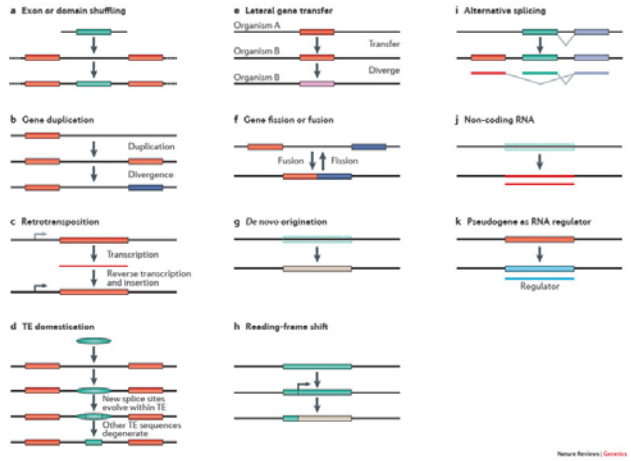
Reconciliation with Duplication, Transfer and Loss

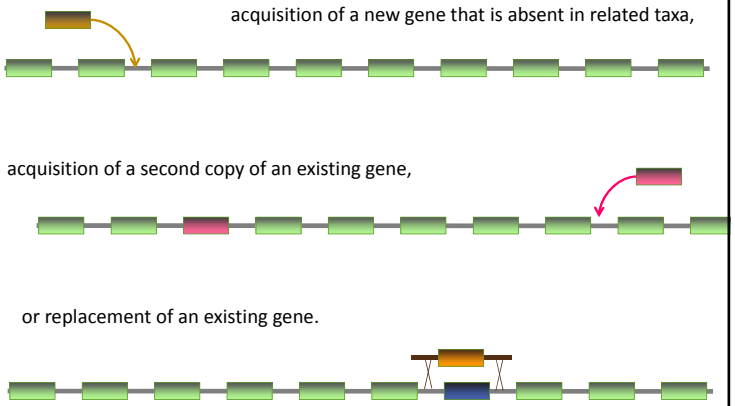
03-327/727 Lecture 28

Maureen Stolzer



Chen, Krinsky, Long, NRG, 2013

Horizontal transfer events can result in



Koonin et al. ARM (2001) **55** 709-4

Outline

I. Why infer transfers?

II. Reconciliation with transfers

III. Demo

Horizontal Gene Transfer

- HGT in Eukaryotes
- HGT in Bacteria
- How to detect HGT

Classical view

- Never
- Occasionally

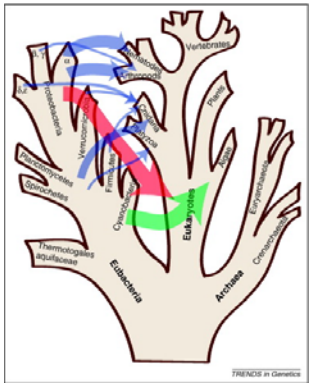
Horizontal Gene Transfer

- HGT in Eukaryotes
- HGT in Bacteria
- How to detect HGT

Emerging view

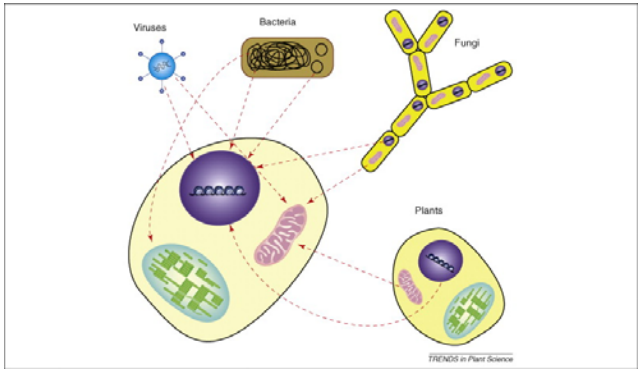
- More and more examples are coming to light
- Rampant!

