

Evolution of Tandemly Arrayed Genes in Multiple Species

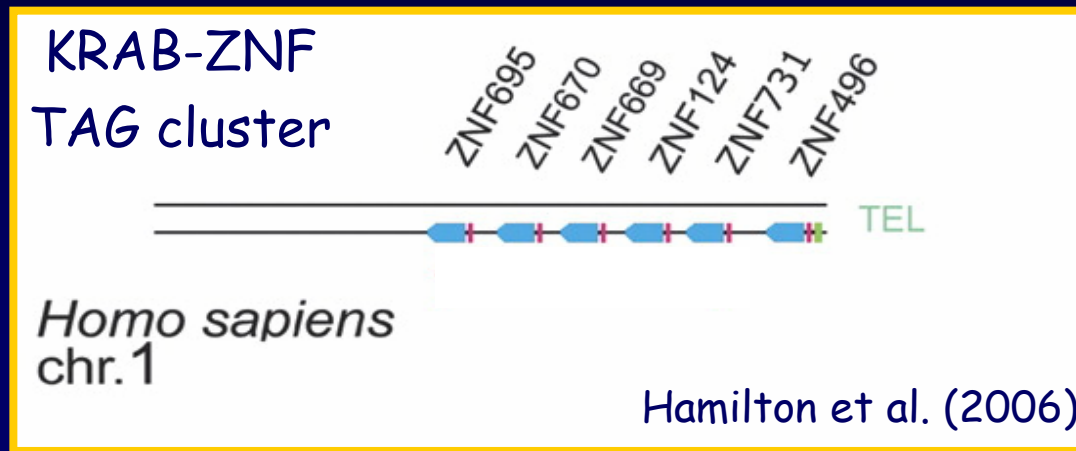
Mathieu Lajoie

Denis Bertrand

Nadia El-Mabrouk

Université de Montréal

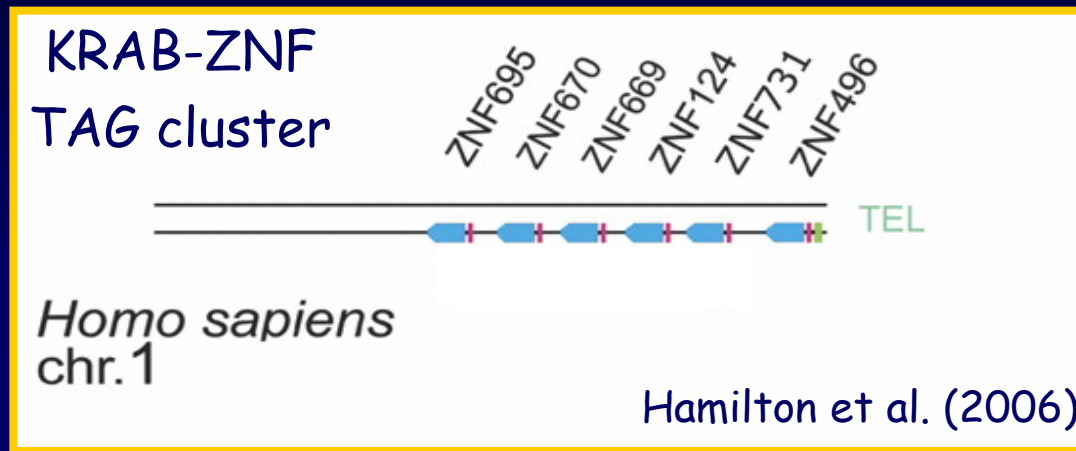
Tandemly Arrayed Genes (TAGs)



TAGs usually:

- Are similar in sequence and function
- Share a common ancestor
- Arise from tandem duplications
- Are good candidates for genetic innovation

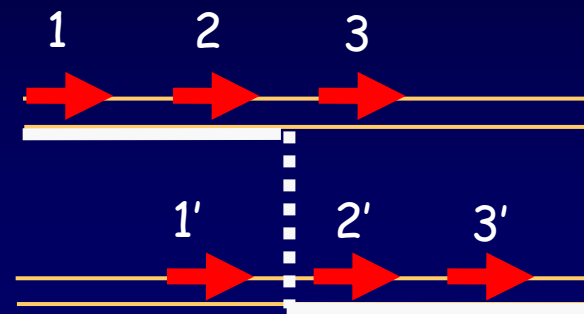
Tandemly Arrayed Genes (TAGs)



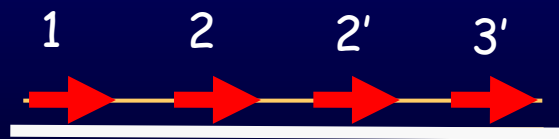
- Many gene families are organized in clusters of TAGs:
 - Major histocompatibility complex (MHC)
 - Olfactory receptor genes (~400 genes in human)
 - KRAB-ZNF genes (~420 genes in human)
- Each TAG cluster contains from 2 to several tens of genes

Tandem Duplication Model

(Fitch 1977)



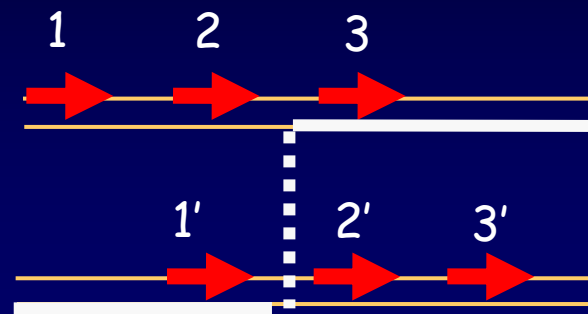
unequal cross-over
during meiosis



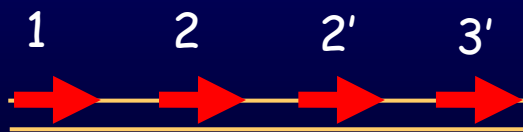
tandem
duplication
(simple)

Tandem Duplication Model

(Fitch 1977)



unequal cross-over
during meiosis

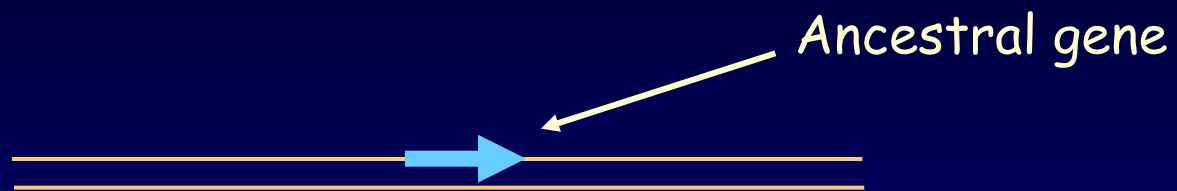


tandem
duplication
(simple)

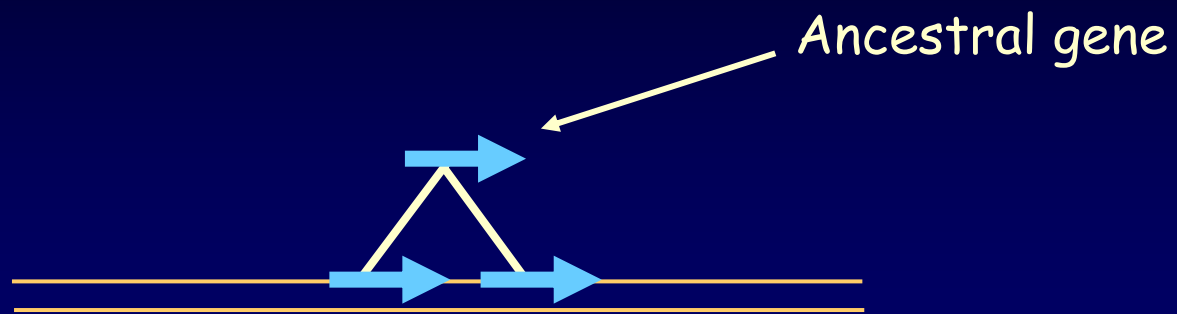


gene loss

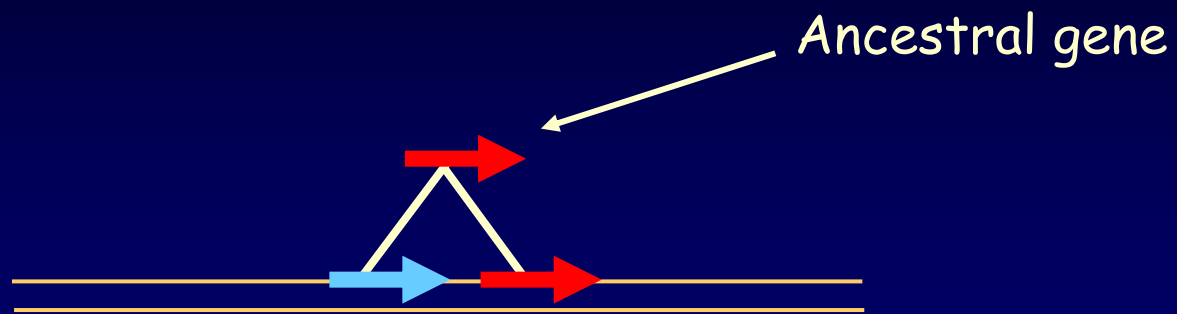
Tandem Duplication History (simple)



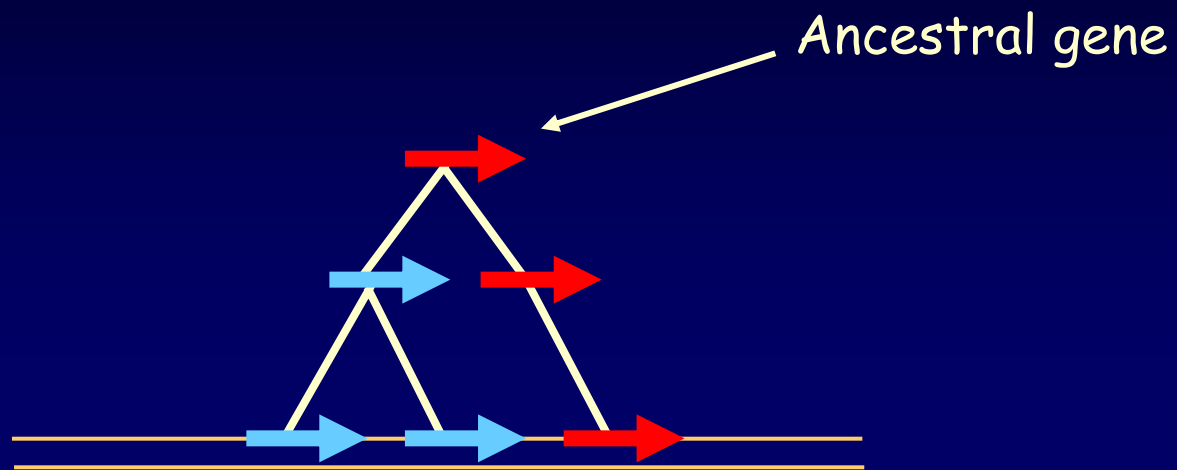
Tandem Duplication History (simple)



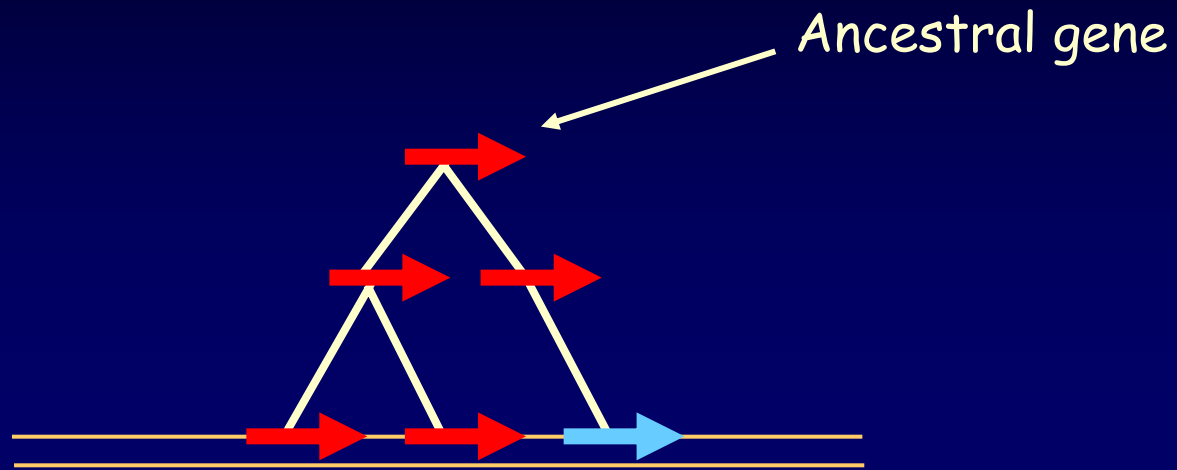
Tandem Duplication History (simple)



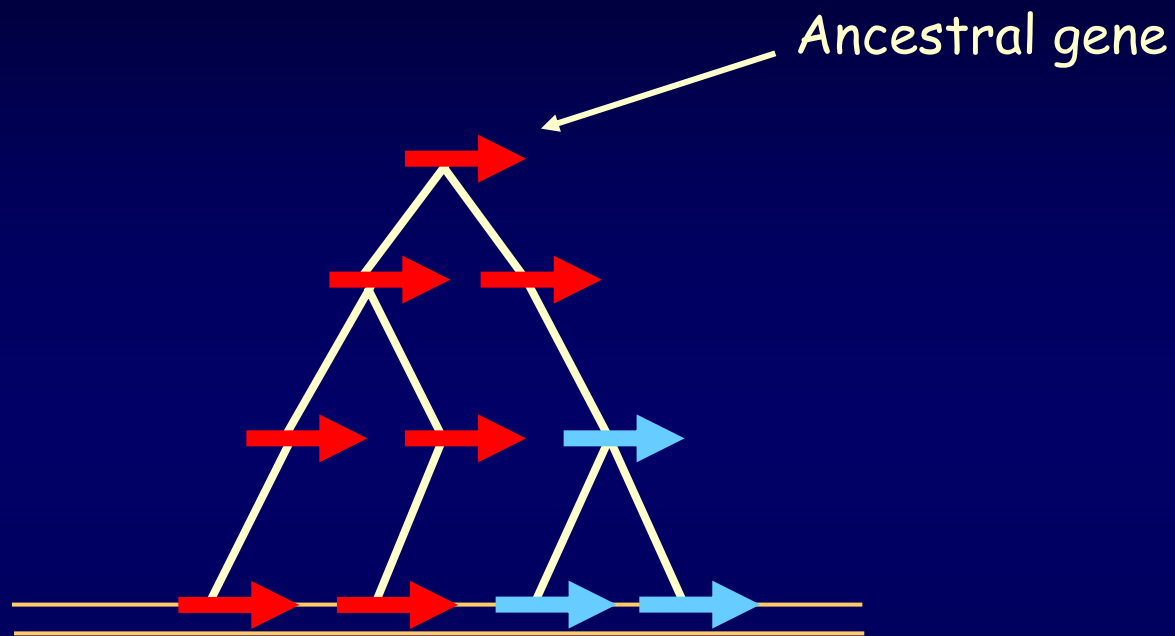
Tandem Duplication History (simple)



