Regarding
How to deal with simultaneous moves of multiple gap-blocks (or sequence-blocks).

By Kiyoshi Ezawa

From: March 16 (Fri), 2018,

Till: June 11 (Mon), 2018.

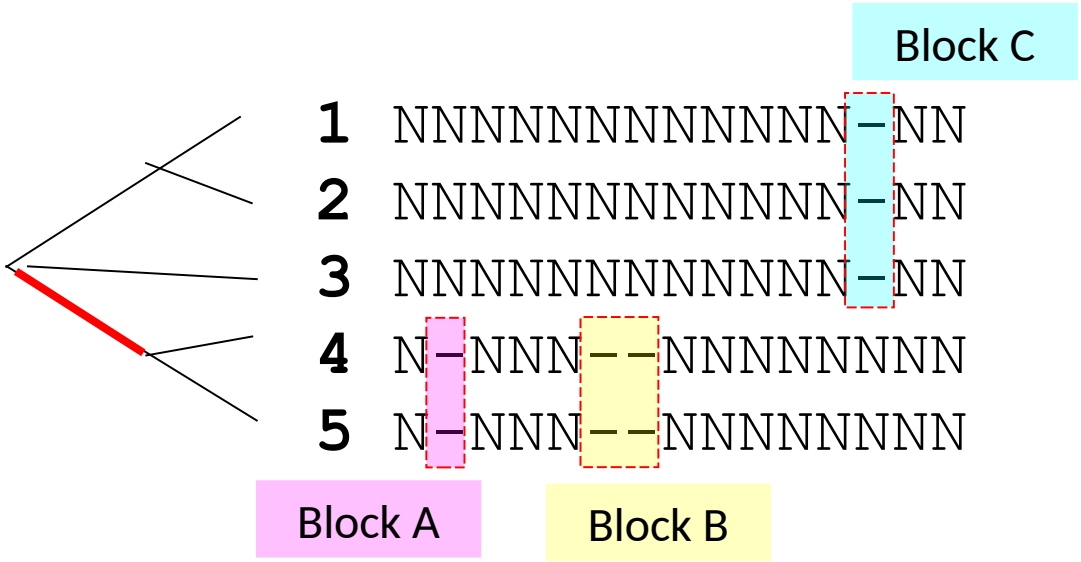
© 2018 Kiyoshi Ezawa. **Open Access** This file is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author (K. Ezawa) and the source (https://www.bioinformatics.org/ftp/pub/anex/Documents/Blueprints/Detailed_Components/simultaneous_moves_of_multiple_blocks.draft1_CC4.pdf), provide a link to the Creative Commons license (above), and indicate if changes were made.

Case Study

- Let us first consider a number of concrete examples,

  and try to deduce some principles.

Case 1 (from Figure 3B of "blueprint1_ANEX.xxxx.pdf")



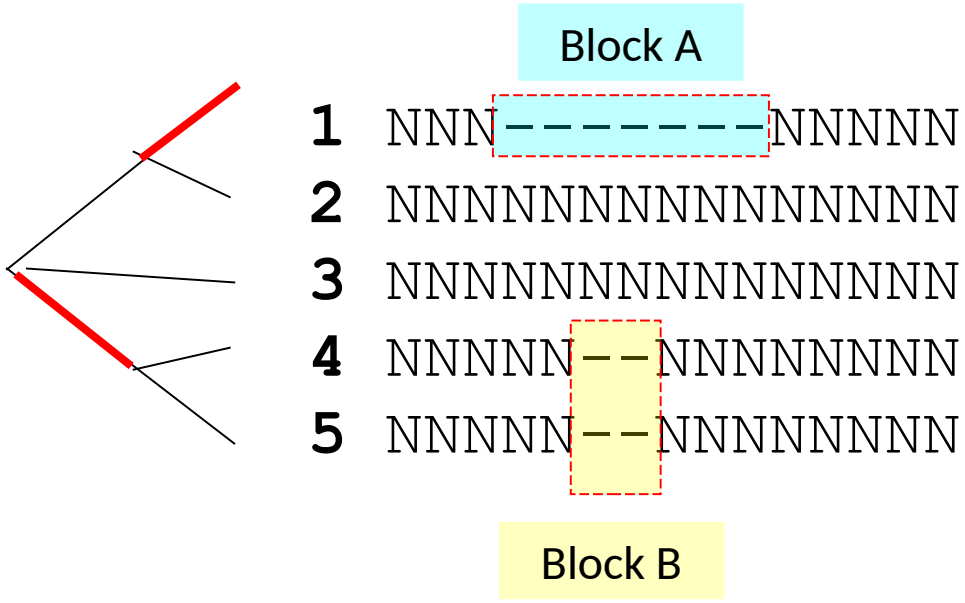
In this case, each indel block (A, B or C) resulted from a single insertion/deletion. Therefore, the move of each gap-block (or sequence-block) coincides with the move of the corresponding indel bock.

For example, block B can move with a negligible impact on the indel component of the alignment probability, as long as it remains between blocks A and C (and at least one gapless column separates it from each of blocks A & C).

If block A and/or block C move(s), the range of the "free move" by block B also changes accordingly.

Besides, block B can also "freely move" after it passes block C, until it reaches yet another block.