Horizontal transfer involving animals



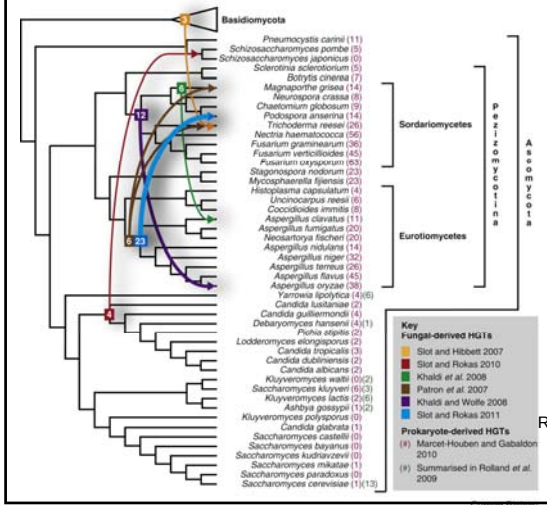
Dunning Hotopp, TIG, 2011

Horizontal transfer involving plants



Bock, Trends Plant Scince, 2009

Horizontal transfer within fungi



Richards, Current Biol. 2011

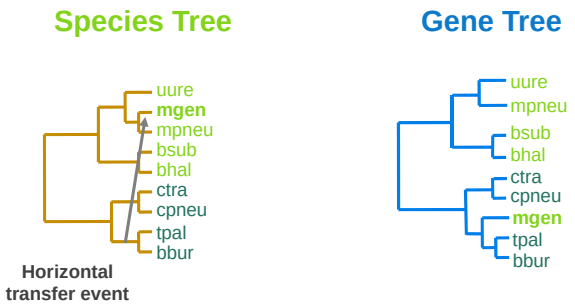
Outline

I. Why infer transfers?

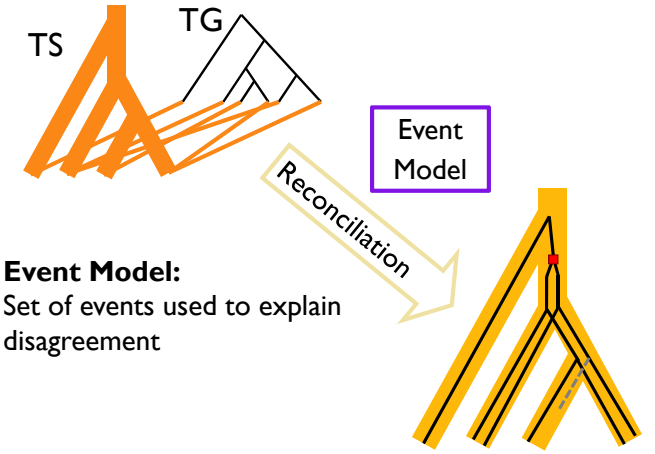
II. Reconciliation with transfers

III. Demo

The signature of horizontal transfer is incongruence between the species tree and the gene tree

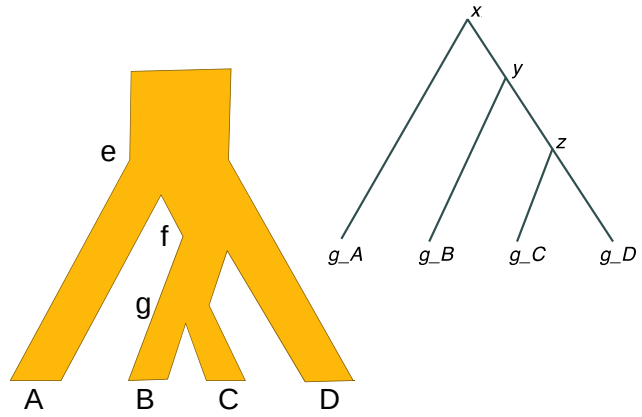


General Reconciliation



Event Model:
Set of events used to explain disagreement

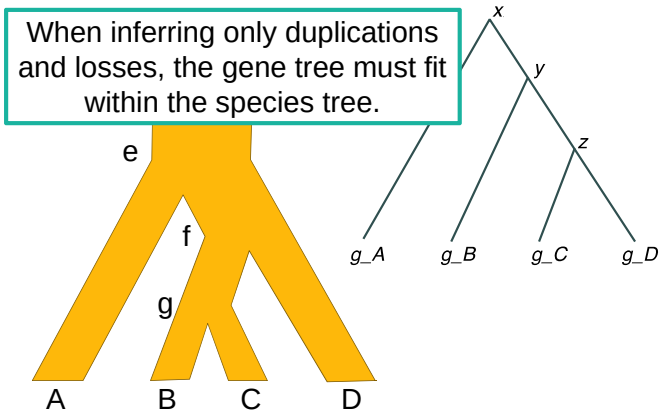
Reconciliation for the Duplication-Transfer Model



13

Reconciliation for the Duplication-Transfer Model

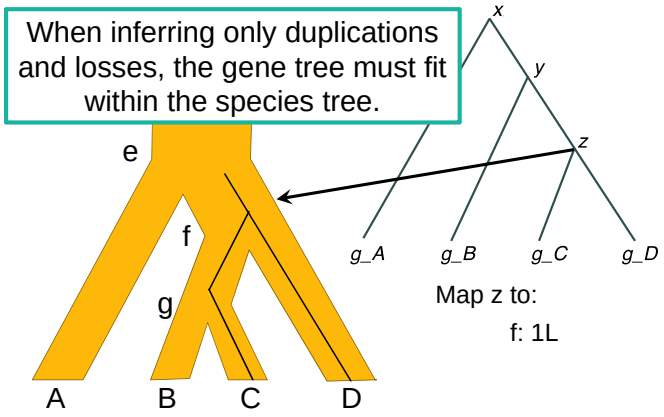
When inferring only duplications and losses, the gene tree must fit within the species tree.



14

Reconciliation for the Duplication-Transfer Model

When inferring only duplications and losses, the gene tree must fit within the species tree.

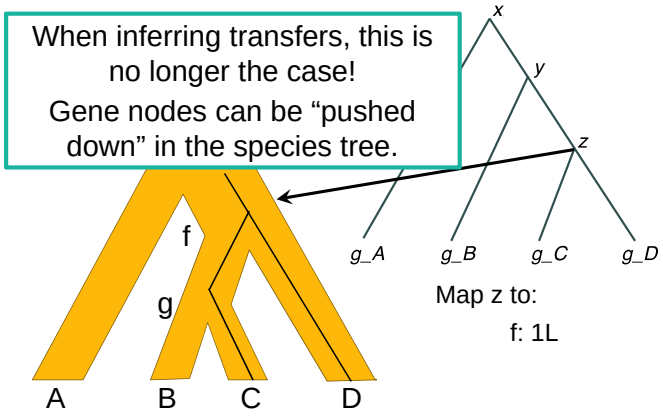


15

Reconciliation for the Duplication-Transfer Model

When inferring transfers, this is no longer the case!

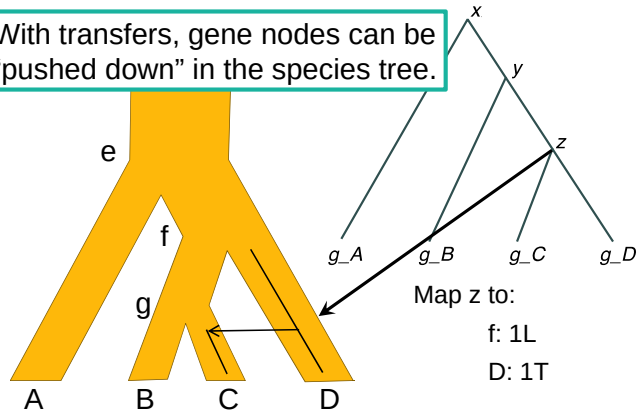
Gene nodes can be “pushed down” in the species tree.