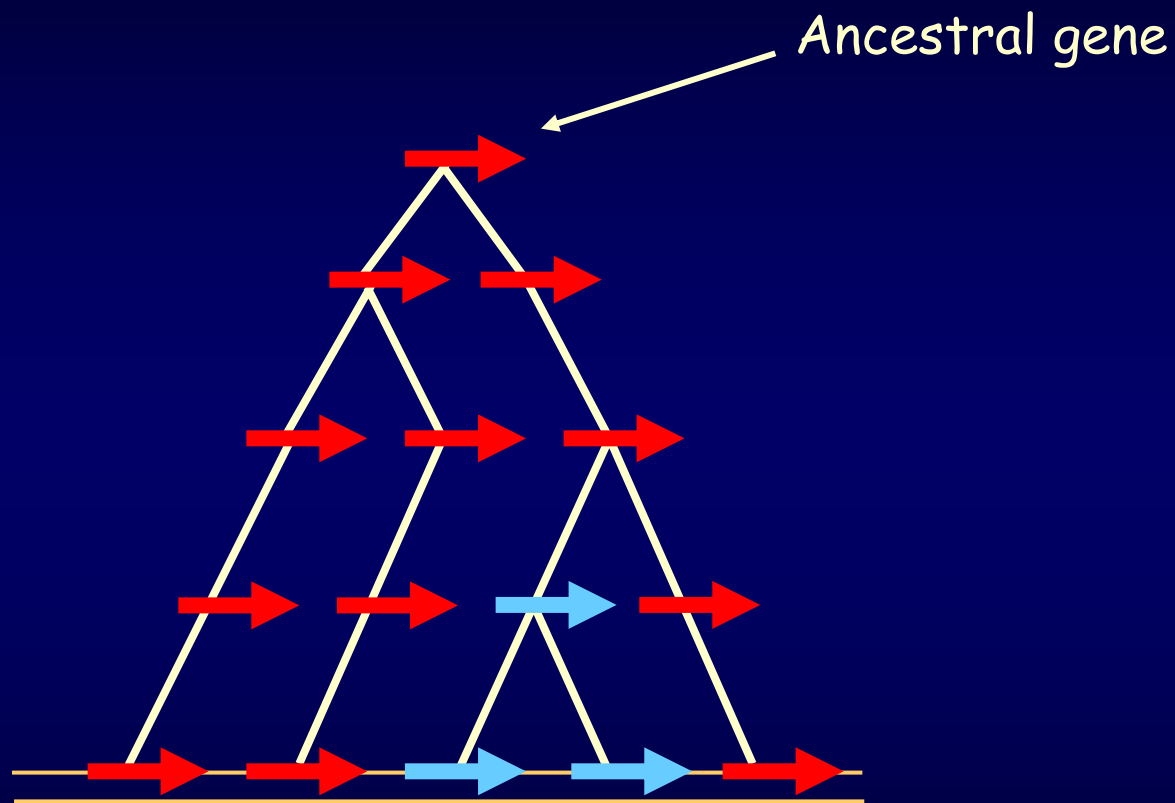
Tandem Duplication History (simple)



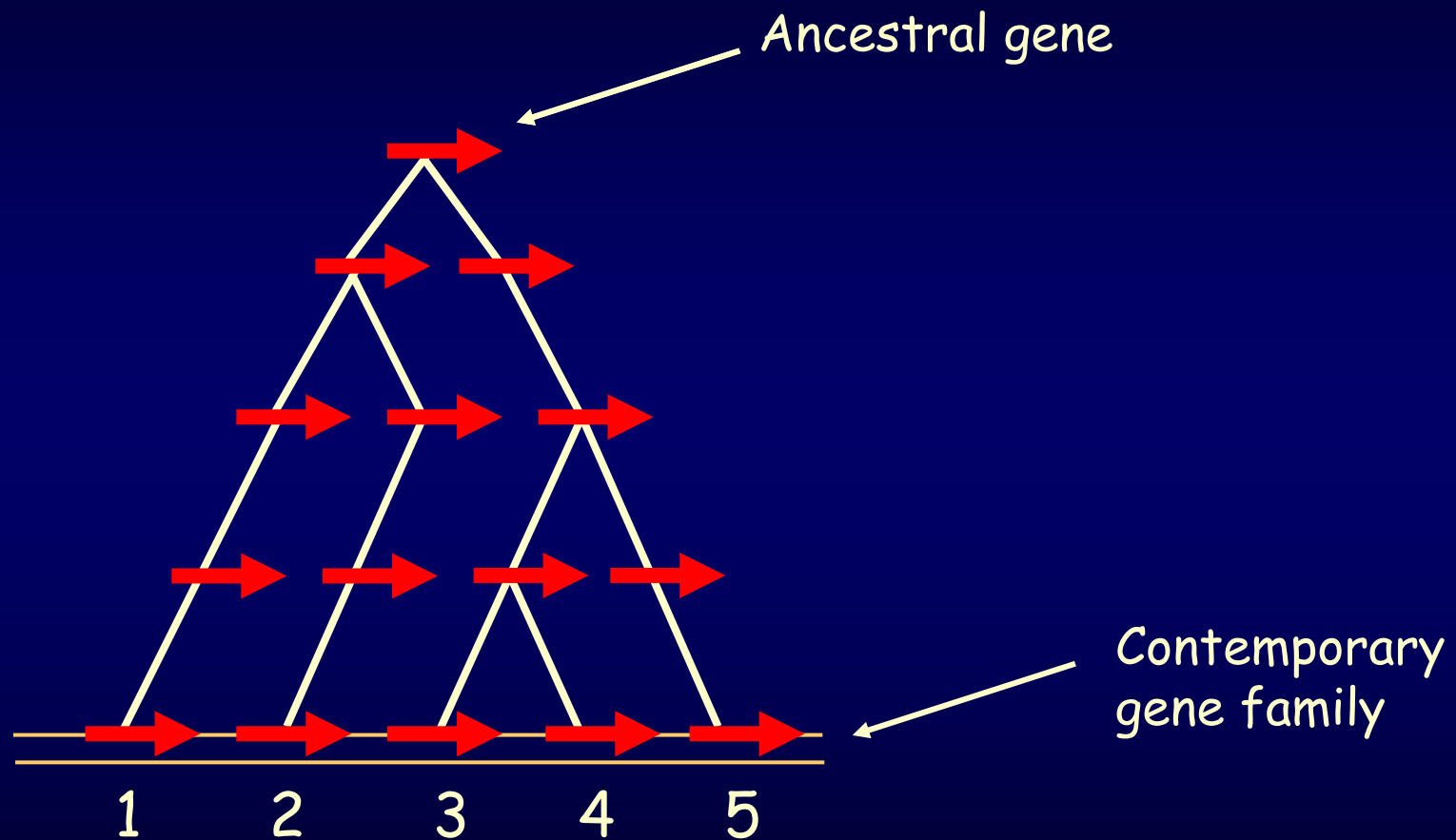
Tandem Duplication History (simple)



Tandem Duplication History (simple)

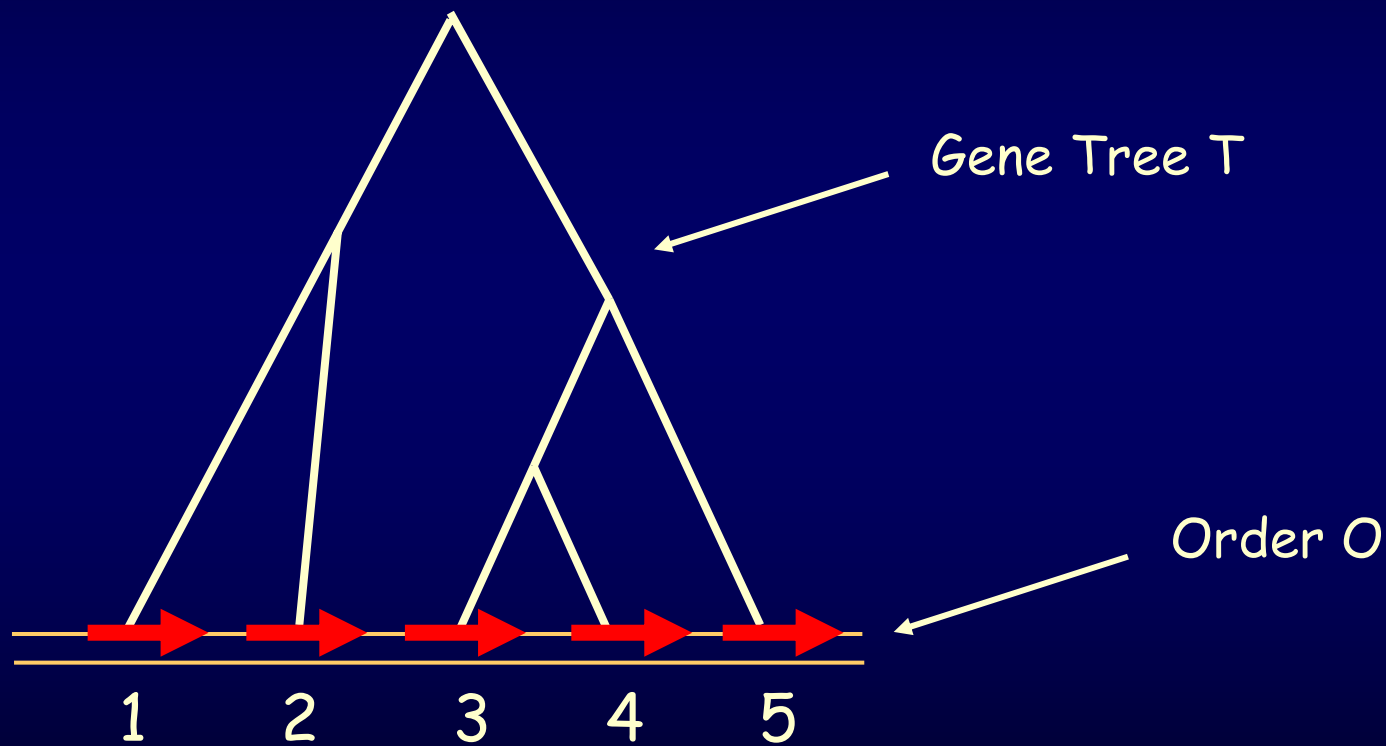


Tandem Duplication History (simple)



Tandem Duplication History (simple)

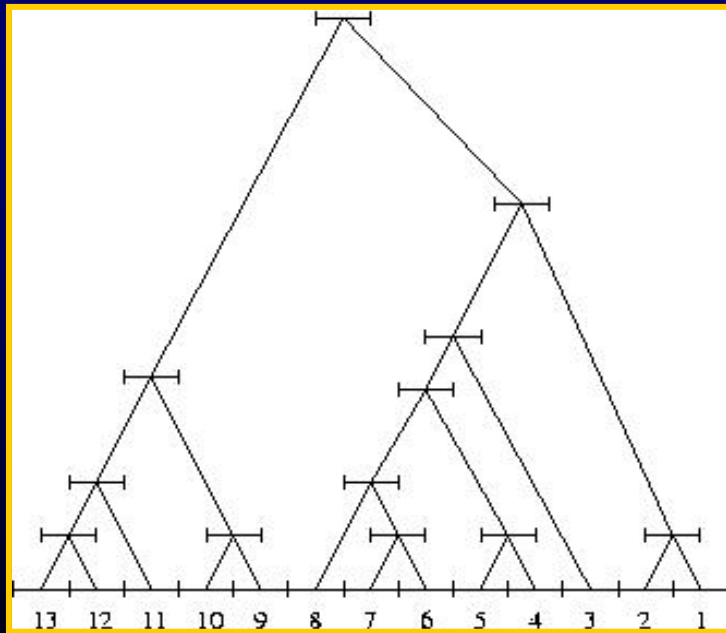
⇒ Induces an ordered gene tree (T, O)



Tandem Duplication Model (unrestricted)

Related Work

- Subject of many studies:
 - Theoretical results (*Jaitly et al.*), (*Gascuel et al.*), (*Zhang et al.*)
 - Inference algorithms (*Tang et al.*), (*Elemento and Gascuel*), (*Bertrand and Gascuel*)

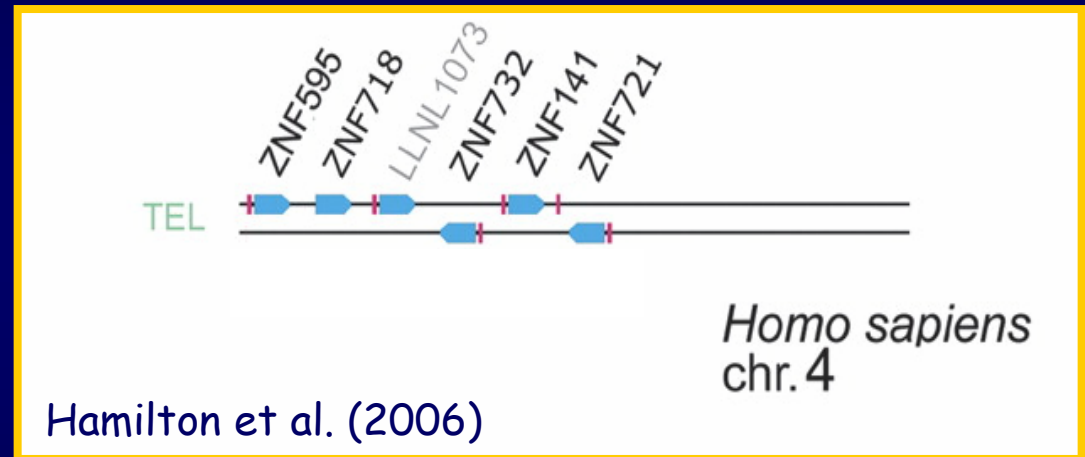


- Successful applications :
 - HOX genes (*Zhang and Nei*)
 - Human TRGV locus (*Elemento et al.*)

Tandem Duplication Model (unrestricted)