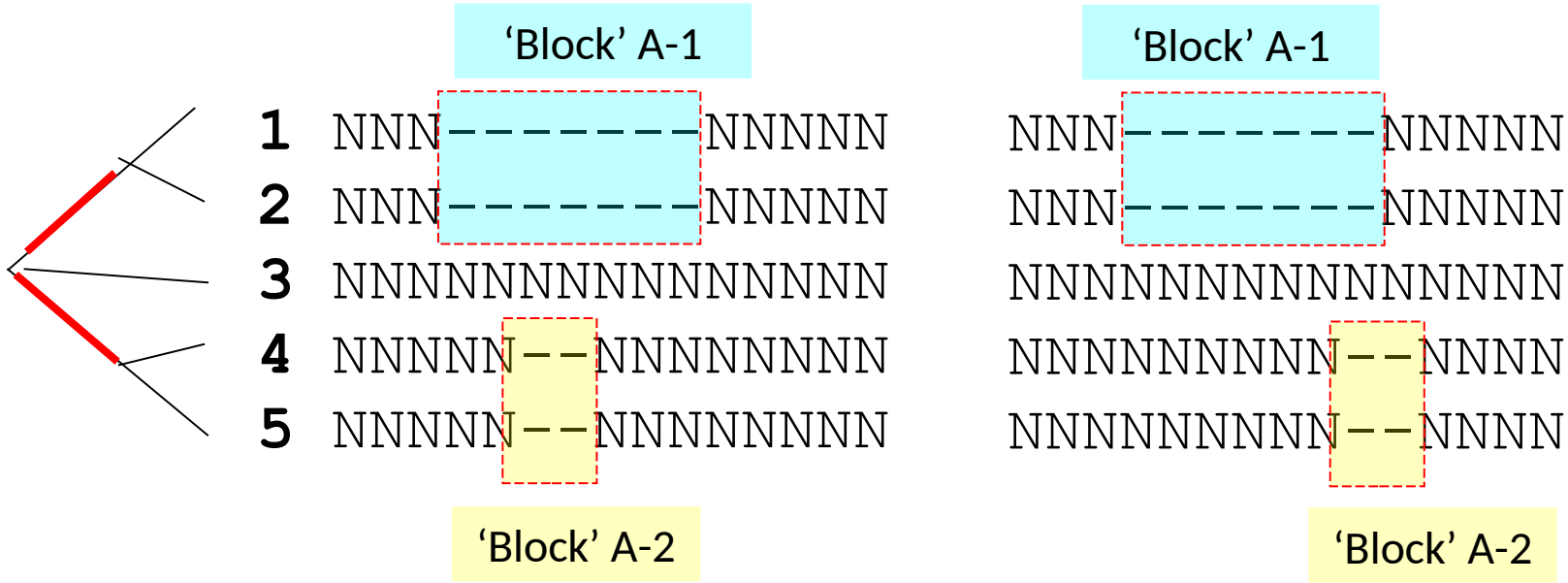
Case 2 (from Figure 3C of "blueprint1_ANEX.xxxx.pdf")



In this case, too, each indel block (A, or B) resulted from a single insertion/deletion. Therefore, the move of each gap-block (or sequence-block) coincides with the move of the corresponding indel bock.

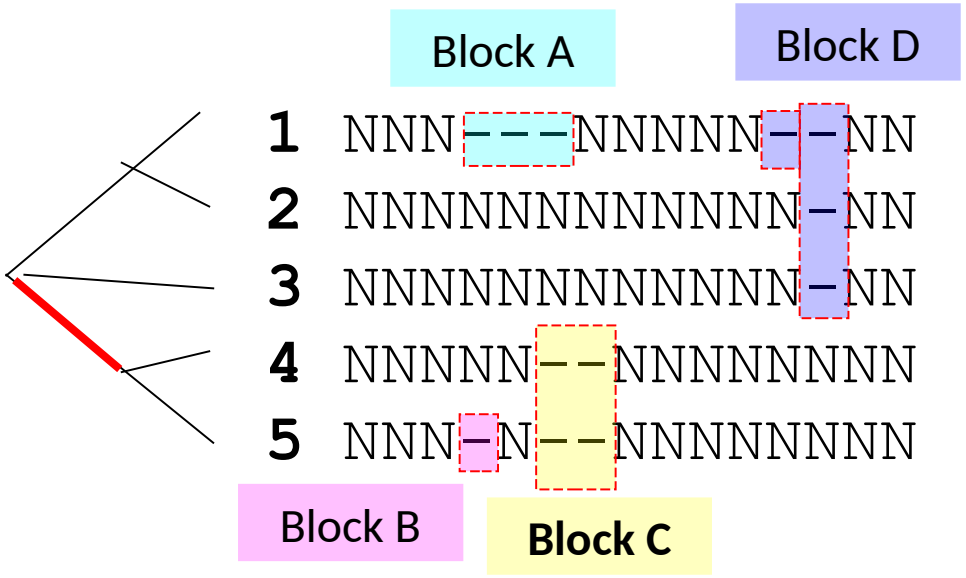
Also in this case, too, because blocks A and B are virtually independent of each other, they can move almost freely, with a negligible impact on the indel component probability, until either of them reaches a yet another block.

Case 3 (from Figure 3D of "blueprint1_ANEX.xxxx.pdf")



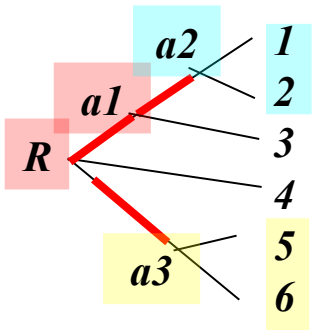
In this case, the two 'block's (A-1 and A-2) are NOT independent of each other.
Rather they constitute a single indel block (see, e.g., the MSA on the left).
As long as the topology of this indel block remains unchanged,
the two 'block's can move almost freely, with a negligible impact on the indel component of the alignment probability.
When the topology changes (as in the right MSA), the indel component also changes.
After that, however, the indel component remain almost unchanged again, as long as the topology remains unchanged.
Finally, the two 'blocks' become independent of each other, forming two indel blocks.