16

Reconciliation for the Duplication-Transfer Model

With transfers, gene nodes can be "pushed down" in the species tree.

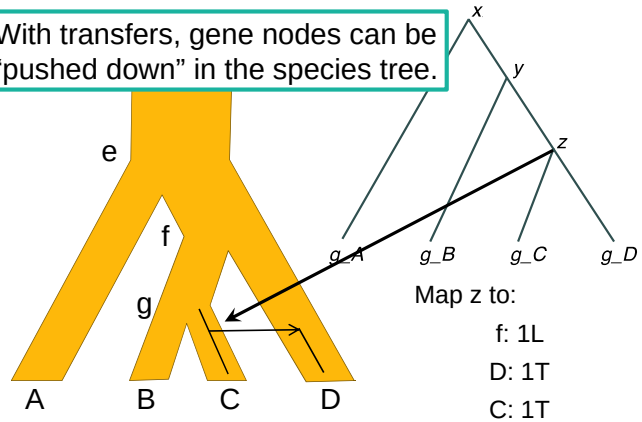


Map z to:
f: 1L
D: 1T

17

Reconciliation for the Duplication-Transfer Model

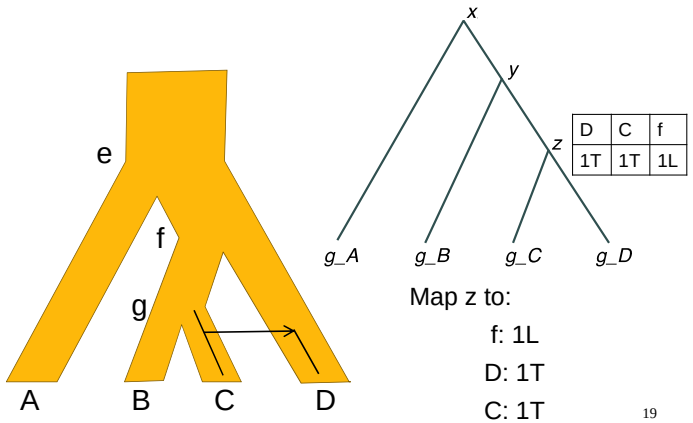
With transfers, gene nodes can be "pushed down" in the species tree.



Map z to:
f: 1L
D: 1T
C: 1T

18

Possible mappings and costs are stored in a table at each gene node.



D	C	f
1T	1T	1L

Map z to:
f: 1L
D: 1T
C: 1T

19

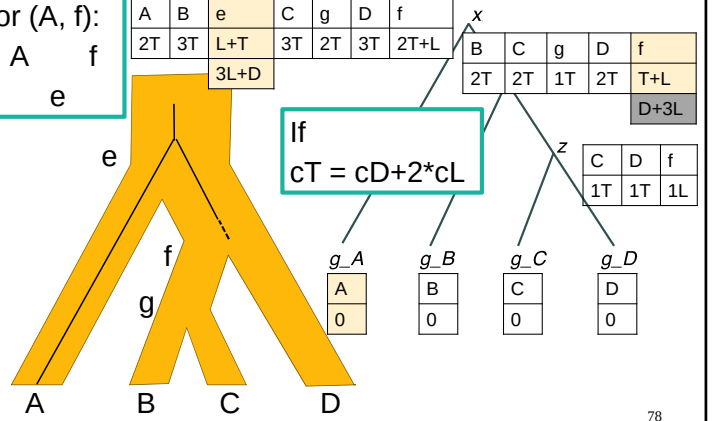
Possible mappings and costs are stored in a table at each gene node.

For (A, f):
A f
e

A	B	e	C	g	D	f
2T	3T	L+T	3T	2T	3T	2T+L
		3L+D				

B	C	g	D	f
2T	2T	1T	2T	T+L
				D+3L

If
cT = cD+2*cL

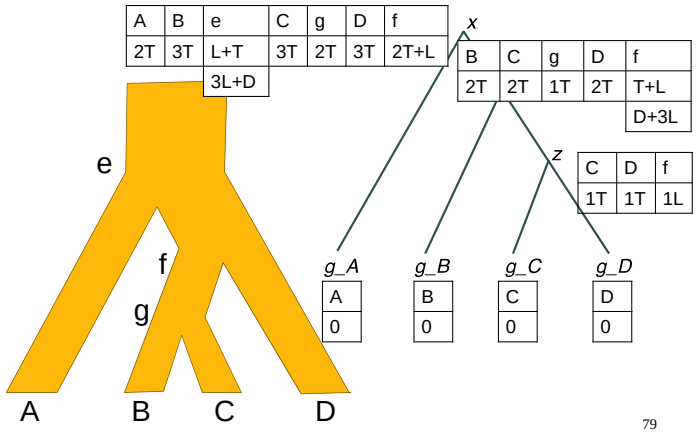


C	D	f
1T	1T	1L

g_A	g_B	g_C	g_D
A	B	C	D
0	0	0	0

78

Where we left off…



Transfer Algorithm

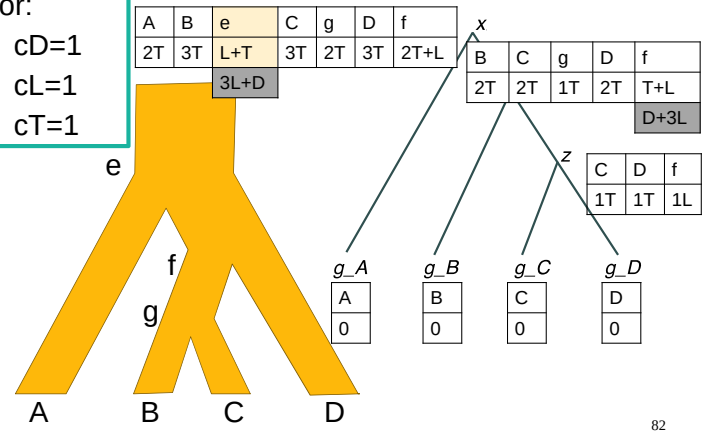
- Pass 1: Consider all possible species mappings at each node in T_G. Store this information in a table; only keep the lowest cost for each species.
- Up next…
  - ✧ Getting a solution
  - ✧ Multiple optimal solutions
  - ✧ Temporal feasibility
  - ✧ Examples and Demos

Pass 2

- Pass 2 – Traceback: Traverse the gene tree from root to leaves to construct a solution.
  - Pull information from each table.
  - Solutions in parent table point to entries in children
- Repeat for each optimal solution.