Limitations

- ❑ Many TAG clusters cannot be explained by a duplication history
- ❑ The model cannot explain the presence of genes in both orientations
- ❑ The model only considers cluster evolution in one species at a time

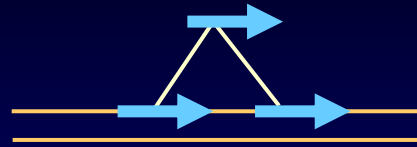


We propose a new evolutionary model accounting for simple Duplication, gene Loss, Inversion and Speciation (DLIS)

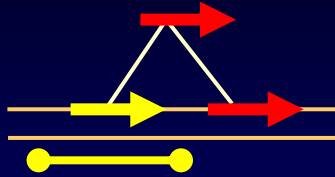
DLIS History



DLIS History

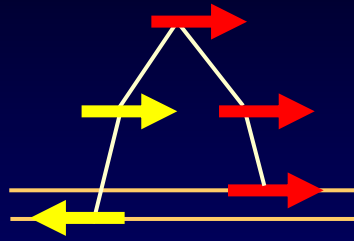


DLIS History

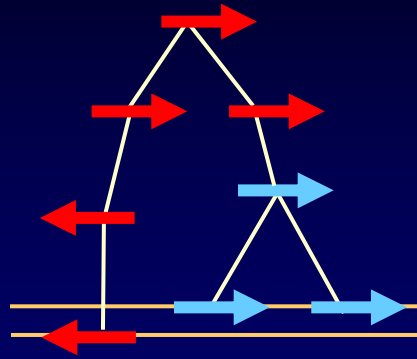


DLIS History

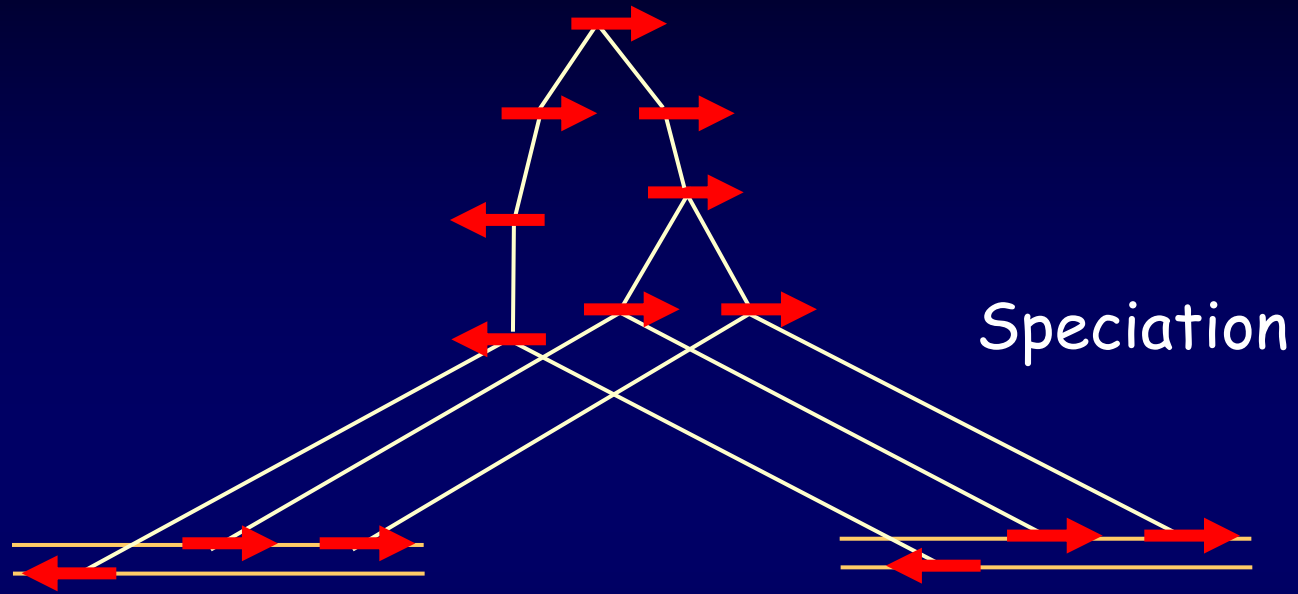
Inversion



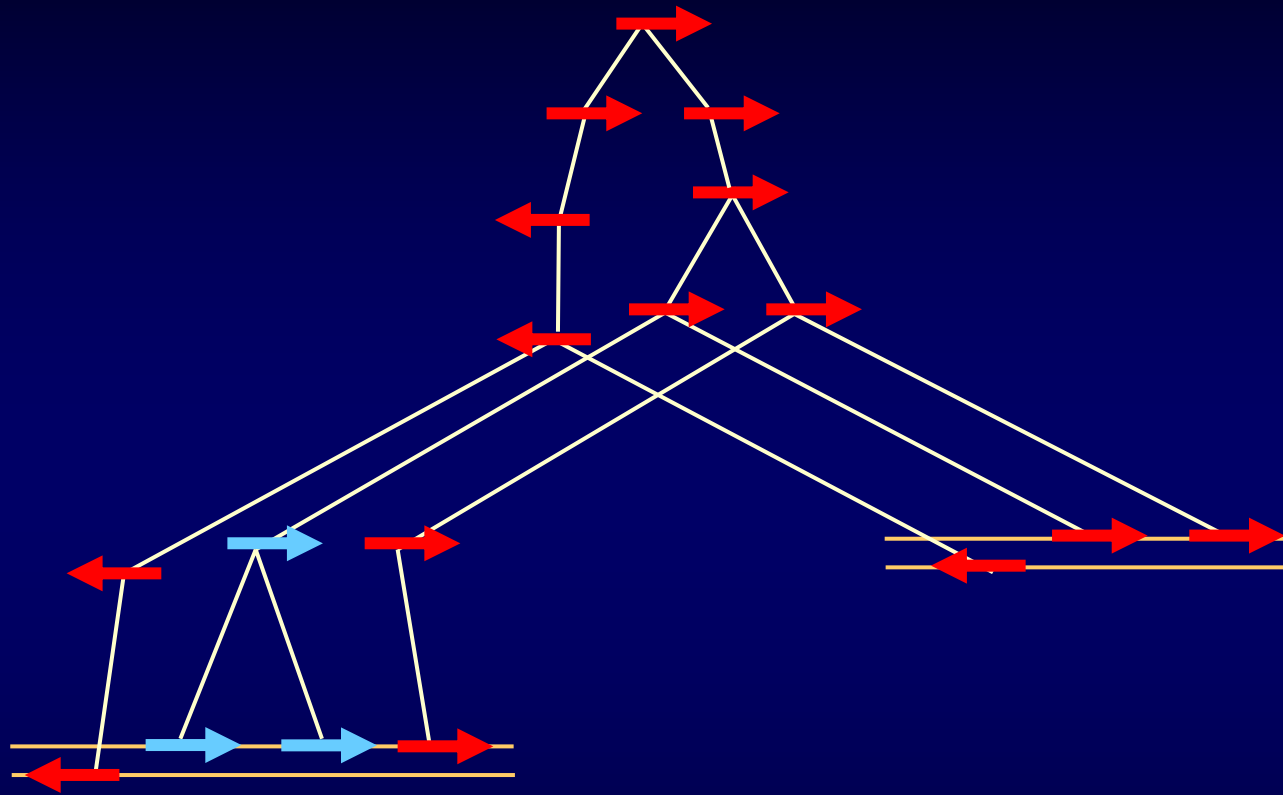
DLIS History



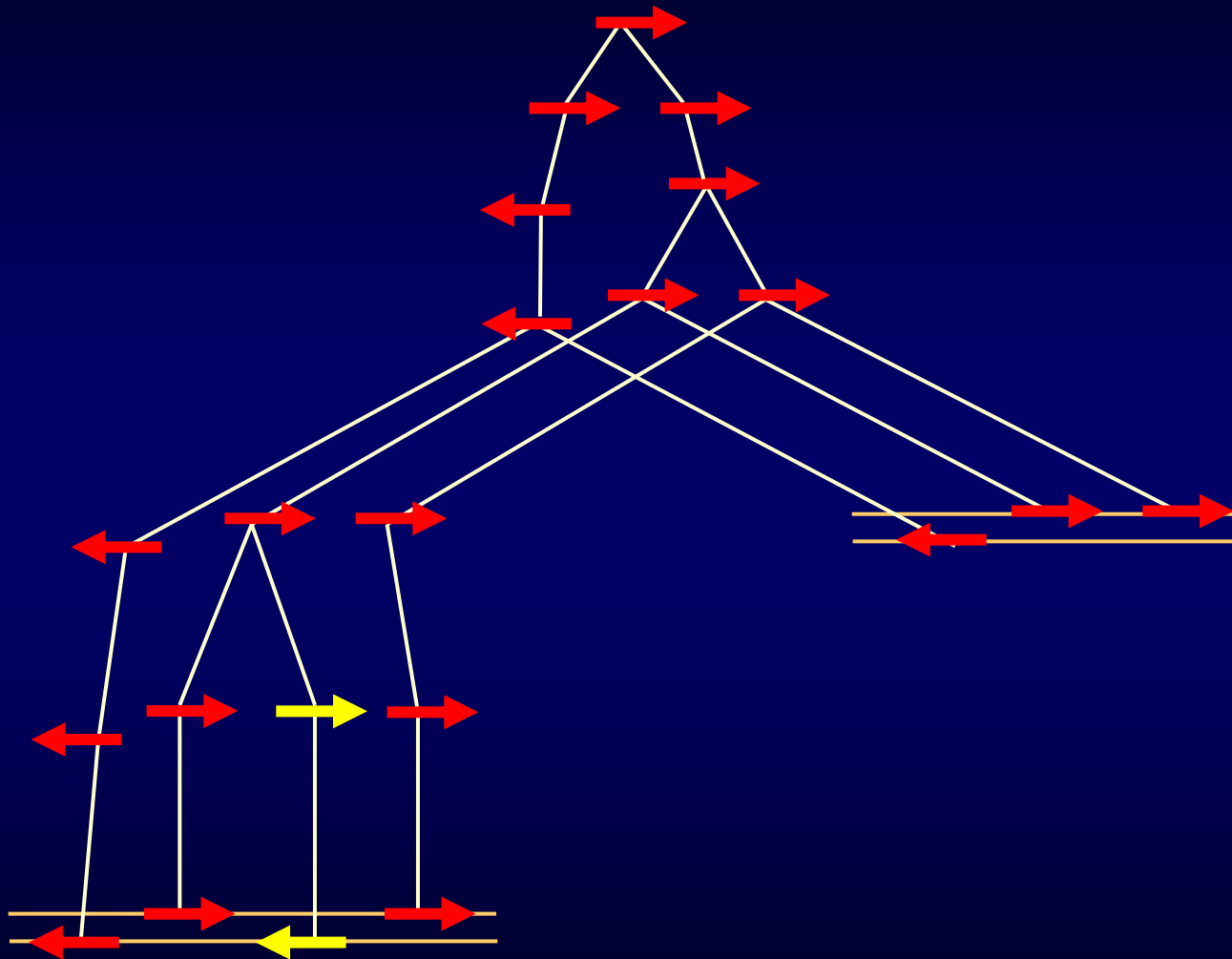
DLIS History



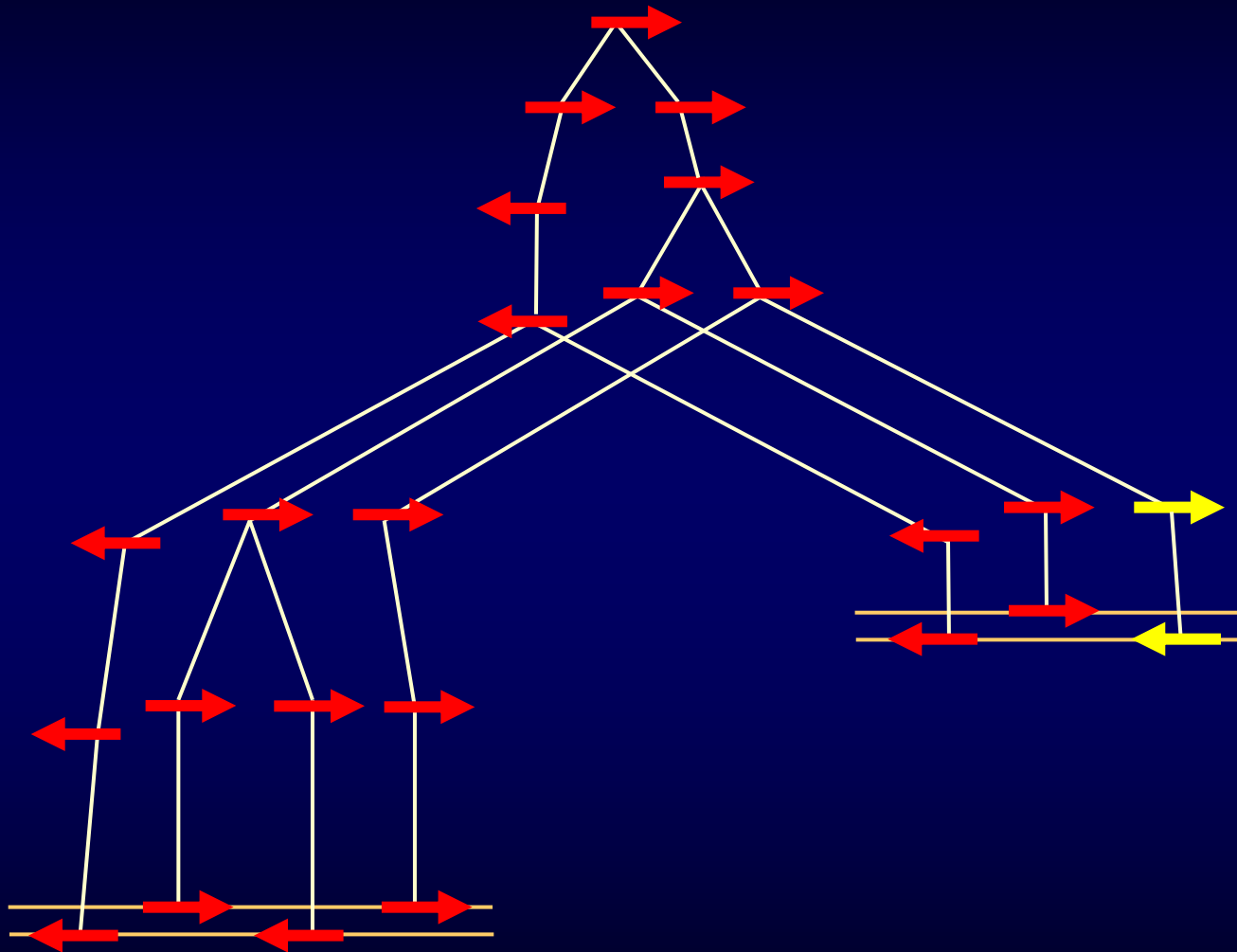