Case 5 (from Figure 3A of "blueprint1_ANEX.xxxx.pdf")



Here, we will pay attention to block C.
It overlaps with block A, but they are independent of each other.
So, as long as we consider the indel component, we can ignore the relative position between blocks C and A.
When getting close together, block B can interfere with block C, and block D can also interfere with it, though in different manners.
When they touch or overlap block C, the indel component must be re-calculated.
However, after either of them "passes through" block C, the indel component becomes nearly equal to the original again.
(The same applies even to the relative position between blocks B and D.)

Case 6 (from Figure S3A of "suppl_blueprint1_ANEX.xxxx.pdf")



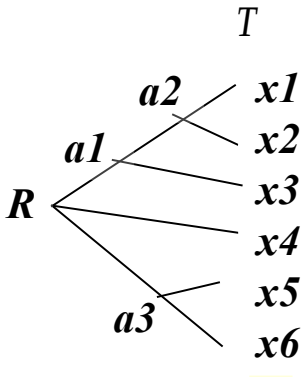
Block A1 Block A2

1		L	---	---	---	---	---	---	---	R
2		L	1	---	---	---	---	6	7	R
3		L	1	2	3	4	5	6	7	R
4		L	1	2	3	4	5	6	7	R
5		L	1	2	---	---	---	---	7	R
6		L	---	---	---	---	5	6	7	R

Block C1 Block C2 Block C3

Here, block A (consisting of A1 and A2) and block C (consisting of C1, C2 and C3) are independent of each other, thus their entire moves have only negligible impacts on the indel component of the alignment probability.
Within block A, although relative moves between blocks A1 and A2 have only negligible impacts on the indel component as long as the topology of block A remain unchanged, their relative move could change the indel component if block A changes its topology.
Within block C, as well, relative moves between blocks C1, C2 and C3 do not significantly change the indel component if the block topology remain unchanged,
while a topology change significantly changes the indel component.

Case 7 (from Figure SS1A of "suppl_addendum.xxxx.pdf")



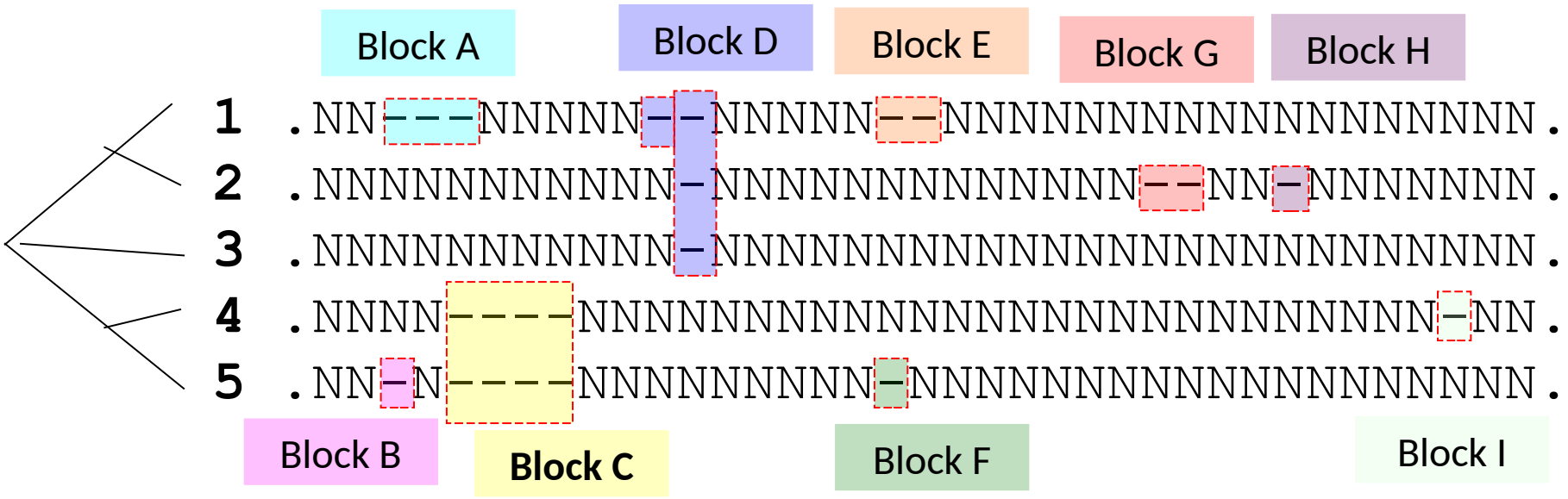
Block A1 Block A2 Block A3

x1		L	A	---	---	---	---	---	R
x2		L	A	B	---	D	E	---	R
x3		L	A	B	C	D	E	---	R
x4		L	A	B	C	D	E	F	R
x5		L	A	B	---	---	E	---	R
x6		L	---	---	---	D	E	---	R