Pass 2: Traceback

For:
cD=1
cL=1
cT=1



Pass 2: Traceback

For:
cD=1
cL=1
cT=1

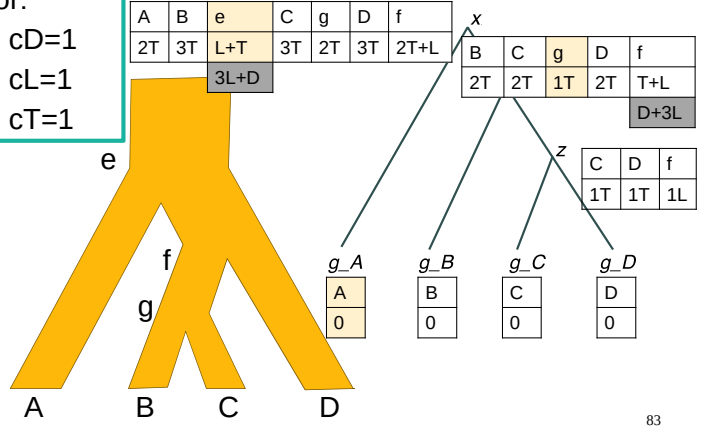
A	B	e	C	g	D	f
2T	3T	L+T	3T	2T	3T	2T+L
		3L+D				

x

B	C	g	D	f
2T	2T	1T	2T	T+L
				D+3L

z

C	D	f
1T	1T	1L



g_A	g_B	g_C	g_D
A	B	C	D
0	0	0	0

83

Pass 2: Traceback

For:
cD=1
cL=1
cT=1

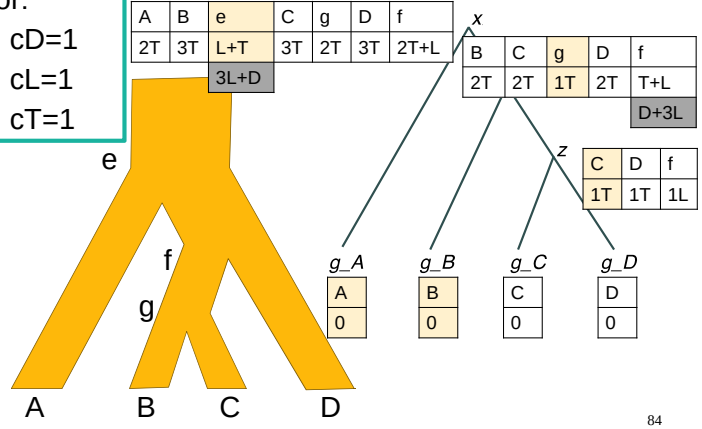
A	B	e	C	g	D	f
2T	3T	L+T	3T	2T	3T	2T+L
		3L+D				

x

B	C	g	D	f
2T	2T	1T	2T	T+L
				D+3L

z

C	D	f
1T	1T	1L



g_A	g_B	g_C	g_D
A	B	C	D
0	0	0	0

84

Pass 2: Traceback

For:
cD=1
cL=1
cT=1

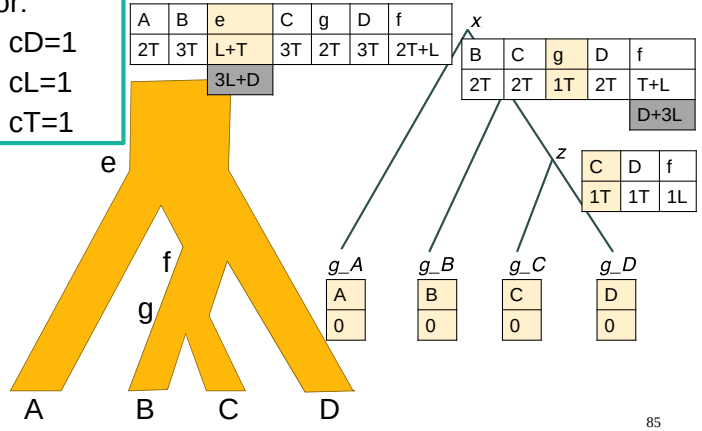
A	B	e	C	g	D	f
2T	3T	L+T	3T	2T	3T	2T+L
		3L+D				

