

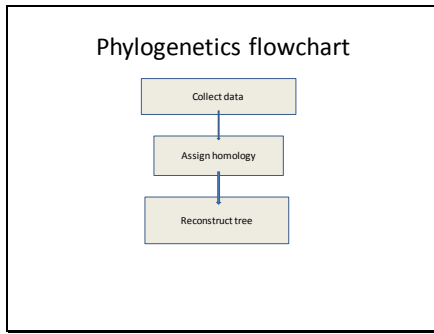
Slide 1

Optimization alignment

tree reconstruction from
unhomologized data

Rasmus Hovmoller

Slide 2

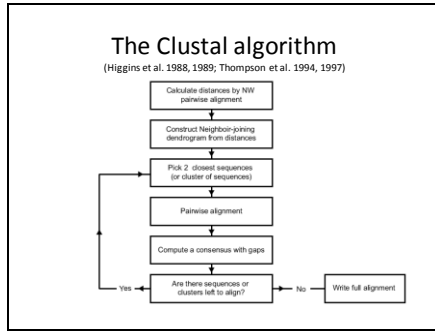


Slide 3

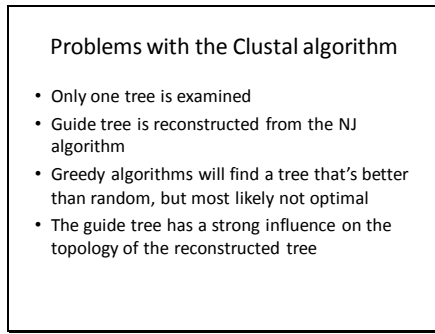
Static sequence alignment

- N-P hard problem
- How do deal with ambiguities?
- Optimality in alignments
 - how to prefer one alignment over another

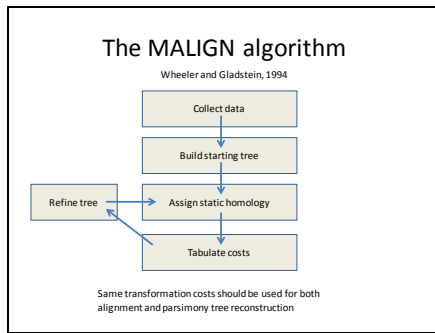
Slide 4



Slide 5



Slide 6






Slide 7

Calculating tree cost

- Assign weights to each type of transformation
 - base substitution
 - gap insertion or deletion
- Given the data and a tree, how many transformations of each type are needed to explain the data
- Using parsimony, find the tree that requires the lowest total transformation costs

Slide 8

Malign tree costs

	A	B	C
Addition sequence			
Multiple alignments	T1: -TTC- T2: -TACG T3: TTGCG	T1: TT-C- T2: TA-CG T3: TTGCG	T1: TT-C- T2: TA-CG T3: TTGCG
Costs	Gaps : 2 Subst: 2 Total: 22	Gaps : 2 Subst: 1 Total: 21	Gaps : 2 Subst: 1 Total: 21

Gap cost=10 Substitution cost= 1.

Slide 9

Why not cut out the middle-man?

- In MALIGN, the alignment costs are calculated directly from trees
- The optimal tree will have the lowest costs
- Why create a static alignment at all?

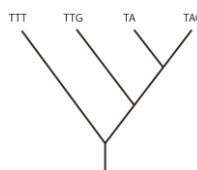
Slide 10

Bridging the gap

- Gaps do not exist in reality
- Gaps can not be inferred from data
- A gap is a placeholder for an indel event
- The pattern of gaps contain information about evolutionary events
- The information from indel data should be used to infer the tree