Block C1 Block C2 Block C3

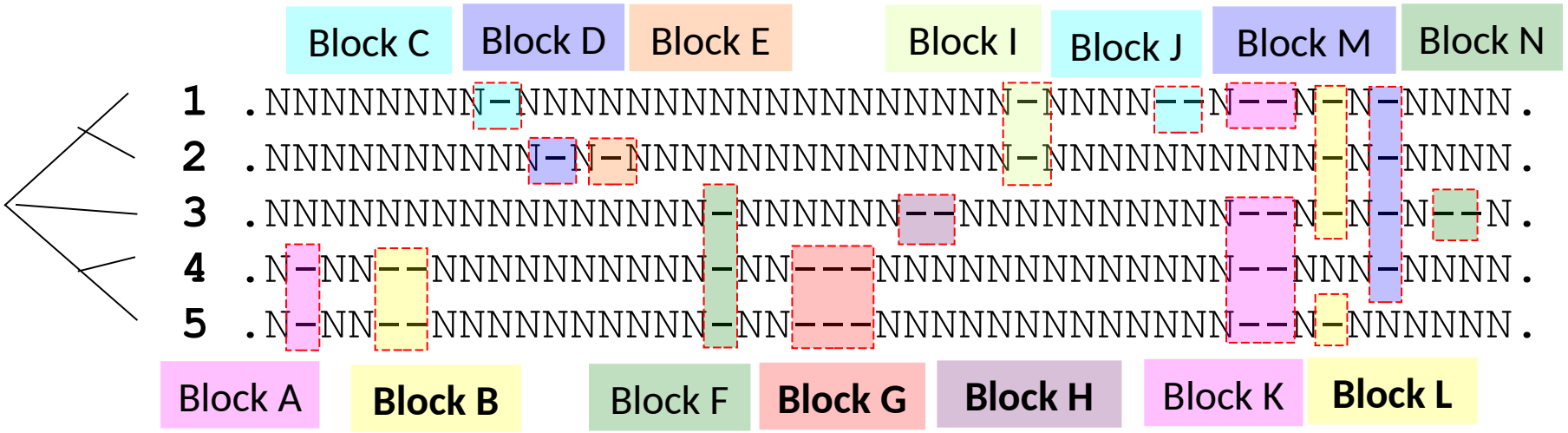
Here, block A (consisting of A1, A2 and A3) and block C (consisting of C1, C2 and C3) are independent of each other, **as long as they do not merge,**
thus their entire moves have only negligible impacts on the indel component of the alignment probability. A major difference between cases 6 and 7 is that case 7 has **an insertion-type block (A3)**. Thus, in the following, we discuss the impact of the move of **block A3**.
In conclusion, the move of block A3 will NOT significantly impact the indel component, as long as it is separated from block C2 by one or more column. (This applies also moves that would separate block A3 from all other block.)
When block A3 comes immediately next to block C2, however,
the indel component will significantly change, because another, deletion-based, history could also explain the gap configuration.

Case 8 (from Figure SS3 of "suppl2_blueprint_ANEX.xxxx.pdf")



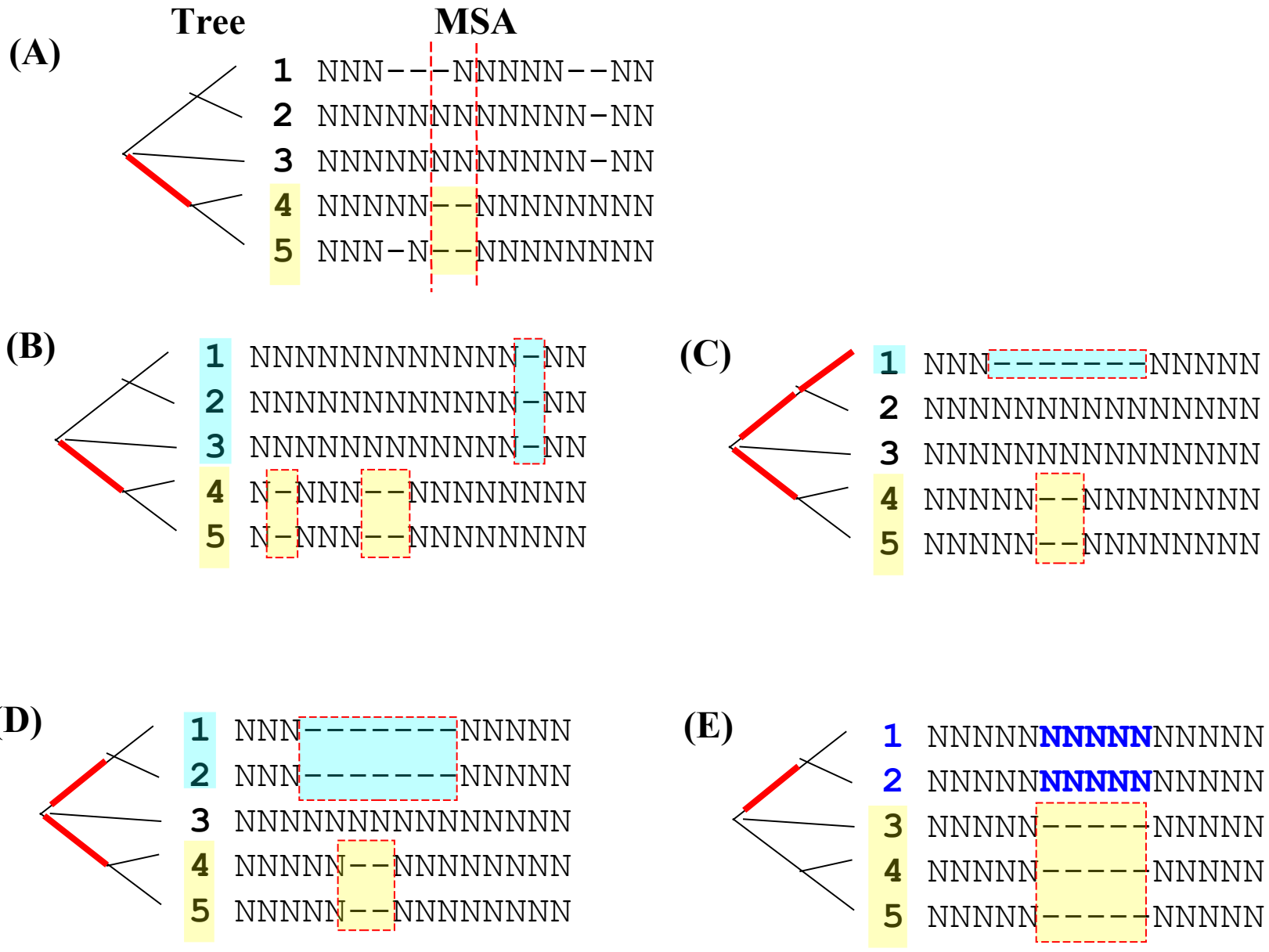
In this case, the relative positions among blocks seem almost non-influential to the indel component,
except the relations involving C, D and F, and the relations between E and G or H.
The move of **block C** could impact the indel component if it touches block B or block F (or I), and if it interacts with block D.
Likewise, the move of **block D** could impact the indel component if it touches block A or E, if it vertically aligns with block B or F (or I), or if it interacts with block C.
Meanwhile, the move of **block E** could impact the indel component if it vertically aligns with block G or H.