DLIS History



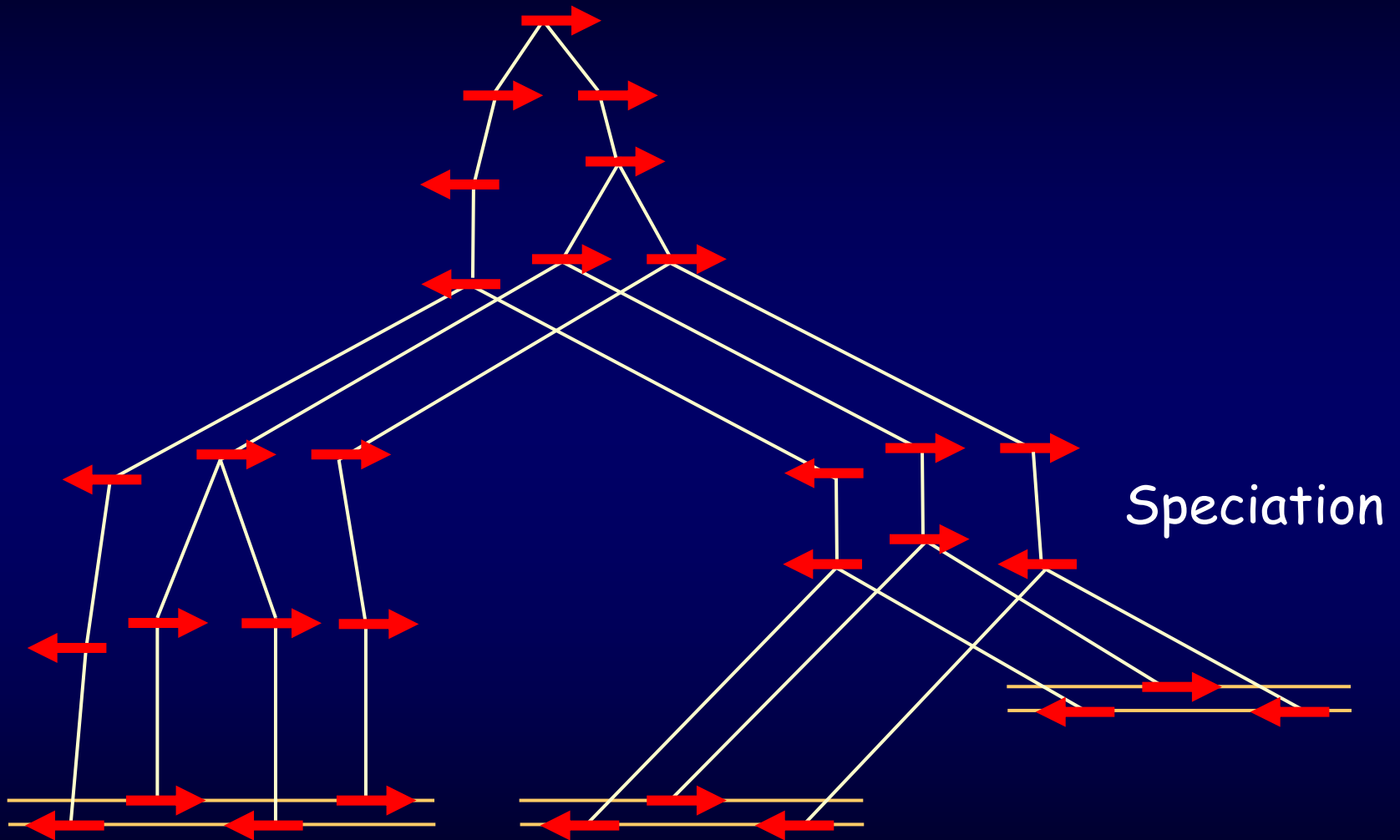
DLIS History



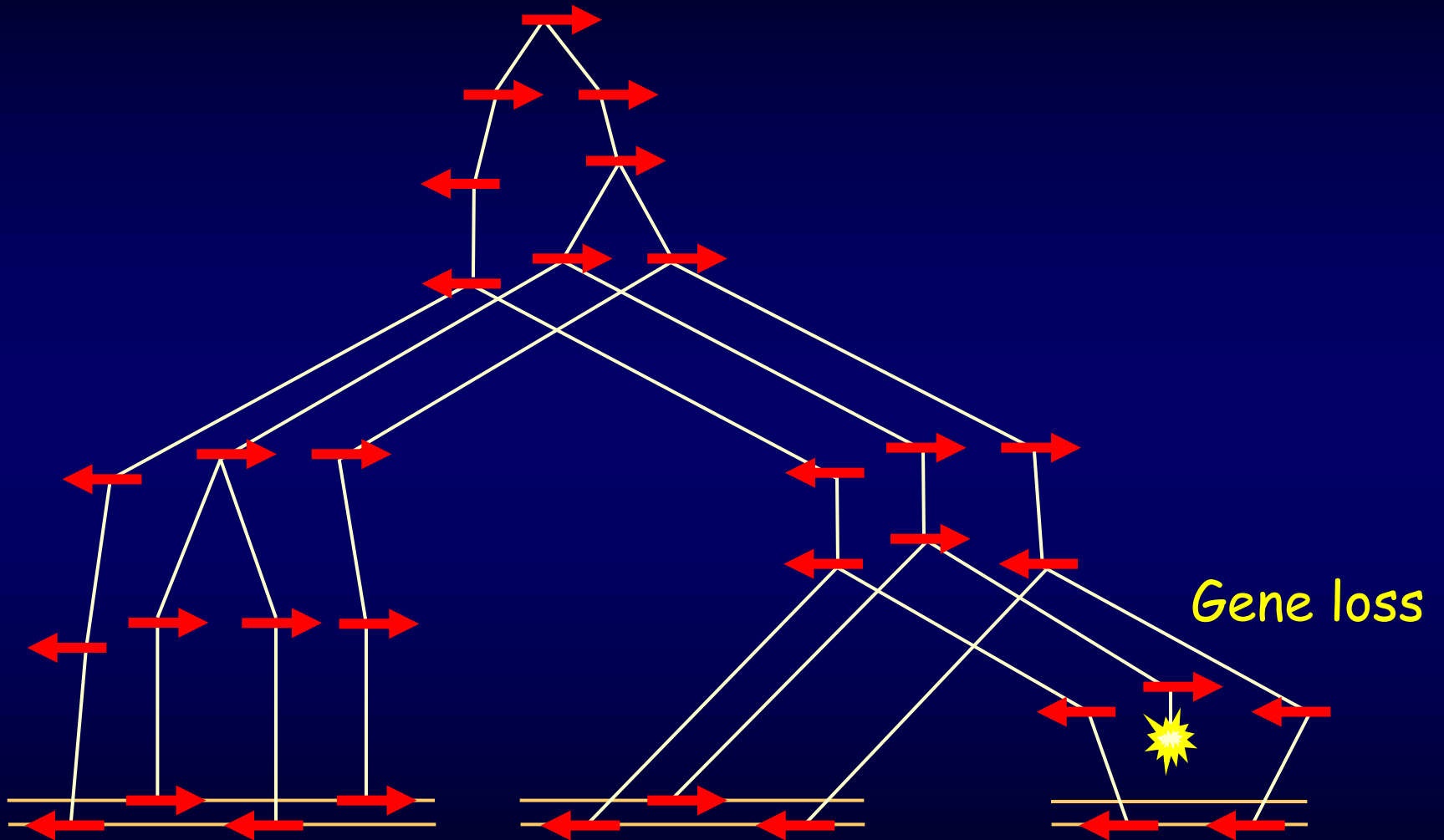
DLIS History



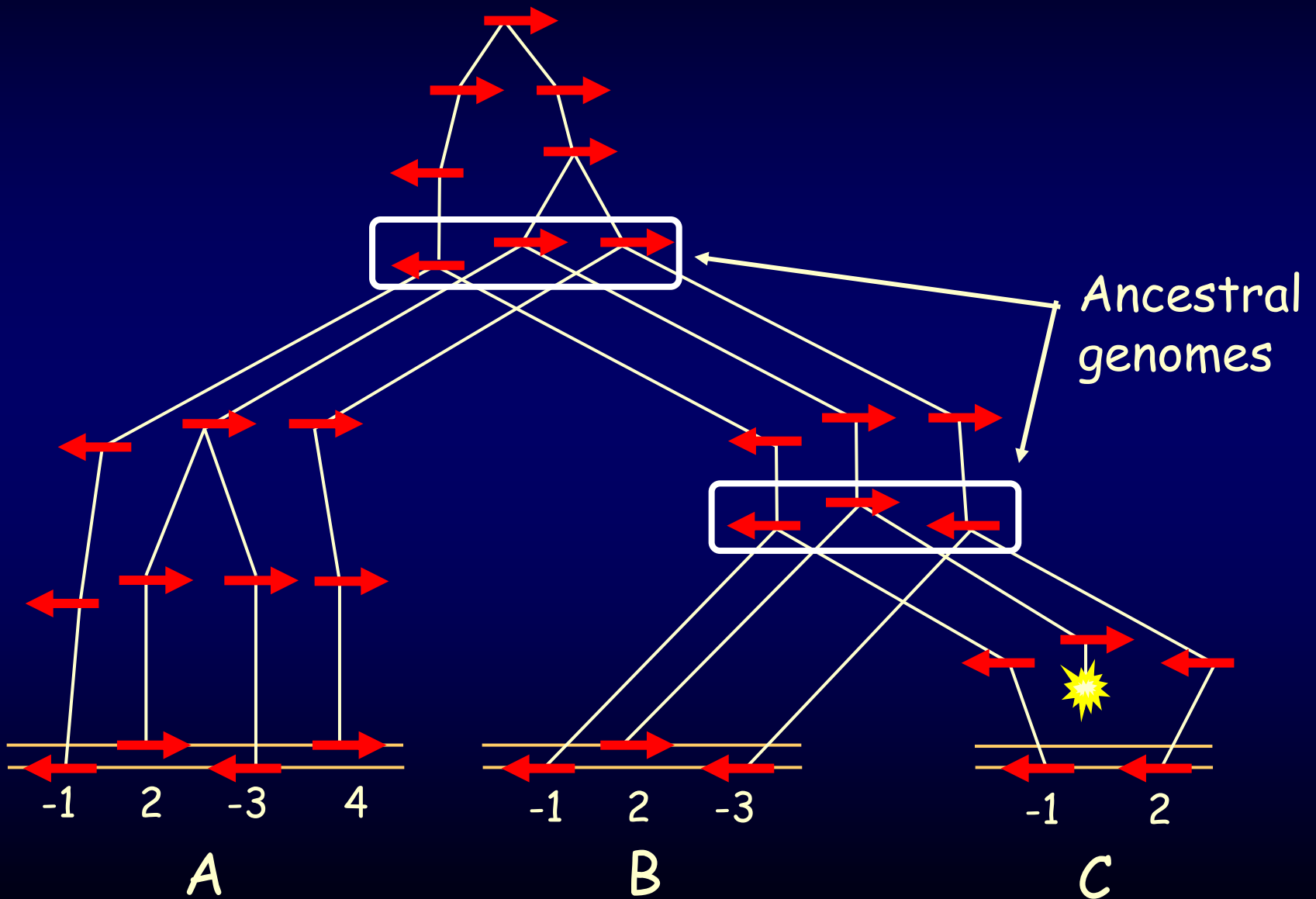
DLIS History



DLIS History

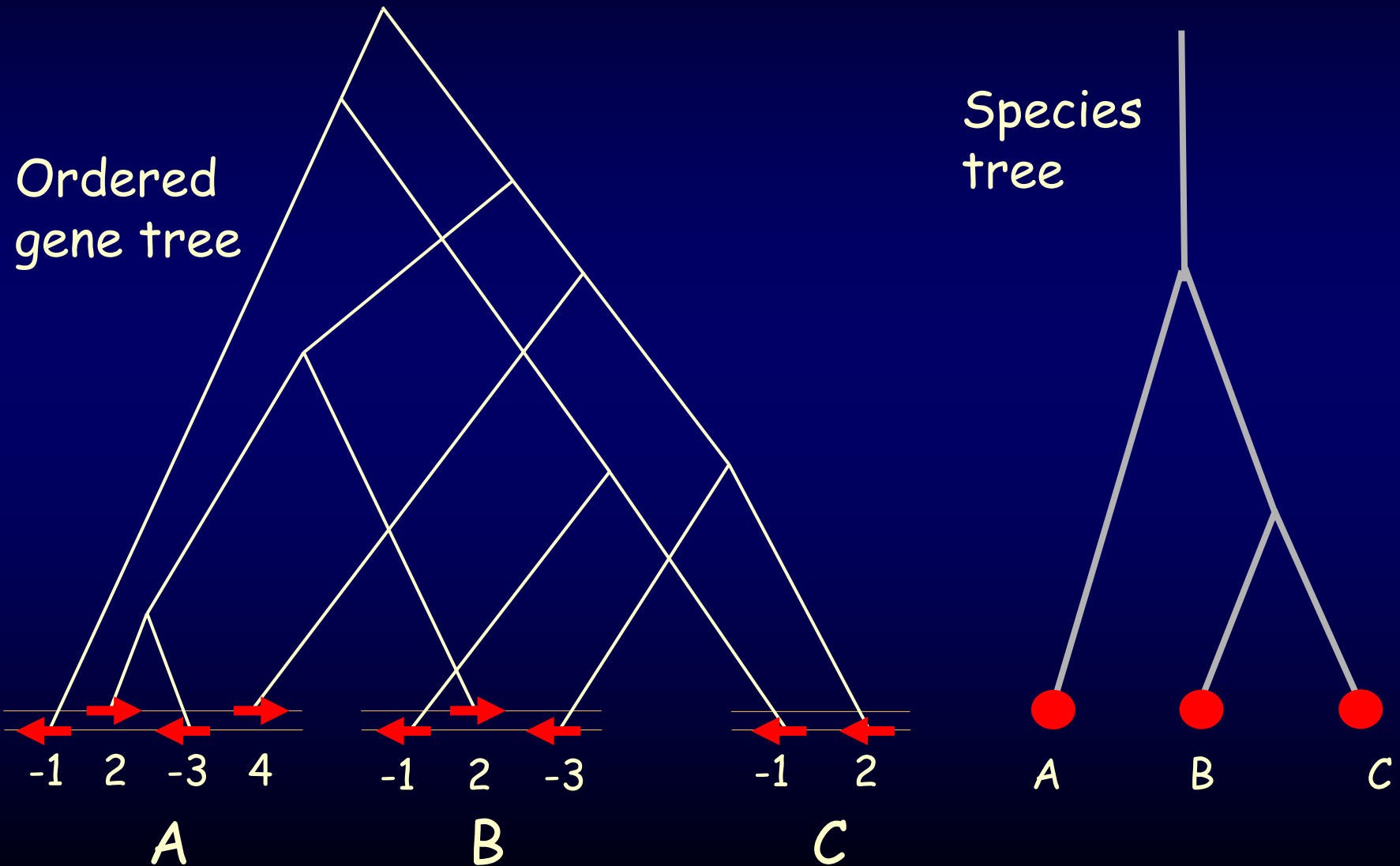


DLIS History



DLIS History

⇒ Induces unique gene and species trees

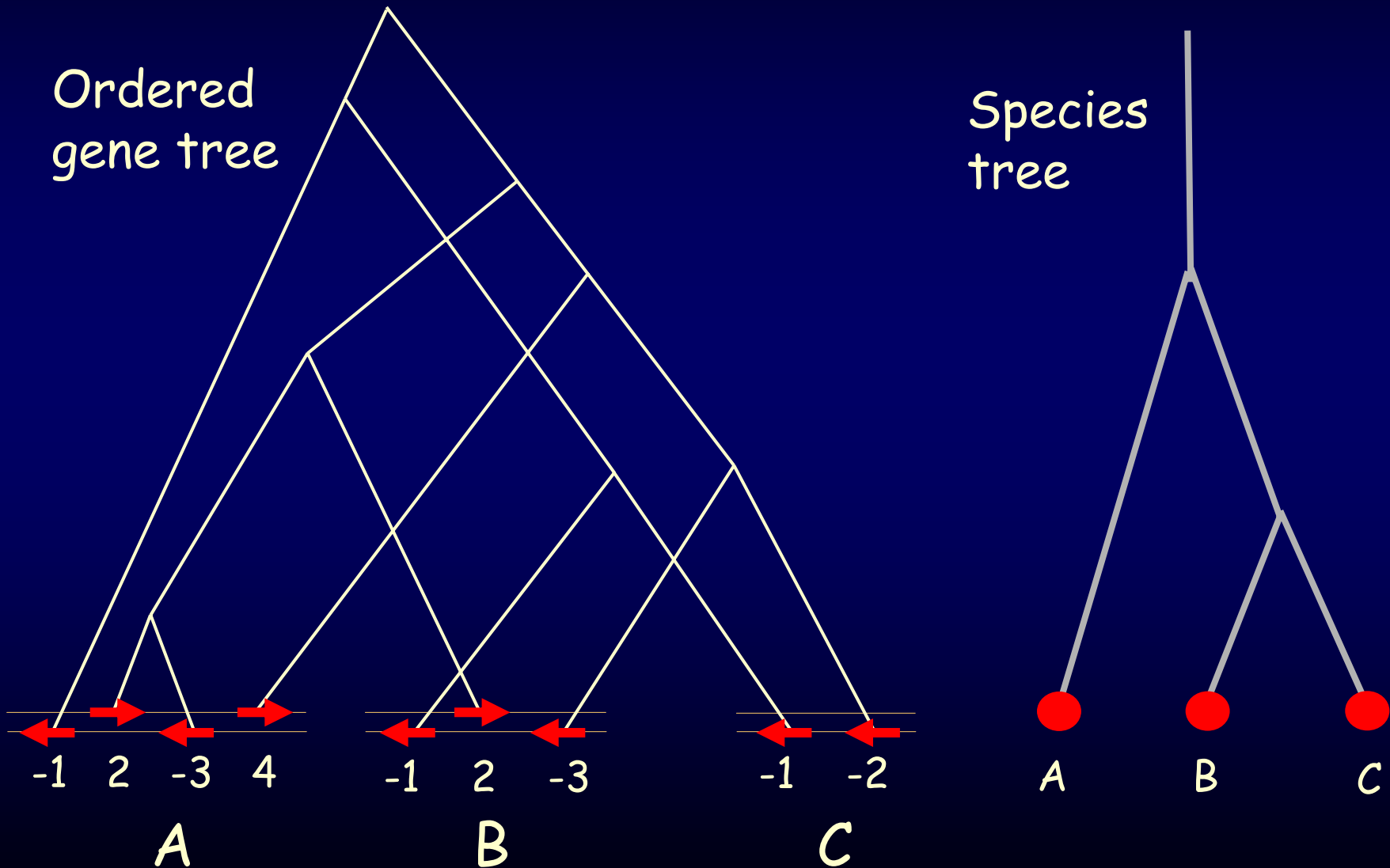


An Inference Problem

- ❑ **Lemma 1:** Any ordered gene tree can be explained by a DLIS history for a given species tree
- ❑ **Problem:** Find a DLIS history involving a minimum number of evolutionary events
- ❑ **General approach:**
 1. Infer gene contents in the ancestral species by reconciliation (gene loss and duplication)
 2. Infer the gene orders which minimize the total number of inversions in the species tree