x

B	C	g	D	f
2T	2T	1T	2T	T+L
				D+3L

z

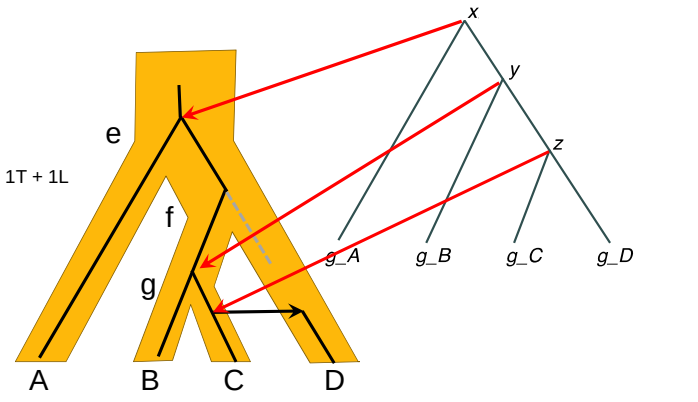
C	D	f
1T	1T	1L



g_A	g_B	g_C	g_D
A	B	C	D
0	0	0	0

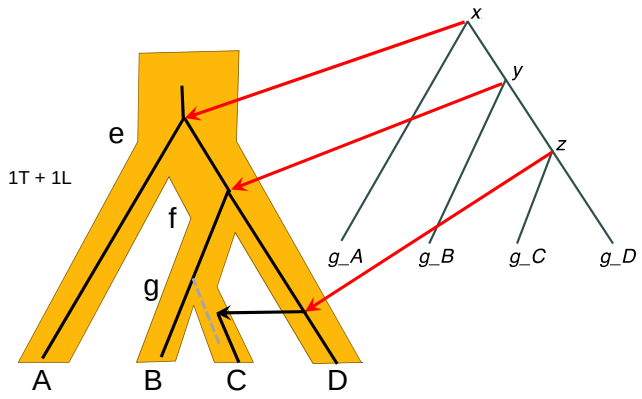
85

Solution



86

Alternative Solution

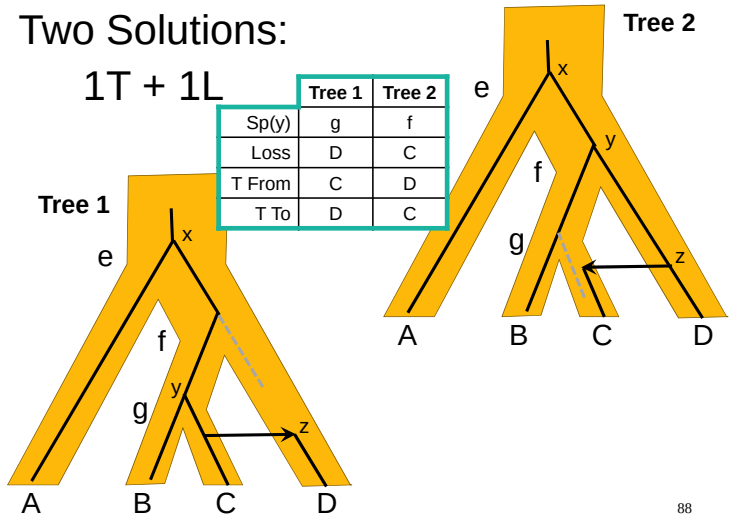


87

Two Solutions:

1T + 1L

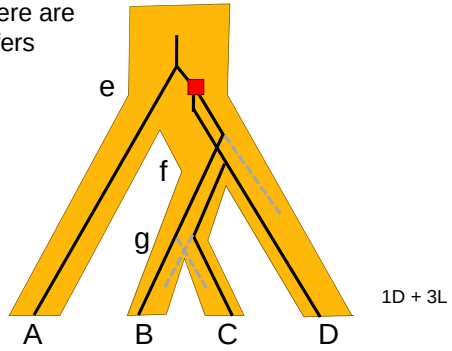
	Tree 1	Tree 2
Sp(y)	g	f
Loss	D	C
T From	C	D
T To	D	C



88

Another Solution:

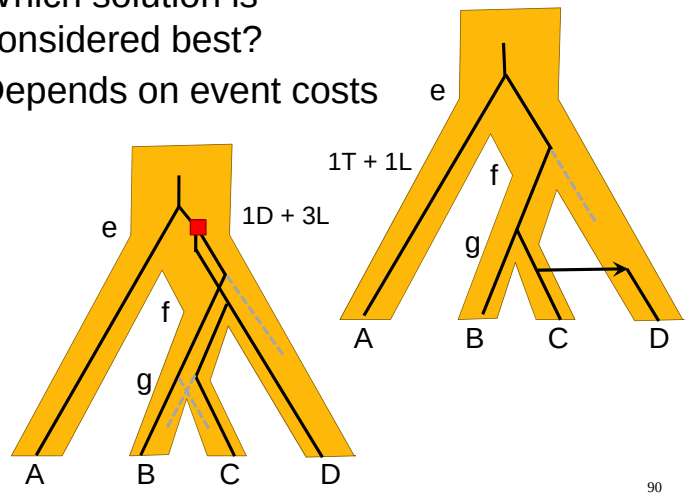
When there are no transfers inferred:



89

Which solution is considered best?

Depends on event costs

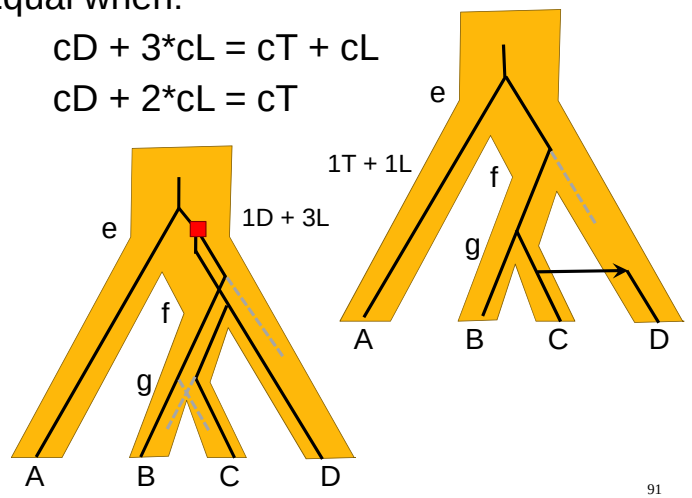


90

Equal when:

cD + 3*cL = cT + cL

cD + 2*cL = cT



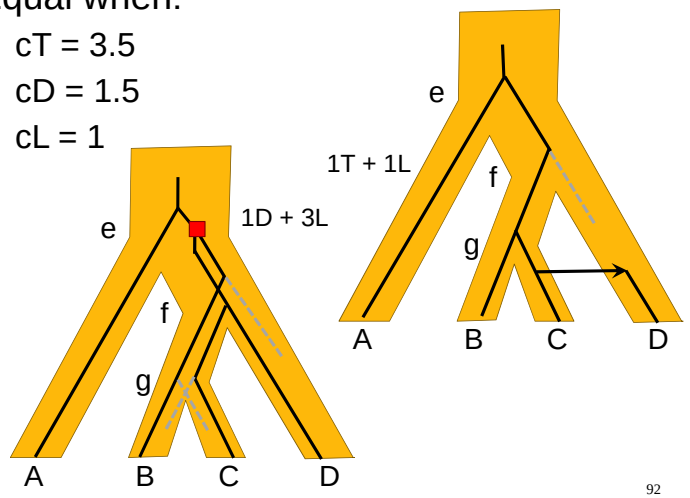
91

Equal when:

cT = 3.5

cD = 1.5

cL = 1



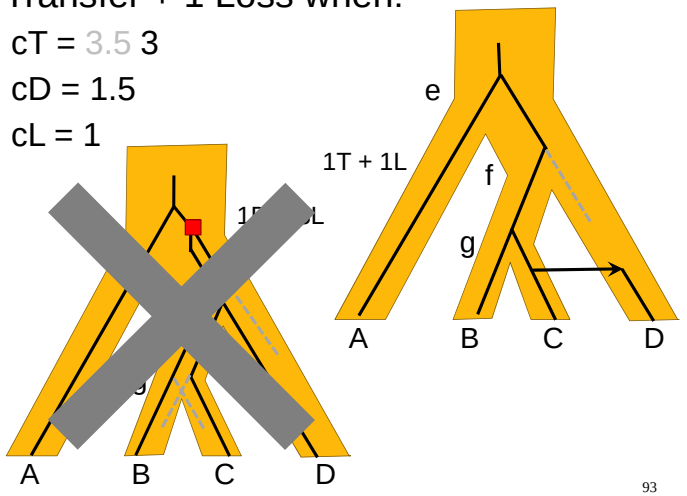
92

1 Transfer + 1 Loss when:

cT = 3.5 3

cD = 1.5

cL = 1



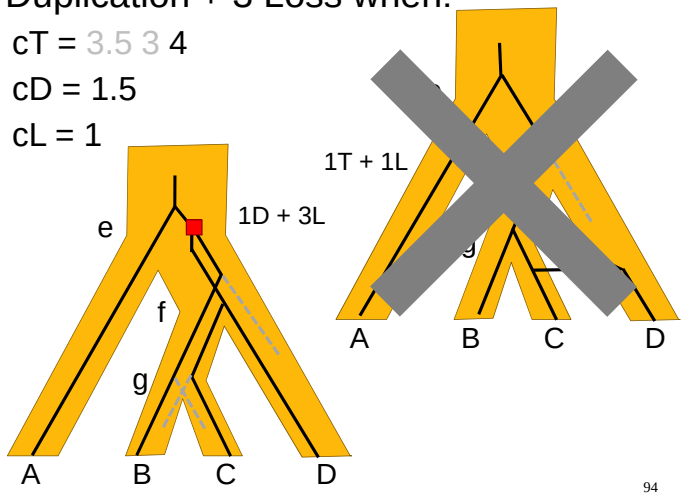
93

1 Duplication + 3 Loss when:

cT = 3.5 3 4

cD = 1.5

cL = 1



94

Three Passes

- Pass 1 – Dynamic Program: Traverse T_G in postorder. At each node g, fill in table:
- Pass 2 – Traceback: Traverse TG in pre-order to construct a solution. Repeat for each optimal solution.
- Pass 3 – Cycle checking: For each given solution, check temporal feasibility.

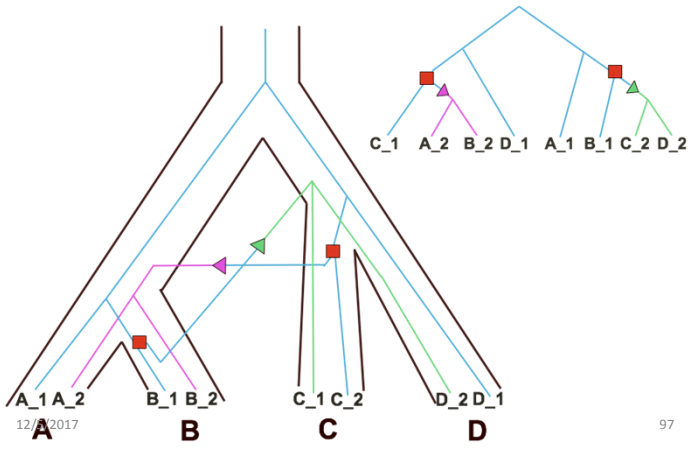
95

Temporal Feasibility?

- Trees imply temporal constraints:
  - Ancestral species pre-date their descendants
  - Ancestral genes pre-date their descendants
- Transfers add a new constraint!
  - Donor and recipient species must have co-existed at some point
- If more than 1 transfer inferred, these constraints could conflict

96

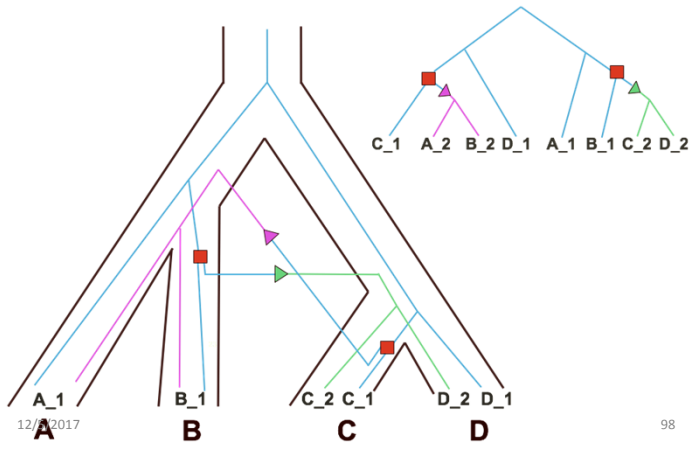
Temporal Infeasibility



12/5/2017

97

Temporal Infeasibility



12/5/2017

98

Pass 3

- Pass 3 –For each given solution, check temporal feasibility.
- If a cycle is found, we do not attempt to fix it.
- Only a problem if *all* optimal solutions have a cycle.
- Finding an optimal solution without cycles is proven to be NP-complete