Case 9 (from Figure SS4A of "suppl2_blueprint_ANEX.xxxx.pdf")

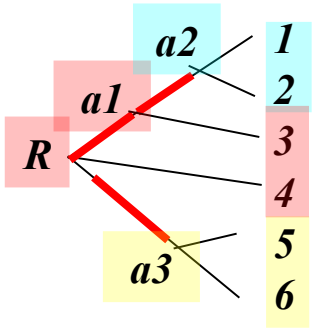


In this case, the relative positions between/among neighboring blocks are more likely to impact the indel component of the alignment probability than in case 8.

Case 10 (from Figure YYY of "zzz.xxxx.pdf")


(A)
Tree
MSA
1 NNN--NNNNN--NN
2 NNNNNNNNNNNN-NN
3 NNNNNNNNNNNN-NN
4 NNNNNN--NNNNNNNN
5 NNN-N--NNNNNNNN
(B)
1 NNNNNNNNNNNN-NN
2 NNNNNNNNNNNN-NN
3 NNNNNNNNNNNN-NN
4 N-NNN--NNNNNNNN
5 N-NNN--NNNNNNNN
(C)
1 NNN-------NNNNN
2 NNNNNNNNNNNNNNN
3 NNNNNNNNNNNNNNN
4 NNNNN--NNNNNNNN
5 NNNNN--NNNNNNNN
(D)
1 NNN-------NNNNN
2 NNN-------NNNNN
3 NNNNNNNNNNNNNNN
4 NNNNN--NNNNNNNN
5 NNNNN--NNNNNNNN
(E)
1 NNNNNNNNNNNNNNN
2 NNNNNNNNNNNNNNN
3 NNNNN-----NNNNN
4 NNNNN-----NNNNN
5 NNNNN-----NNNNN

(A) Tree



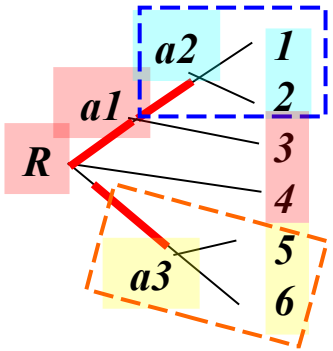
MSA

1		L	---	---	---	---	---	---	---	R
2		L	1	---	---	---	---	6	7	R
3		L	1	2	3	4	5	6	7	R
4		L	1	2	3	4	5	6	7	R
5		L	1	2	---	---	---	---	7	R
6		L	---	---	---	---	5	6	7	R

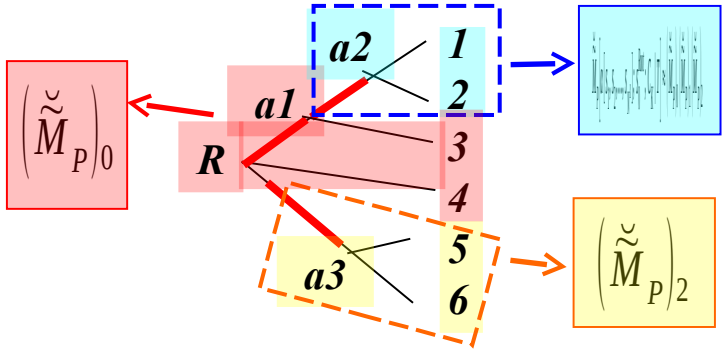
(B) Dollo parsimonious ancestral states

R		L	1	2	3	4	5	6	7	R
a1		L	1	2	3	4	5	6	7	R
a2		L	1	---	---	---	---	6	7	R
a3		L	1	2	---	---	5	6	7	R

(C) Partial indel history zones

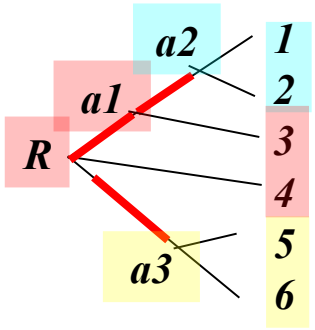


(D) Contributions to multiplication factor



$$\Rightarrow \quad \tilde{\tilde{M}}_P\left[\alpha\left[s_1, s_2, \ldots, s_{N^x}\right]; s_0^{Root}; C_K \mid T\right] \approx \left(\tilde{\tilde{M}}_P\right)_0 \left(\tilde{\tilde{M}}_P\right)_1 \left(\tilde{\tilde{M}}_P\right)_2$$

(A) Tree



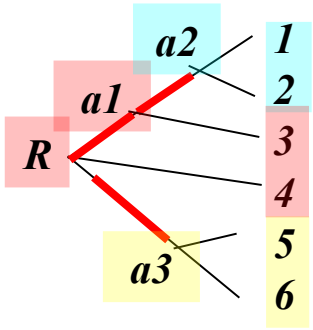
MSA (w/ null column)

1		L	---	---	---	---	---	---	---	---	R
2		L	1	---	---	---	---	---	6	7	R
3		L	1	2	3	---	4	5	6	7	R
4		L	1	2	3	---	4	5	6	7	R
5		L	1	2	---	---	---	---	---	7	R
6		L	---	---	---	---	---	5	6	7	R

(B) Ancestral states

R		L	1	2	3	x1	4	5	6	7	R
a1		L	1	2	3	x1	4	5	6	7	R
a2		L	1	---	---	---	---	---	6	7	R
a3		L	1	2	---	---	---	5	6	7	R