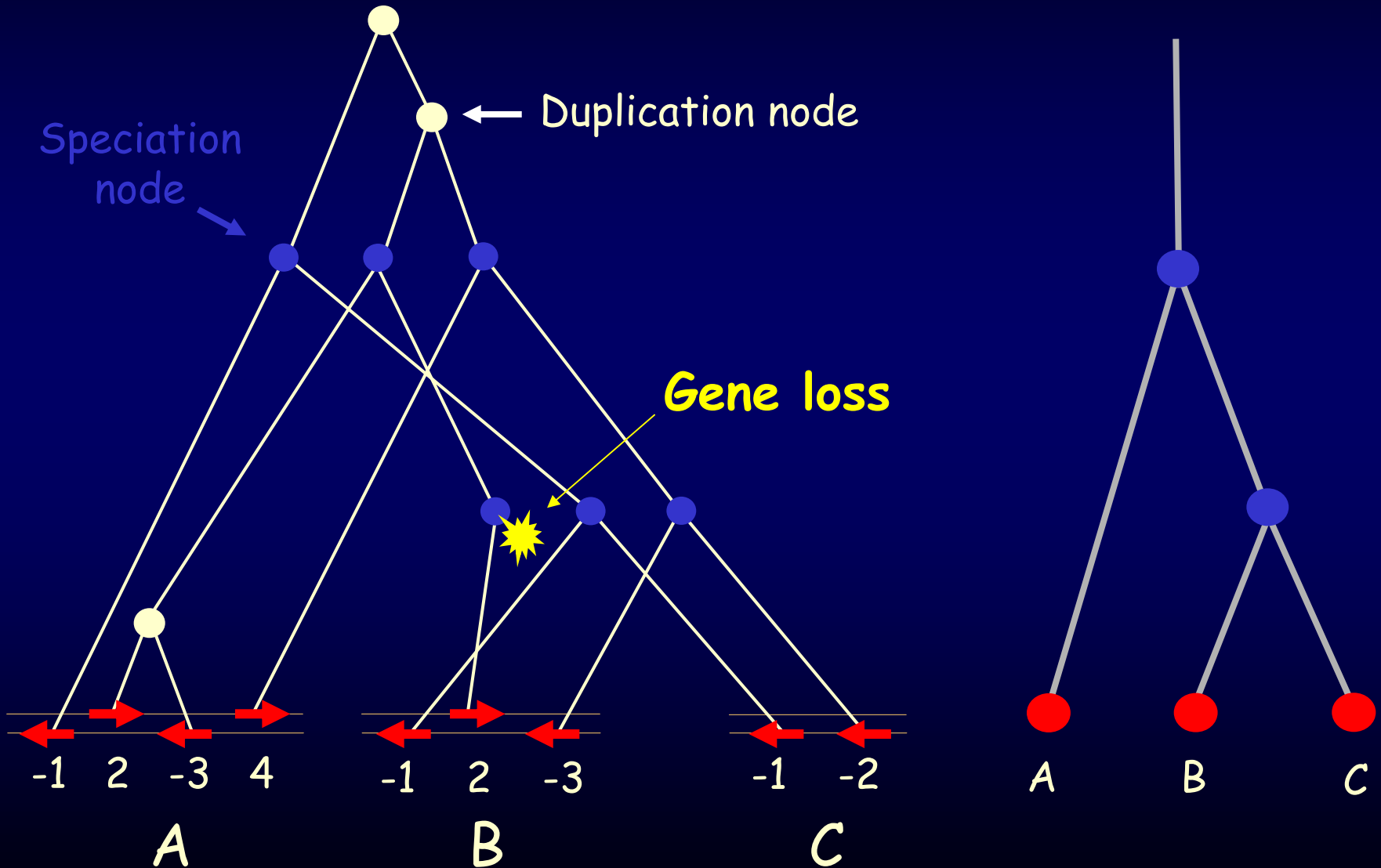
Inferring Ancestral Gene Contents

Species/Gene Tree Reconciliation

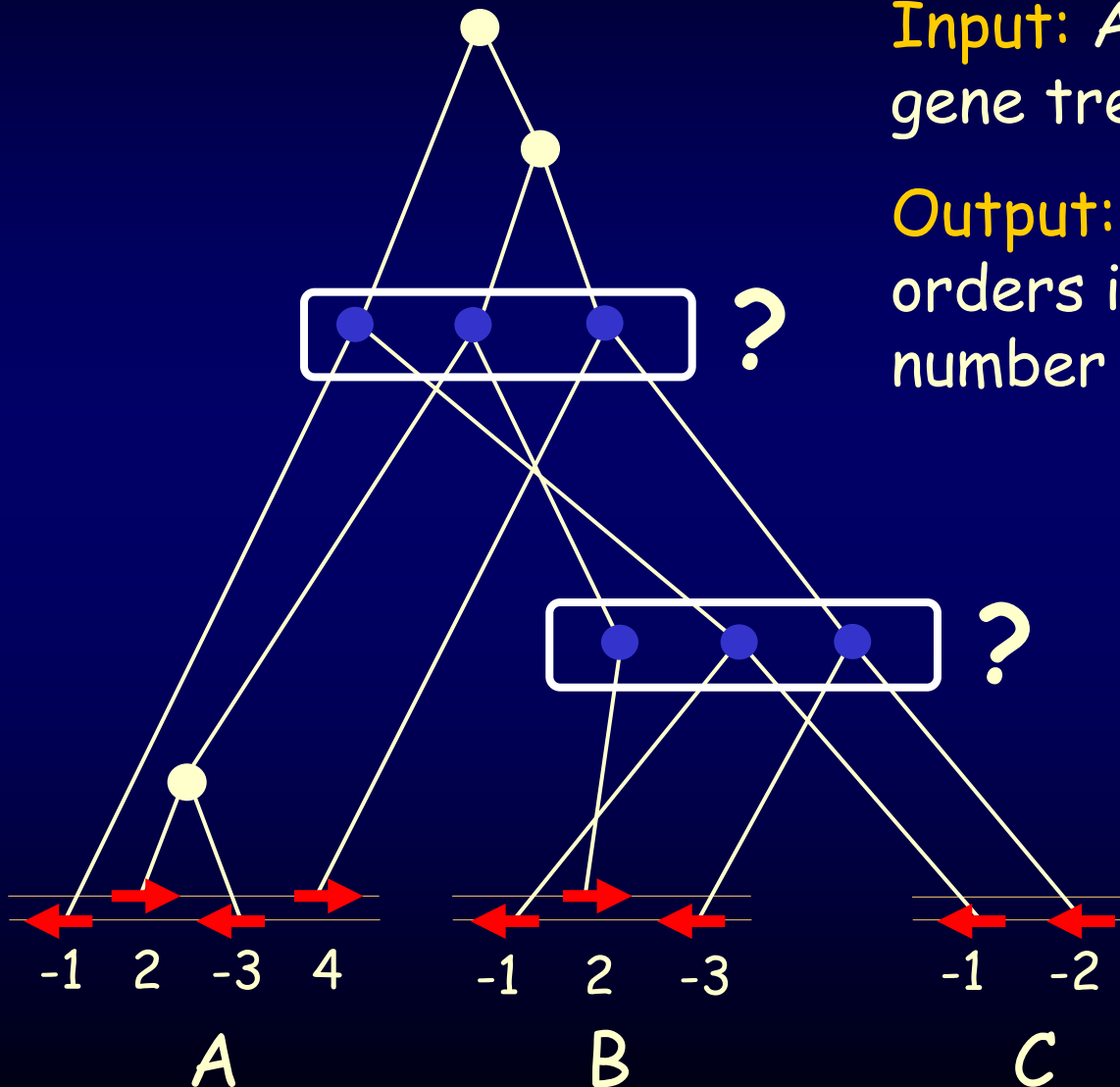


Inferring Ancestral Gene Contents

⇒ Ordered reconciled gene tree



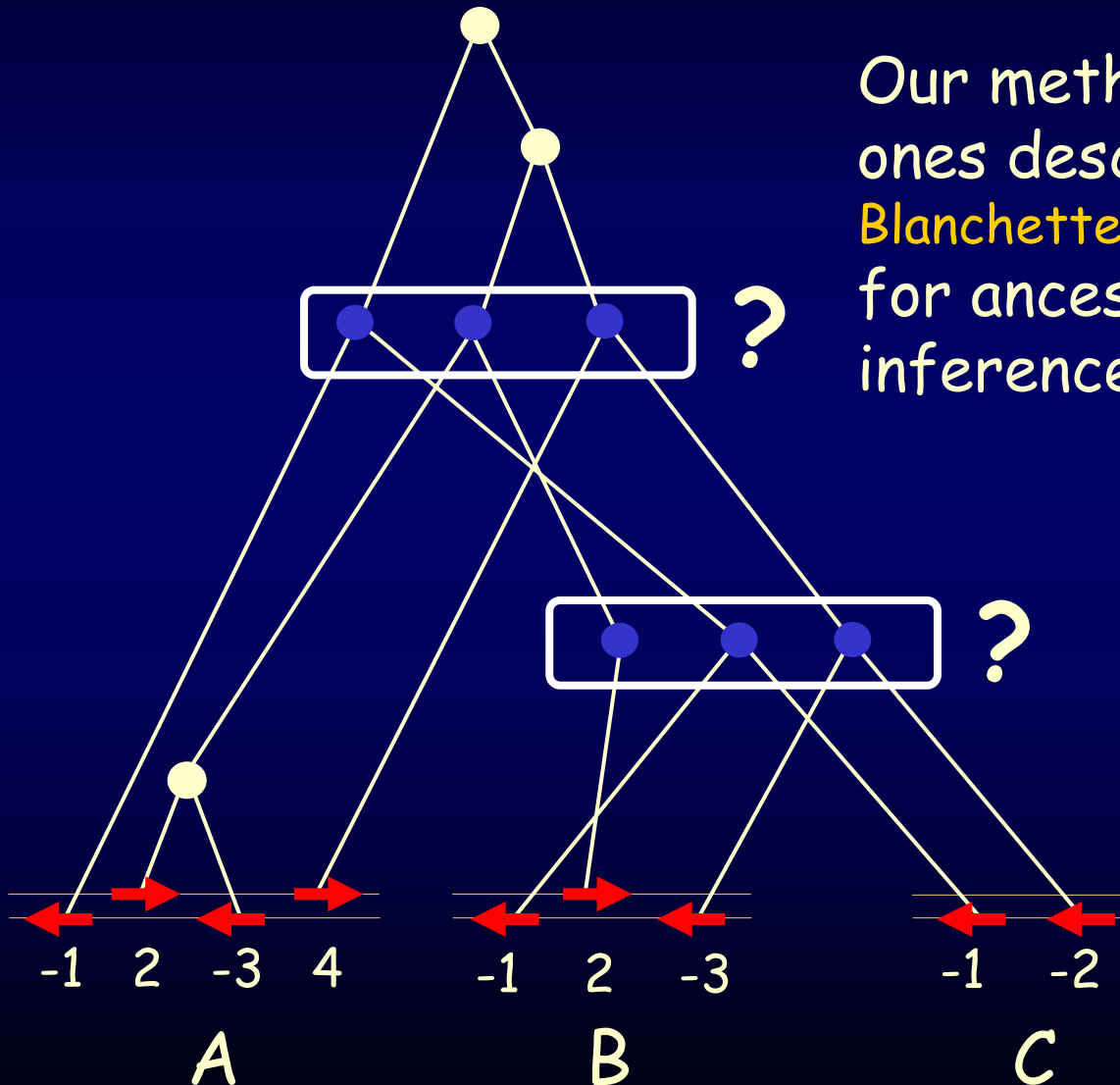
The Minimum DLIS Problem



Input: An ordered reconciled gene tree

Output: Ancestral gene orders inducing a minimum number of inversions

Inferring Ancestral Gene Orders

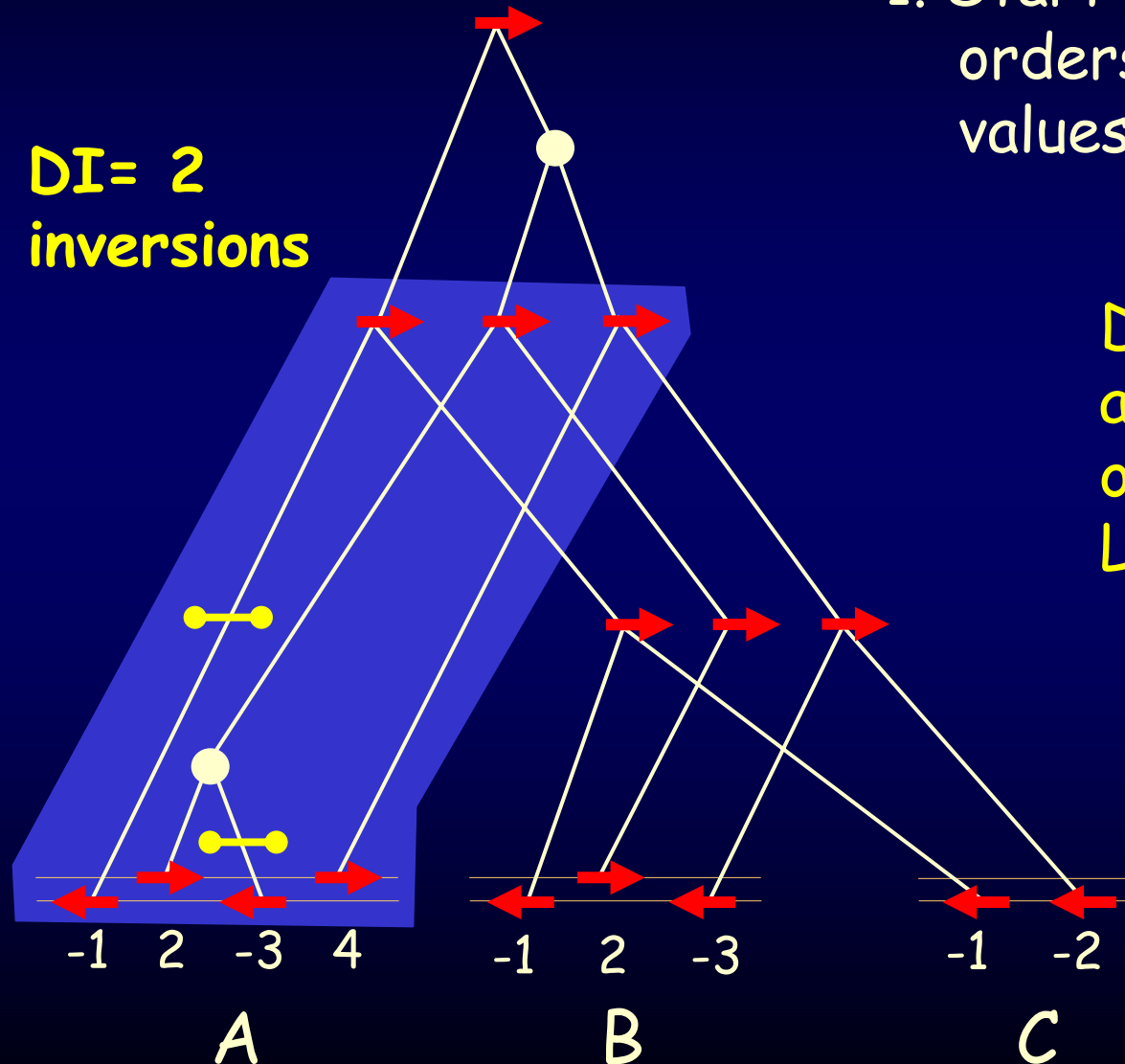