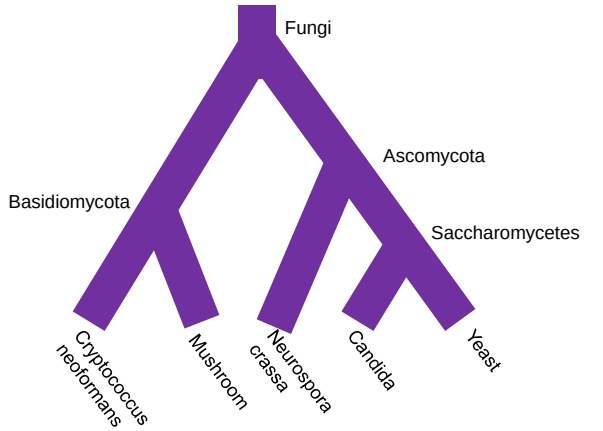
99

Gene Tree-Species Tree Reconciliation with Duplications, Transfers, and Losses

- More than one optimal event history.
- Mathematically optimal histories may be biologically infeasible due to temporal constraints
- No constructive algorithm for finding feasible, optimal histories
- Solution: Generate candidate optimal histories and test them for feasibility.

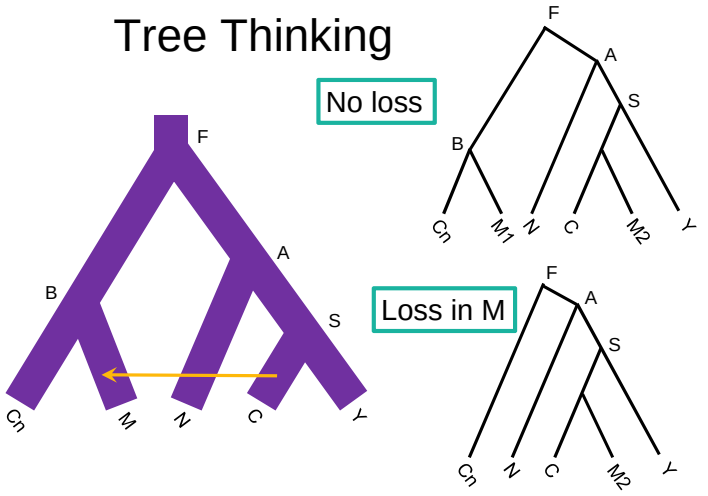
100

Tree Thinking



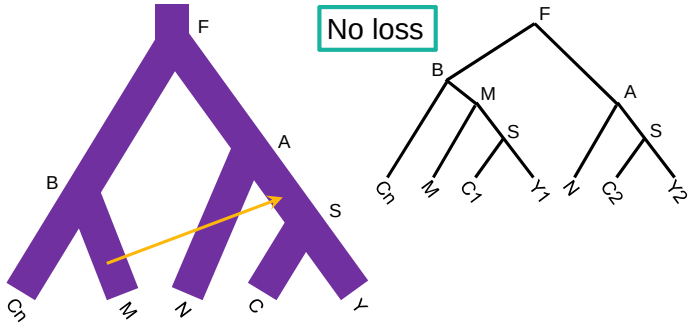
101

Tree Thinking



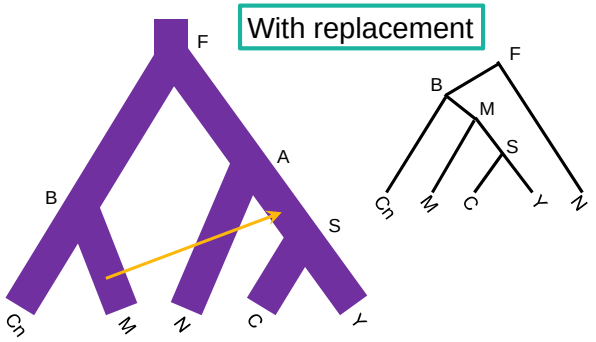
102

Tree Thinking



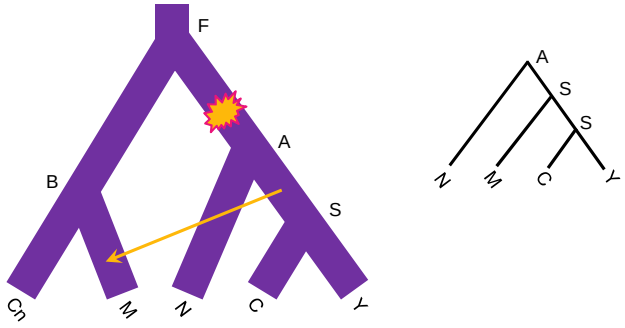
103

Tree Thinking



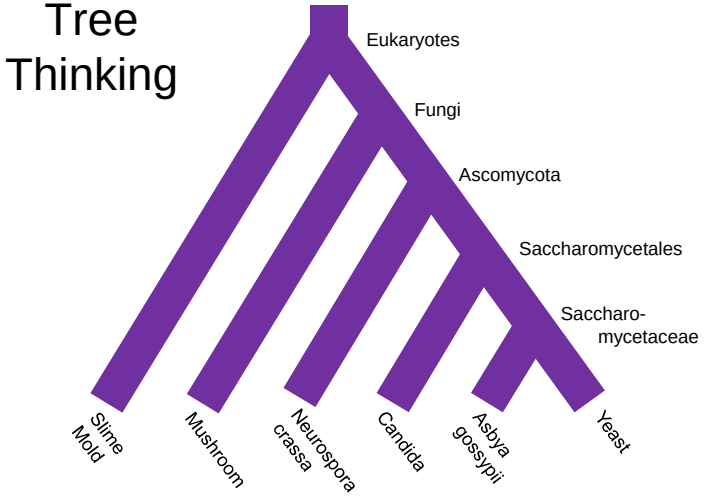
104

Tree Thinking



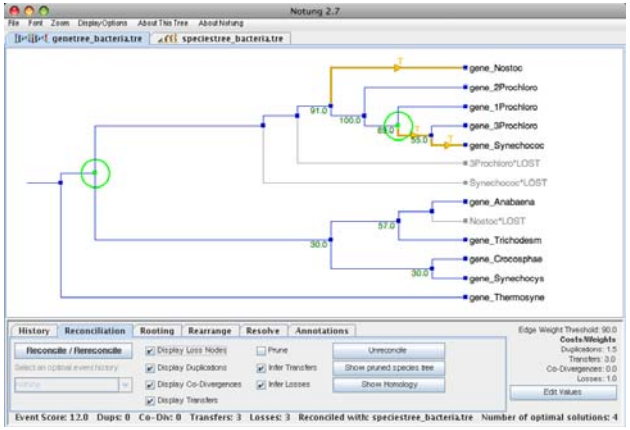
105

Tree Thinking



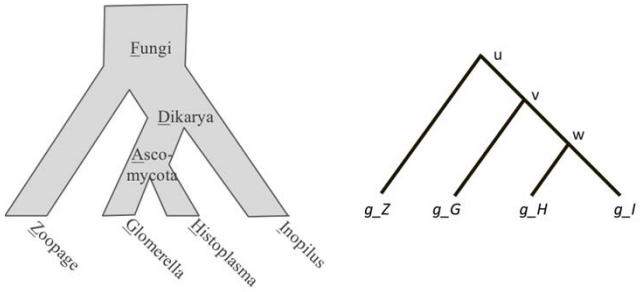