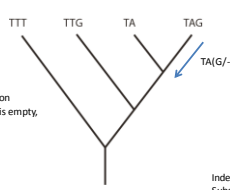
Slide 11

Optimization alignment



Slide 12

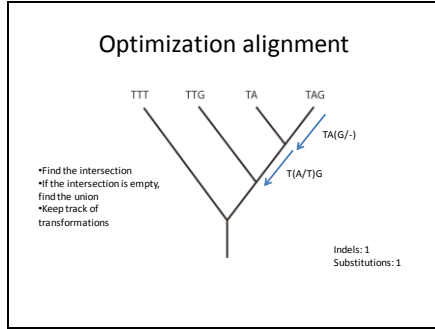
Optimization alignment



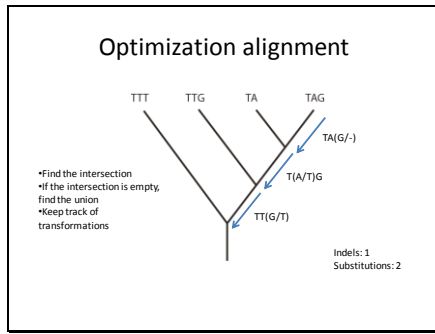
- Find the intersection
- If the intersection is empty, find the union
- Keep track of transformations

Indels: 1
Substitutions: 0

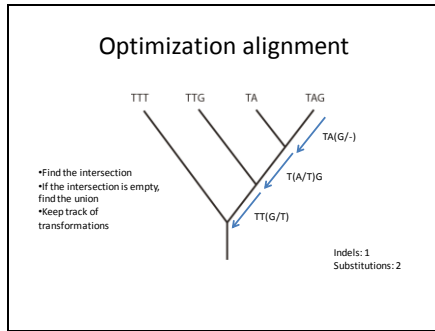
Slide 13



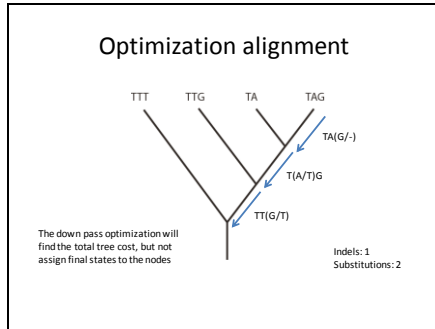
Slide 14



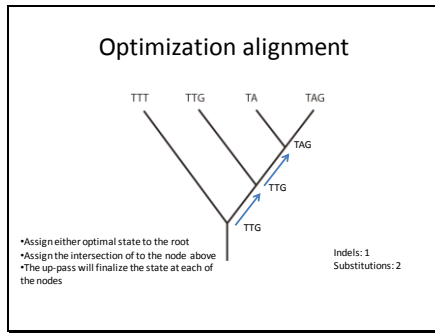
Slide 15



Slide 16



Slide 17



Slide 18

- ### Problems with optimization alignment
- Insertions and deletions are not treated as single events
 - Computationally demanding
 - No static alignment for alternative tree reconstruction
 - Does not use external information from secondary structure

Slide 19

Advantages of optimization alignment

- Explores tree-space
- Will find the lowest cost
- Highly parallelizable
- Any alphabet can be used
- The method is generalizable to all kinds of genetic transformations

Slide 20

Fixed states optimization

- Treat each sequence as a single character
- Calculate transformation costs for all pairs of sequences in the data
- Use the costs as a character state step matrix in a standard tree search
- No unobserved character states are assigned to the nodes

Slide 21

Dataset:
I: TTT
II: TTG
III: TA
IV: TAG

Transformation costs:
Indels: 2
Substitutions: 1

Pairwise transformations

	TTT	TTG	TA	TAG
TTT	-	1s	1s, 1i	2s
TTG	1s	-	1s, 1i	1s
TA	1s, 1i	1s, 1i	-	1i
TAG	2s	1s	1i	-