Our method is similar to the ones described in *Sankoff and Blanchette ; Moret et al. (GRAPPA)* for ancestral gene orders inference

Inferring Ancestral Gene Orders

1. Start with arbitrary gene orders and compute **DI** values for each branch

DI = 2
inversions

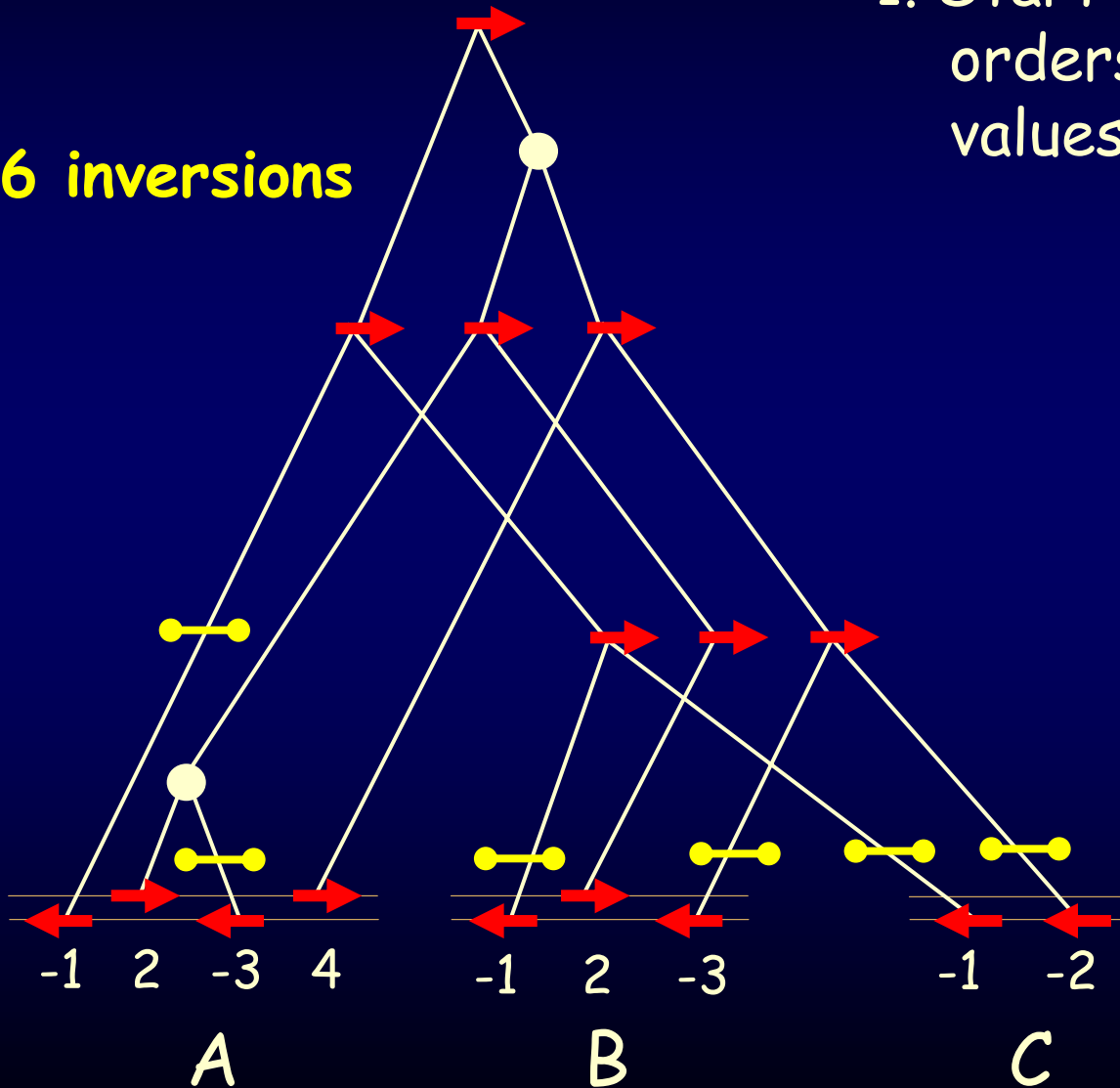


DI computed with
a generalization of
our algorithm in
Lajoie et al. (2006)

Inferring Ancestral Gene Orders

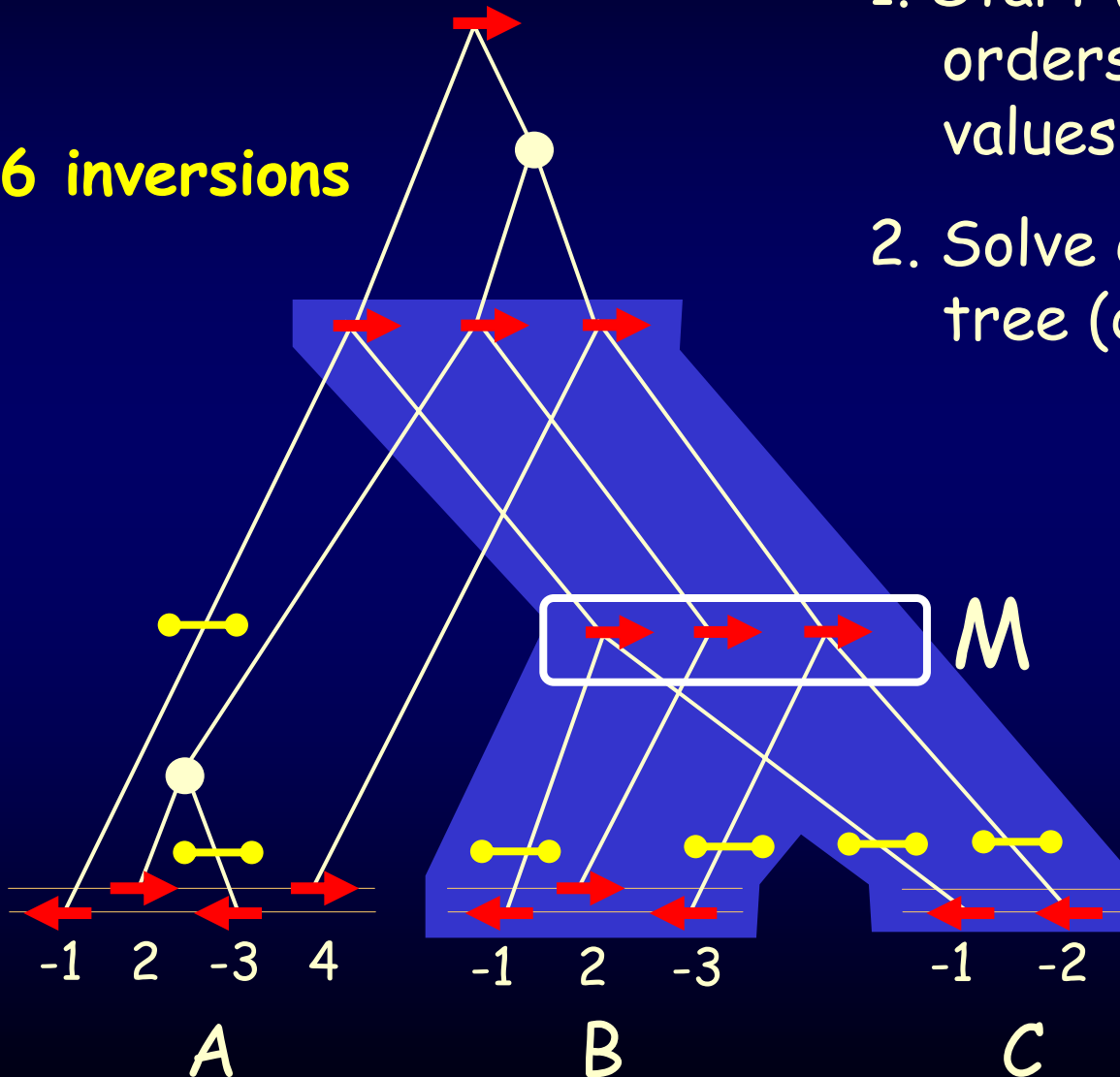
1. Start with arbitrary gene orders and compute DI values for each branch