106

Implemented in Notung 2.7

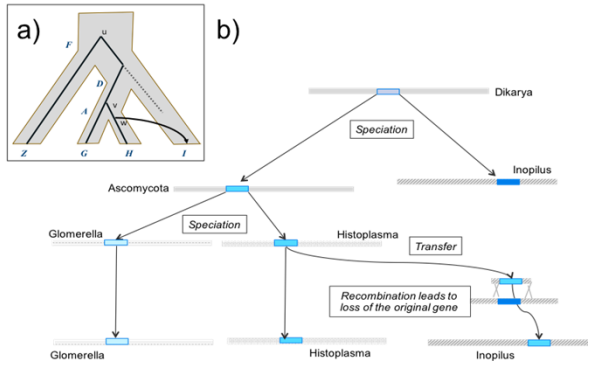


http://www.cs.cmu.edu/~durand/Notung

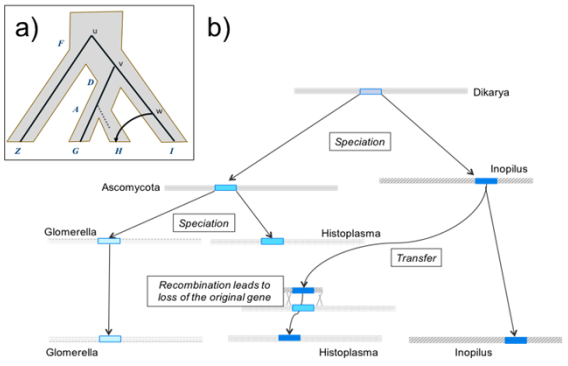
Demo Example from Lecture



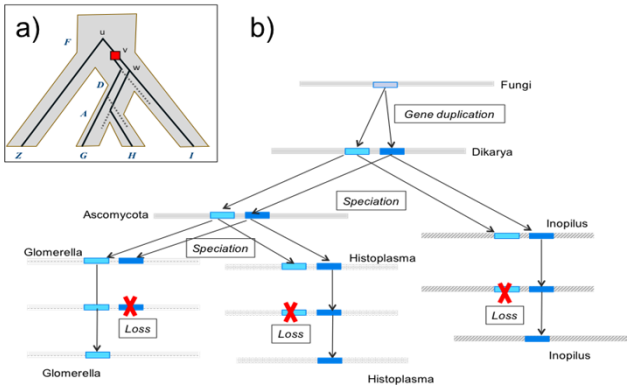
Demo Example from Lecture



Demo Example from Lecture



Demo Example from Lecture



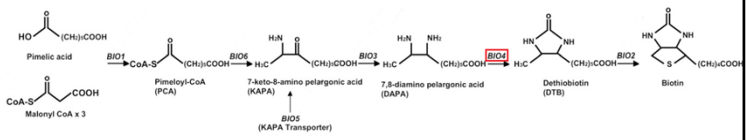
111

Example: BIO4 transfer between Fungi and Bacteria

Plants and some fungi: a bi-functional enzyme (steps 3 and 4 in the Biotin pathway):
1) 7,8-diaminopelargonic acidsynthase (DAPAS) and
2) dethiobiotin synthetase (DTBS)

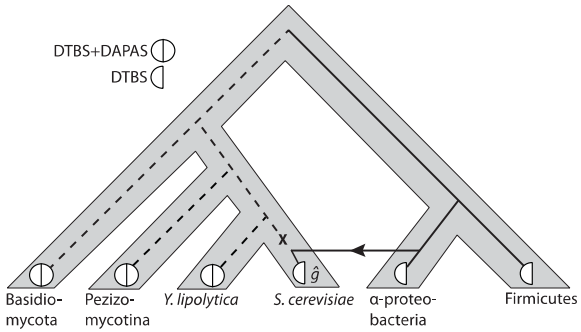
Bacteria: not bi-functional:
1) (DTBS) – step 4 in pathway

Yeast: not bi-functional:
1) (DTBS) – step 4 in pathway



112

Example: BIO4 transfer between Fungi and Bacteria



113