(A) Tree



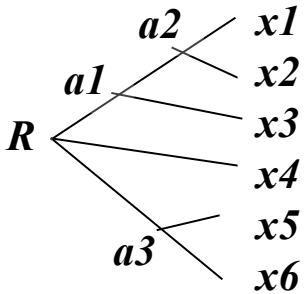
MSA (w/ null column)

1		L	---	---	---	---	---	---	---	---	R
2		L	1	---	---	---	---	---	6	7	R
3		L	1	2	3	---	4	5	6	7	R
4		L	1	2	3	---	4	5	6	7	R
5		L	1	2	---	---	---	---	---	7	R
6		L	---	---	---	---	---	5	6	7	R

(B) Ancestral states

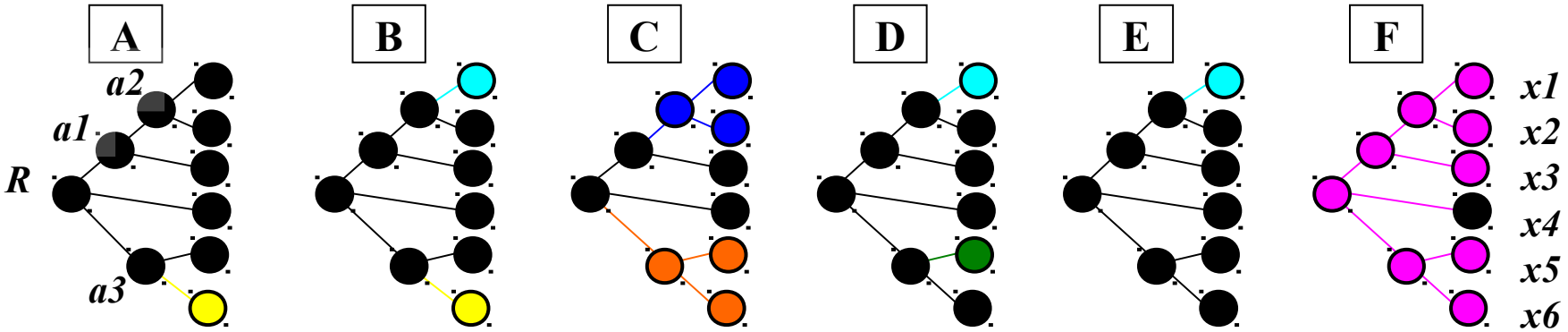
R		L	1	2	3	x1	4	5	6	7	R
a1		L	1	2	3	---	4	5	6	7	R
a2		L	1	---	---	---	---	---	6	7	R
a3		L	1	2	---	---	---	5	6	7	R