Slide 22

Dataset:
 I: TTT
 II: TTG
 III: TA
 IV: TAG

Transformation costs:
 Indels: 2
 Substitutions: 1

Pairwise transformations

	TTT	TTG	TA	TAG
TTT	-	1	3	2
TTG	1	-	3	1
TA	3	3	-	2
TAG	2	1	2	-

Slide 23

Optimization alignment

	TTT	TTG	TA	TAG
TTT	-	1	3	2
TTG	1	-	3	1
TA	3	3	-	2
TAG	2	1	2	-

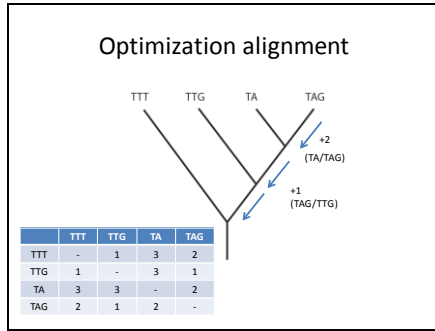
Slide 24

Optimization alignment

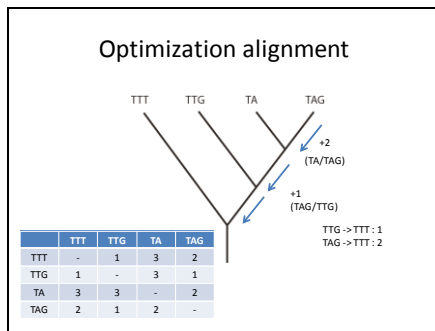
Pick the lowest cost transformation
 TA -> TTG : 3
 TAG -> TTG : 1

	TTT	TTG	TA	TAG
TTT	-	1	3	2
TTG	1	-	3	1
TA	3	3	-	2
TAG	2	1	2	-

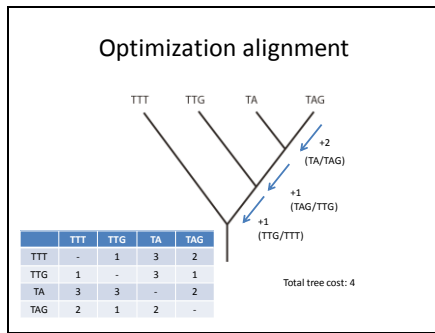
Slide 25



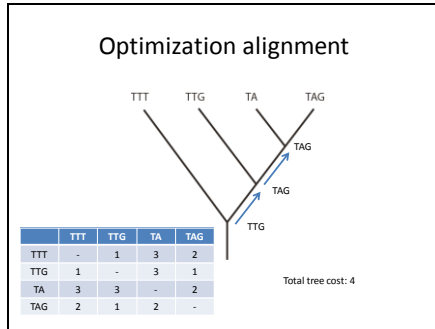
Slide 26



Slide 27



Slide 28



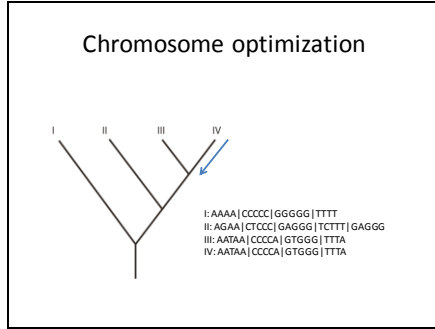
Slide 29

- Chromosome optimization**
- Optimization alignment can be generalized to include all kinds of transformation events
 - Substitution
 - Indels
 - Reversions
 - Duplications
 - Translocation

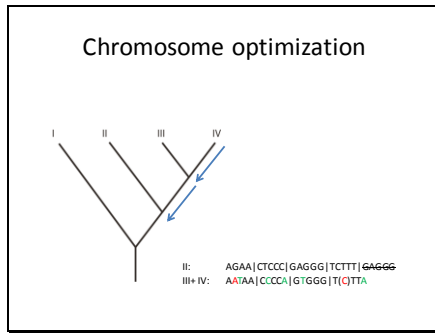
Slide 30

- Chromosome optimization**
- Assign cost for each type of transformation
 - At each node find the optimal set of transformations

Slide 31



Slide 32



Slide 33