6 inversions



Inferring Ancestral Gene Orders

6 inversions

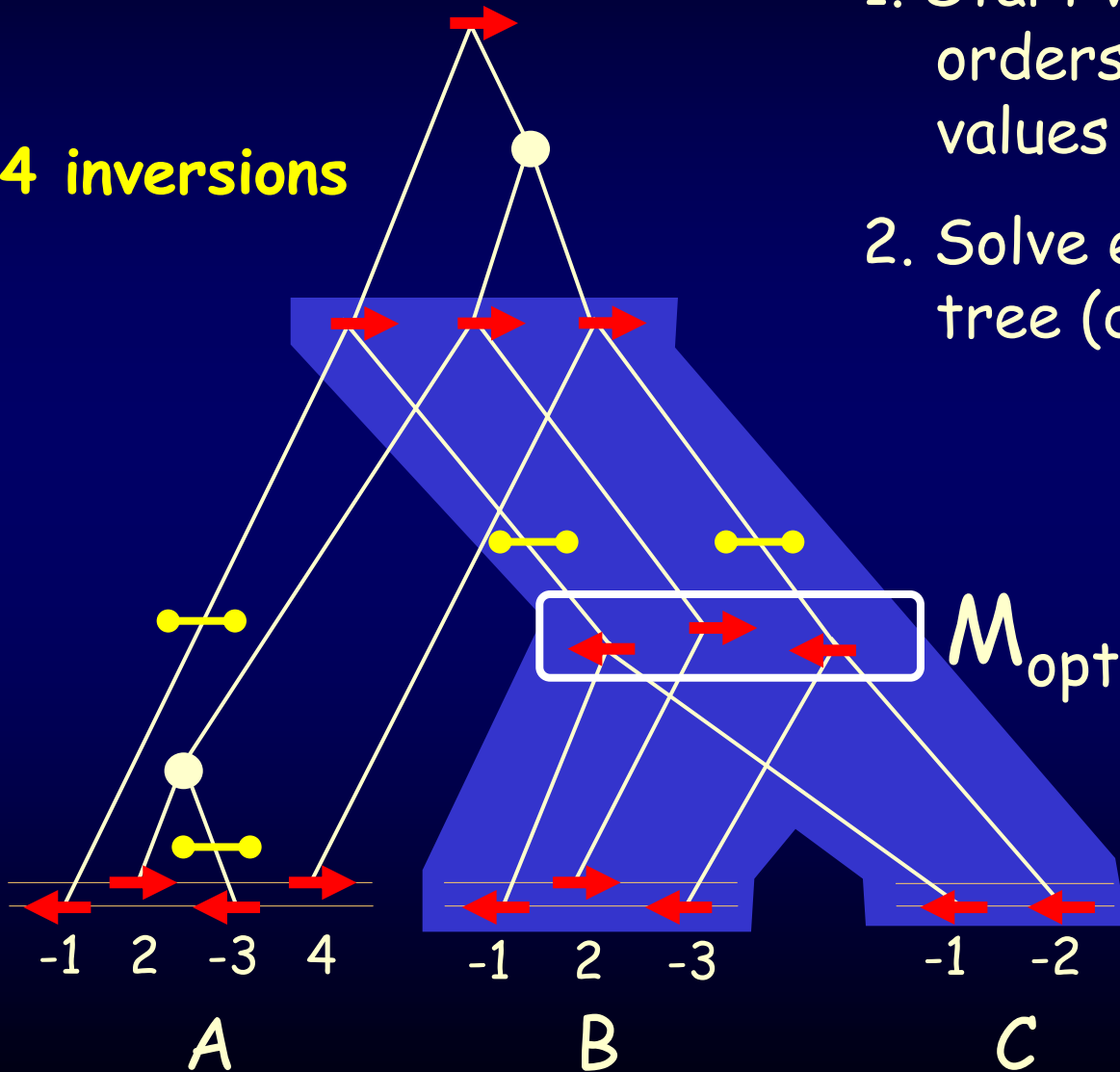


1. Start with arbitrary gene orders and compute DI values for each branch
2. Solve each median in the tree (depth-first search)

$$S_M = DI(M, X) + DI(M, Y) + DI(M, Z)$$

Inferring Ancestral Gene Orders

4 inversions



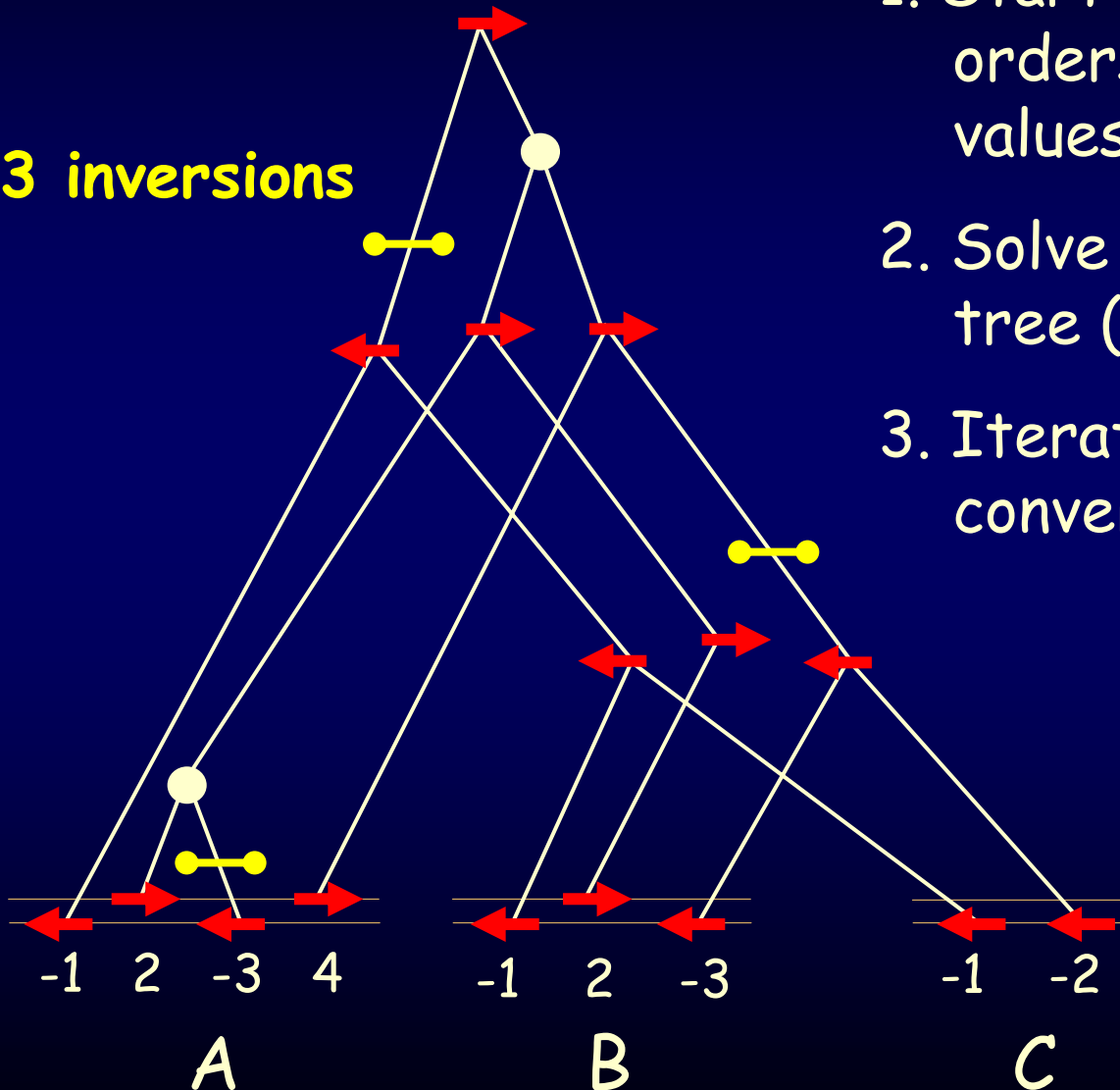
1. Start with arbitrary gene orders and compute DI values for each branch
2. Solve each median in the tree (depth-first search)

$$S_{M_{opt}} = 2$$

M_{opt}

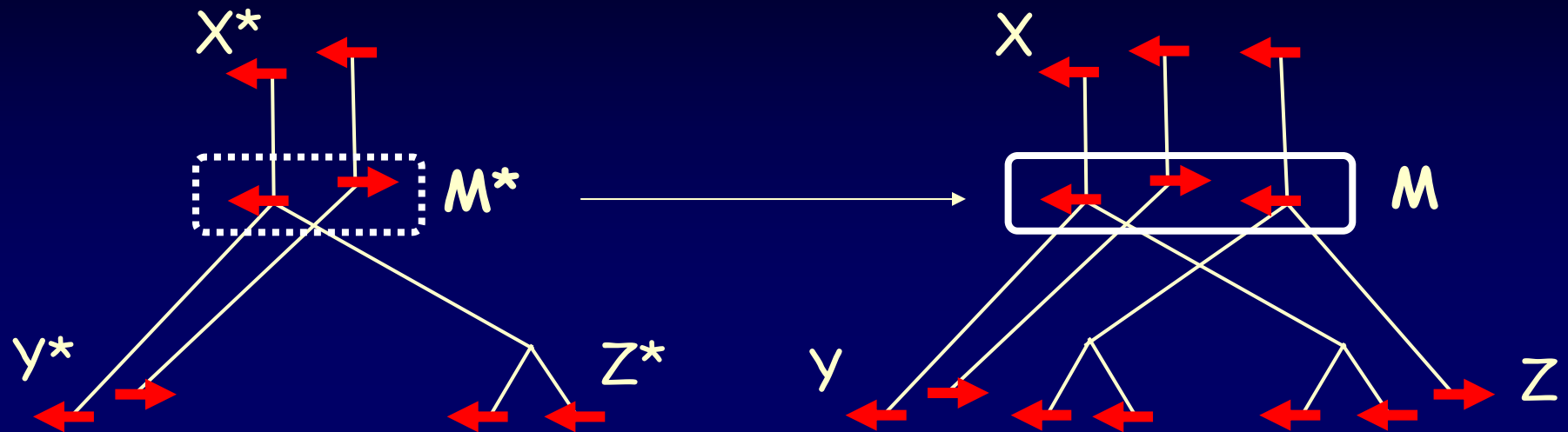
Inferring Ancestral Gene Orders

3 inversions



1. Start with arbitrary gene orders and compute **DI** values for each branch
2. Solve each median in the tree (depth-first search)
3. Iterate step 2 until convergence

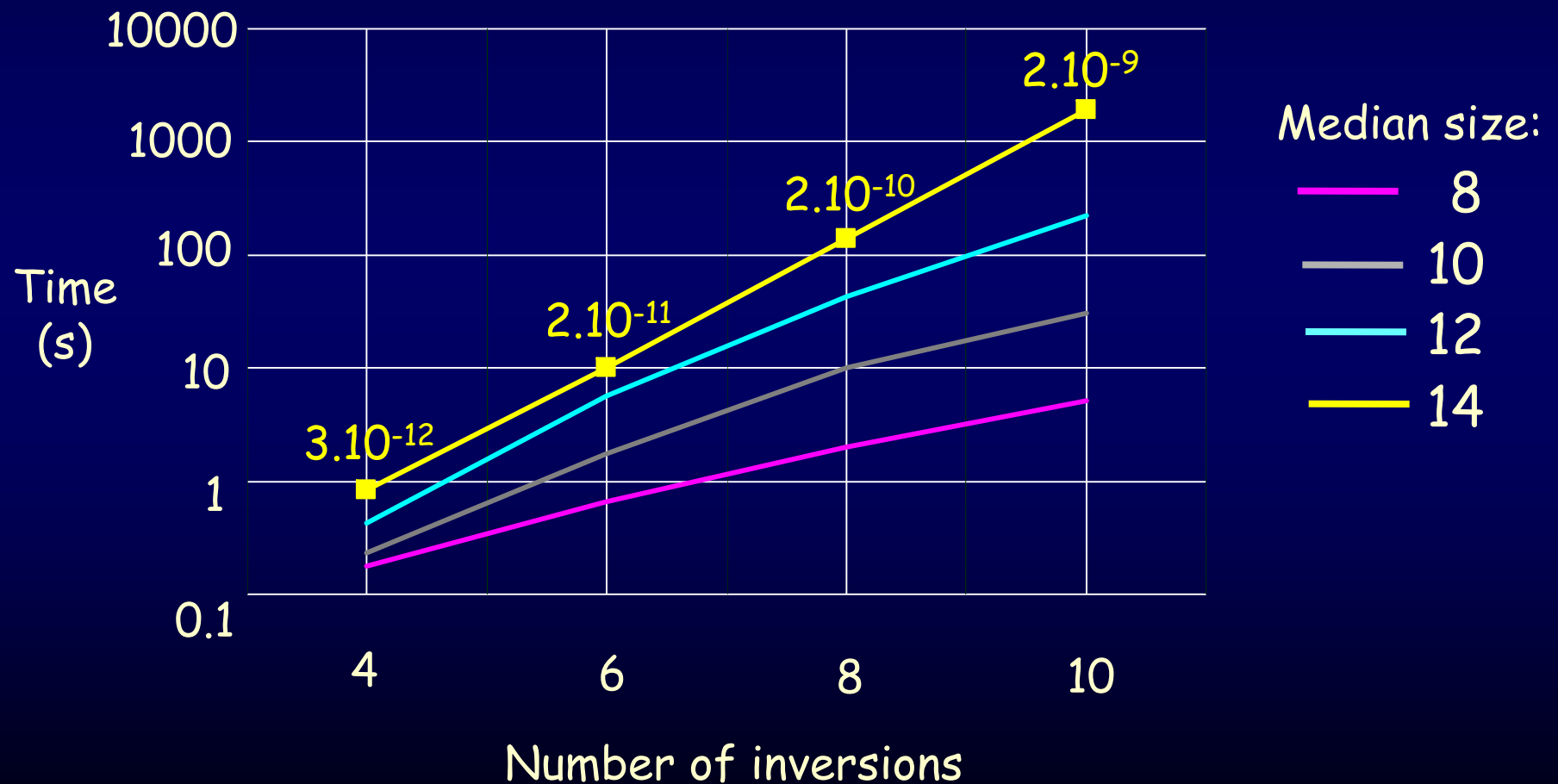
Solving the Median (NP-Hard)



- $2^n n!$ possible orders M
- Simple branch-and-bound strategy
 - Progressively extend the prefix M^* of M
 - Use the lower bound $S_{M^*} \leq S_M$