(A) Tree ( ) Gapped segment (C_K) in MSA

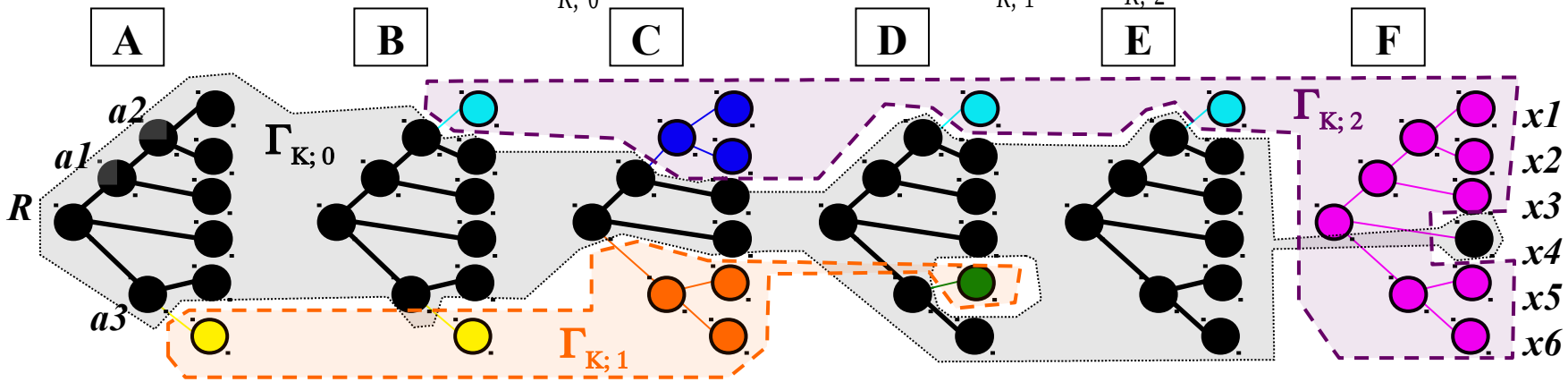


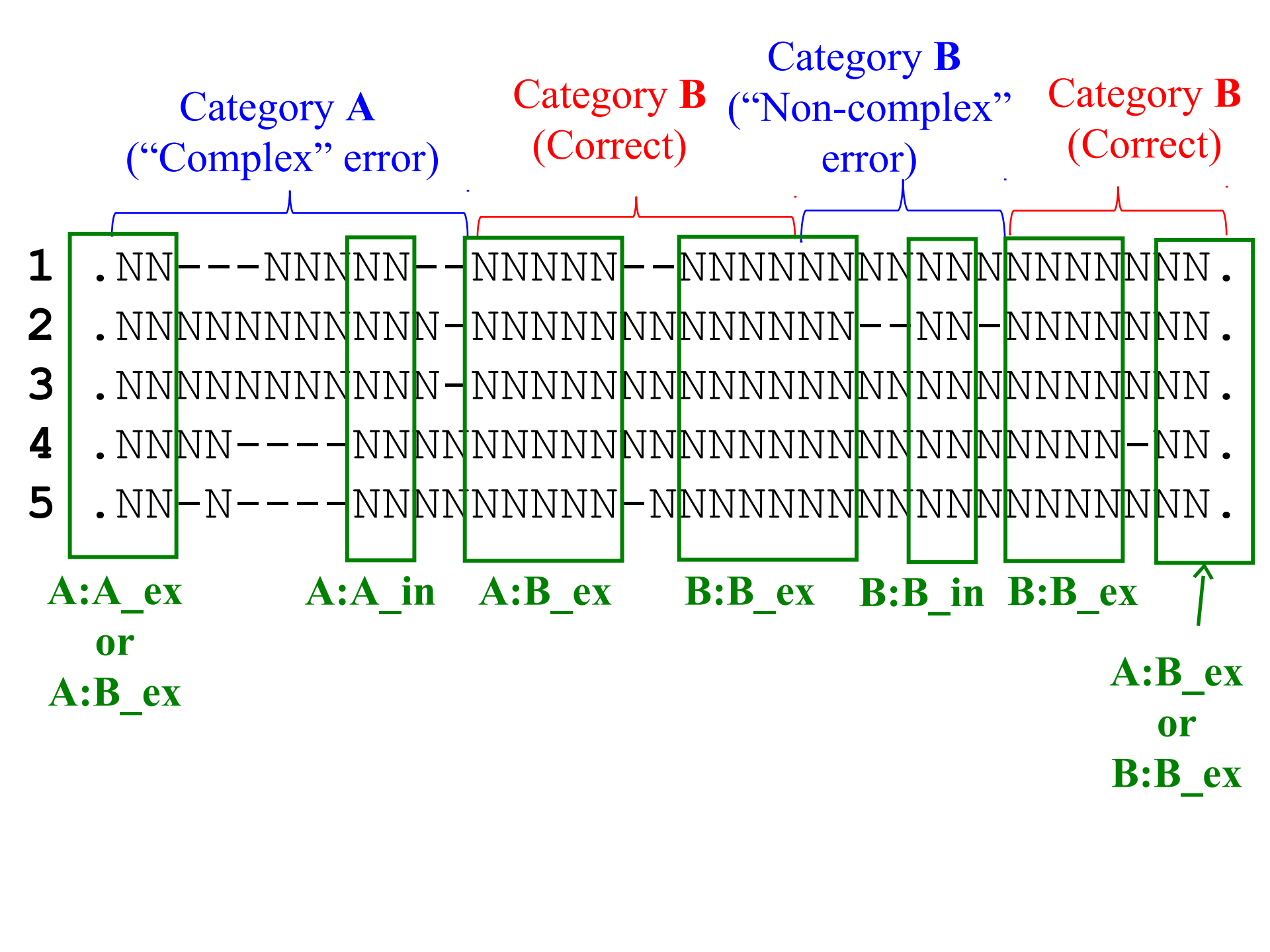
x1		L	A	---	---	---	---	---	R
x2		L	A	B	---	D	E	---	R
x3		L	A	B	C	D	E	---	R
x4		L	A	B	C	D	E	F	R
x5		L	A	B	---	---	E	---	R
x6		L	---	---	---	D	E	---	R

