosome X (kb)

◆Rice

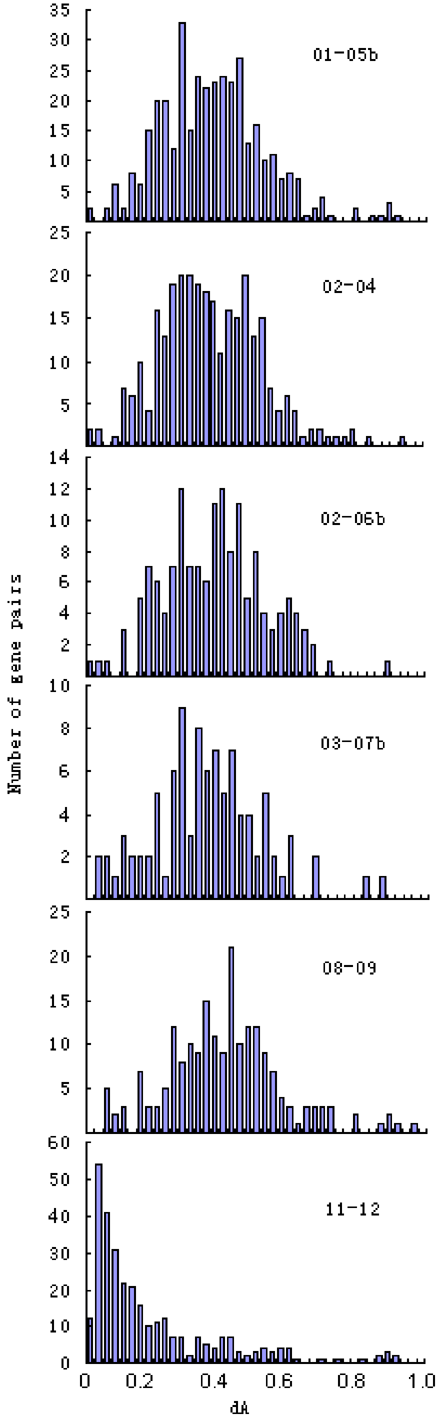

A
chr. 10, Mb
chr. 7, Mb
chr. 12, Mb
chr. 3, Mb
chr. 6, Mb
chr. 4, Mb
chr. 2, Mb
chr. 8, Mb
chr. 4, Mb
chr. 5, Mb
chr. 1, Mb
chr. 9, Mb
chr. 8, Mb
chr. 12, Mb
chr. 11, Mb

Arabidopsis



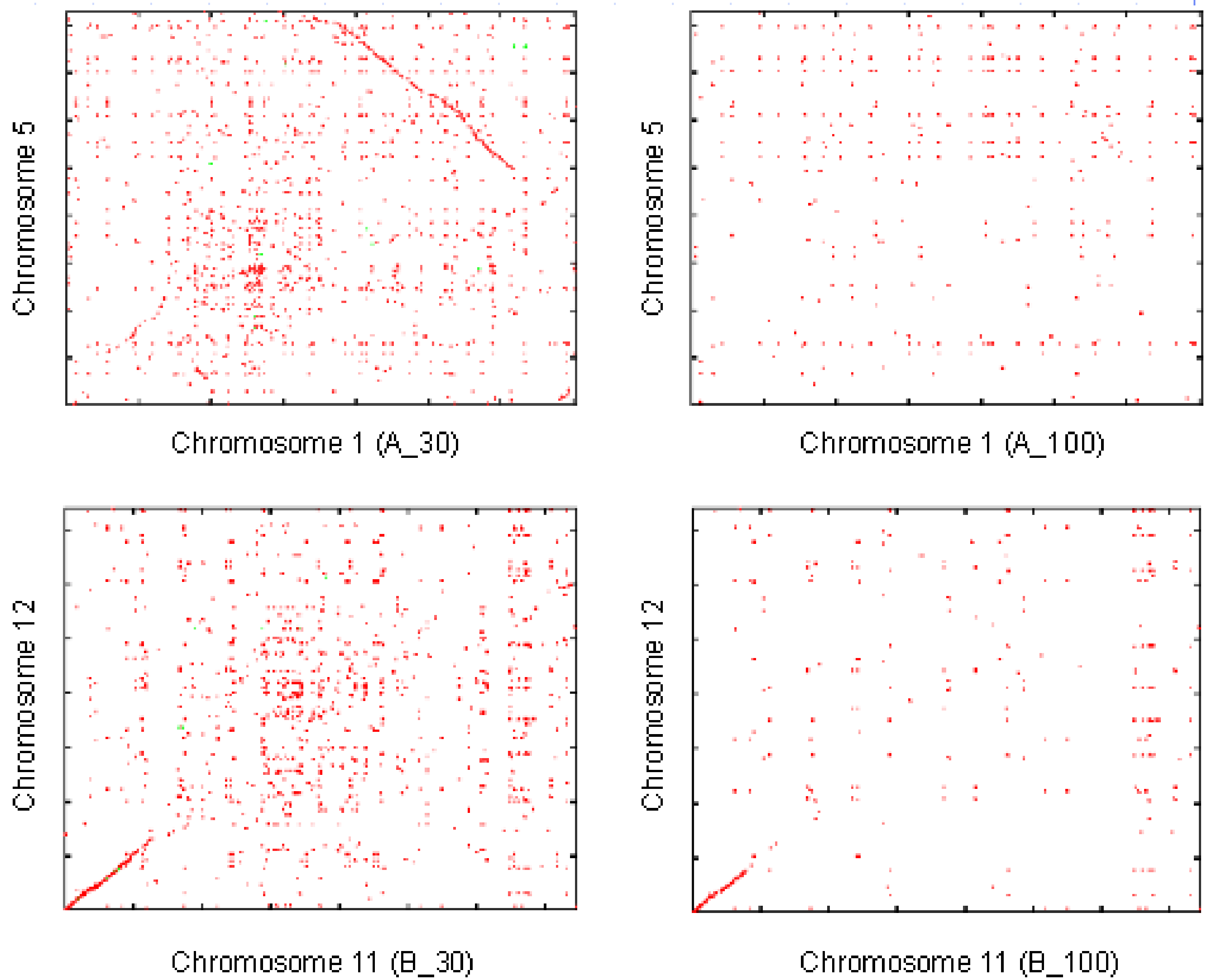
◆Timing (rice)

Age distribution of duplicate genes



◆Timing (rice)

Chromosomal alignment (MUMer)



Exercise

- ?