Branch-and-bound Efficiency

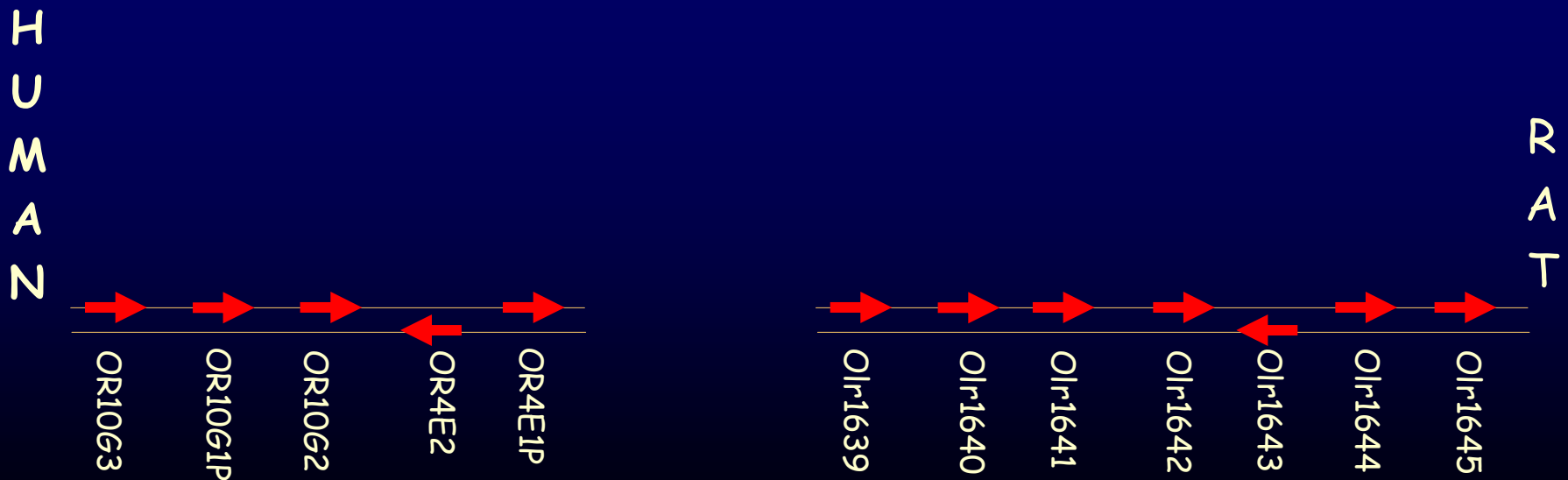
Evolution of 2 TAG clusters of 15 genes
(average for 100 trees)



Application on Biological Data

Human-Rat olfactory receptor gene cluster

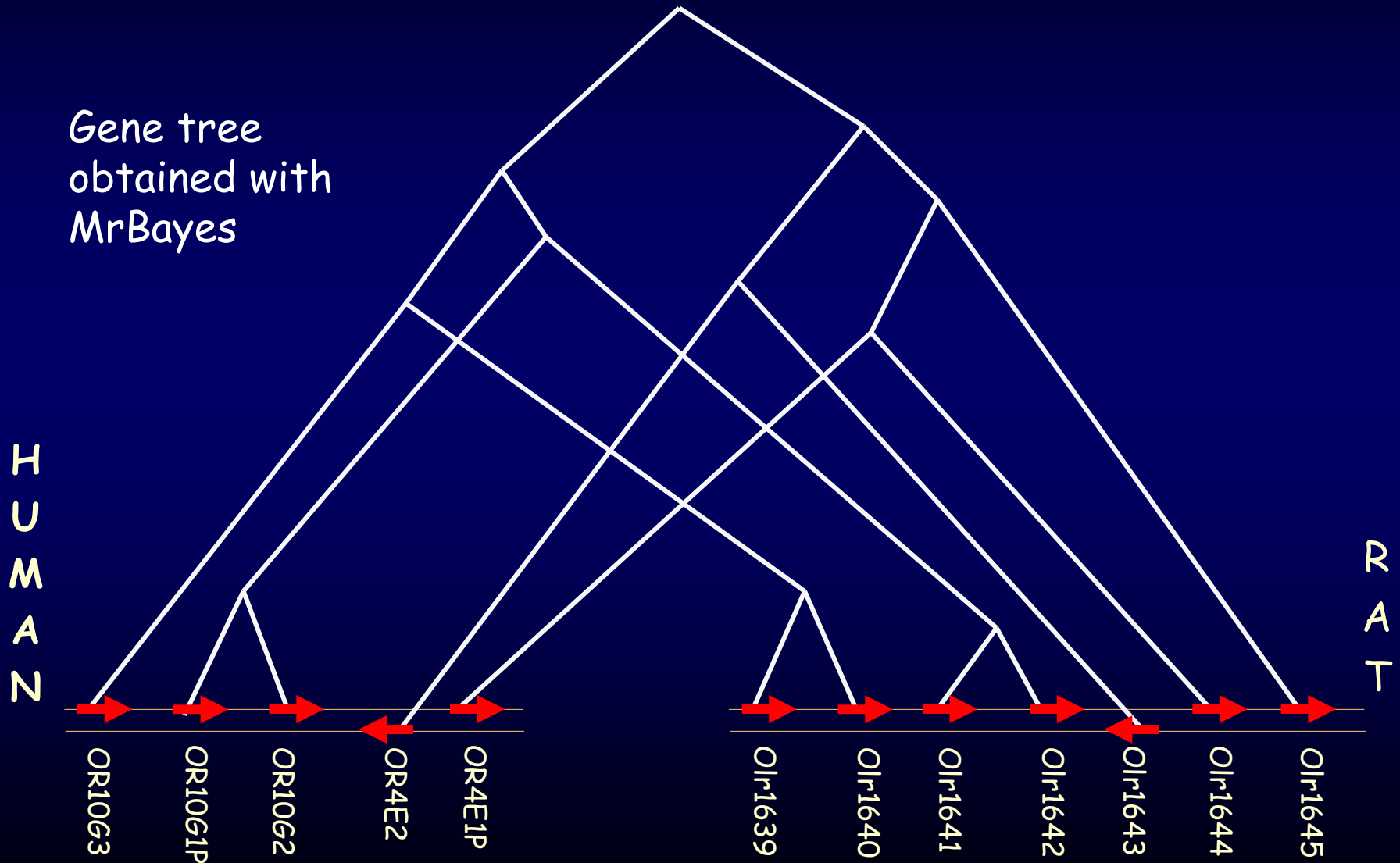
- Sequences / gene orders obtained from the HORDE database
- Alignment with ClustalW



Application on Biological Data

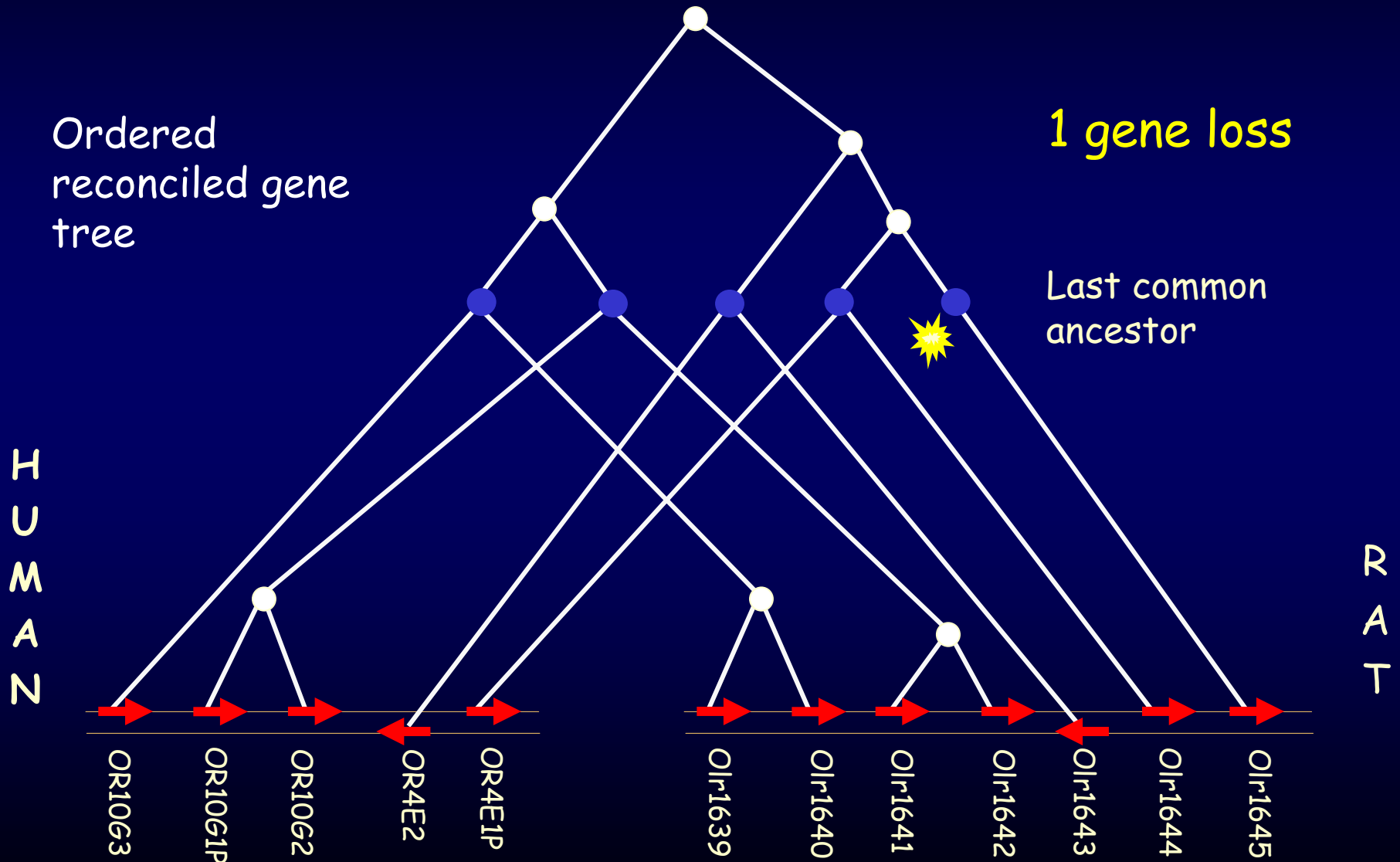
Human-Rat olfactory receptor gene cluster

Gene tree
obtained with
MrBayes



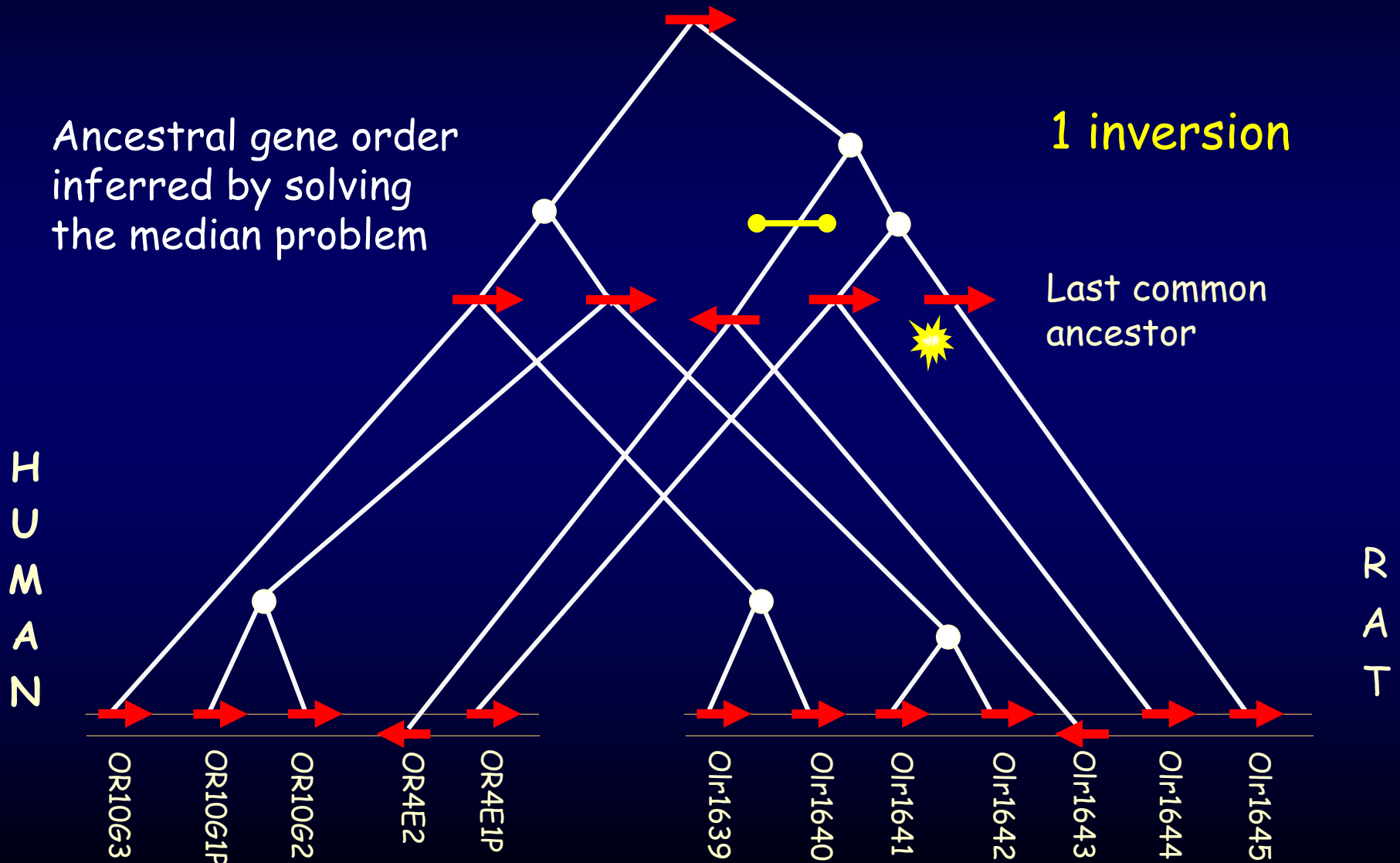
Application on Biological Data

Human-Rat olfactory receptor gene cluster



Application on Biological Data

Human-Rat olfactory receptor gene cluster



Conclusion

- We presented an evolutionary model which includes simple tandem duplication, gene loss, inversion and speciation
- A branch-and-bound algorithm to infer the ancestral gene orders in a species tree
- An example of a TAG cluster which is likely to have evolved in agreement with our model

Future Work

- ❑ Develop fast and accurate heuristics to handle larger TAG clusters
- ❑ Extend the model to multiple duplication (challenging problem)
- ❑ Study how the algorithm responds to model deviations (i.e. multiple duplication)
- ❑ Extensive application of our algorithm on biological data to validate the model and/or identify potential improvements (olfactory receptor genes in the HORDE database)

Acknowledgments

- ❑ Fond Québécois de la Recherche sur la Nature et les Technologies
- ❑ Natural Sciences and Engineering Research Council of Canada
- ❑ Canadian institutes of Health Research