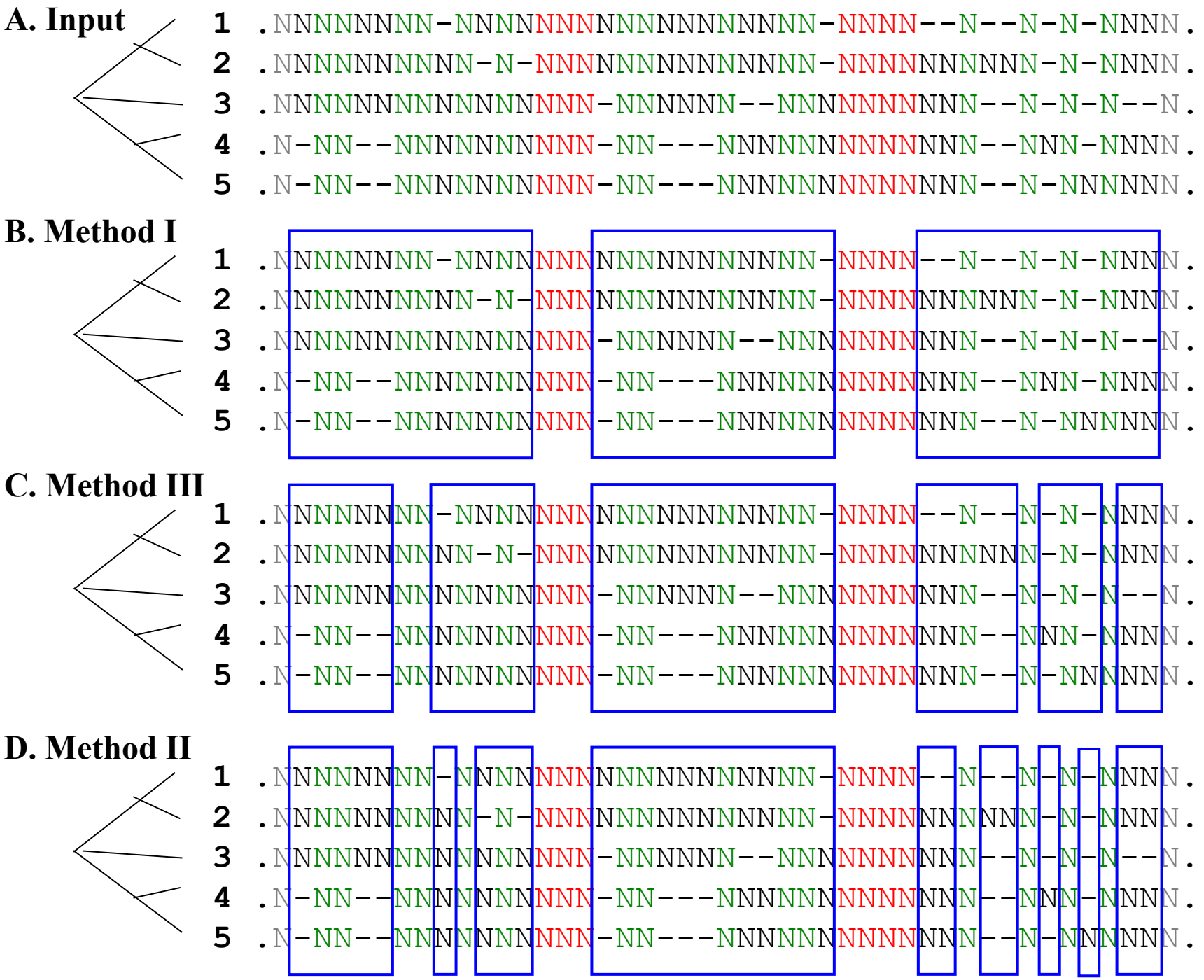
(B) Dollo parsimonious (gap) states

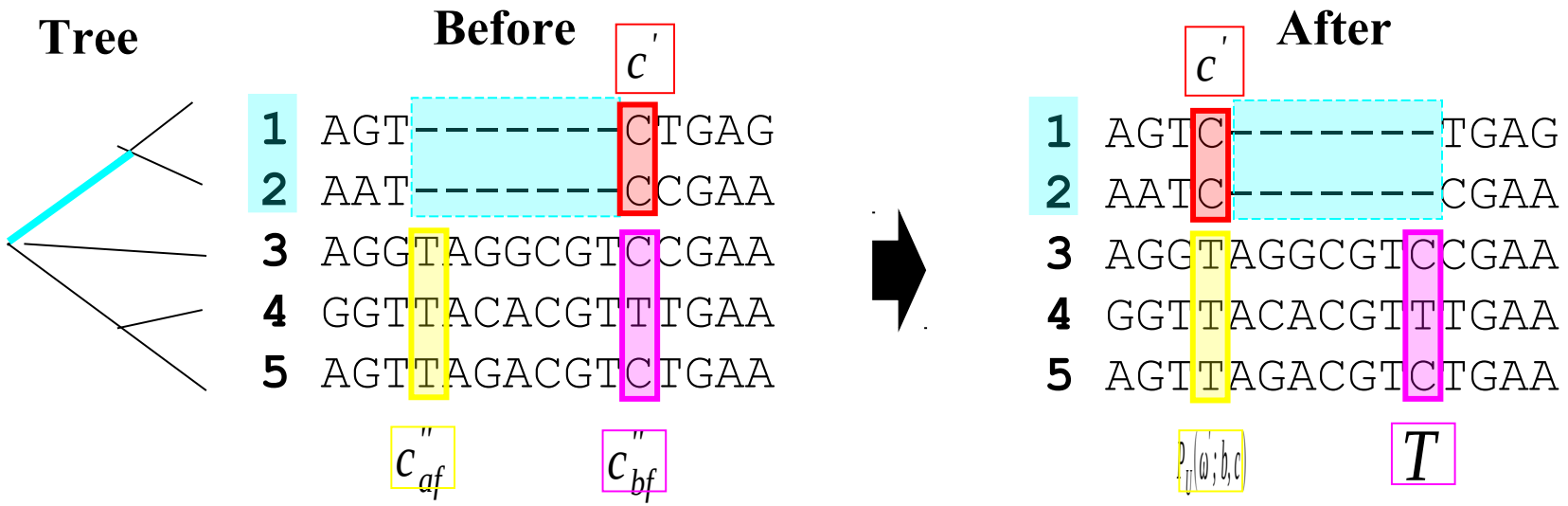


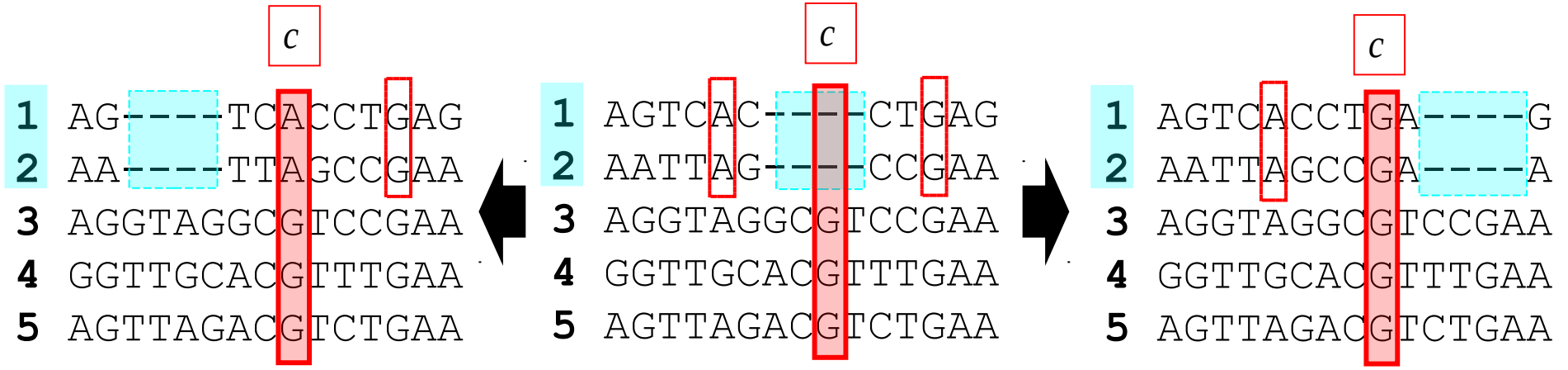
(C) Partitioning network ($\Gamma_{K;0}$) and indel blocks ($\Gamma_{K;1}$ & $\Gamma_{K;2}$)

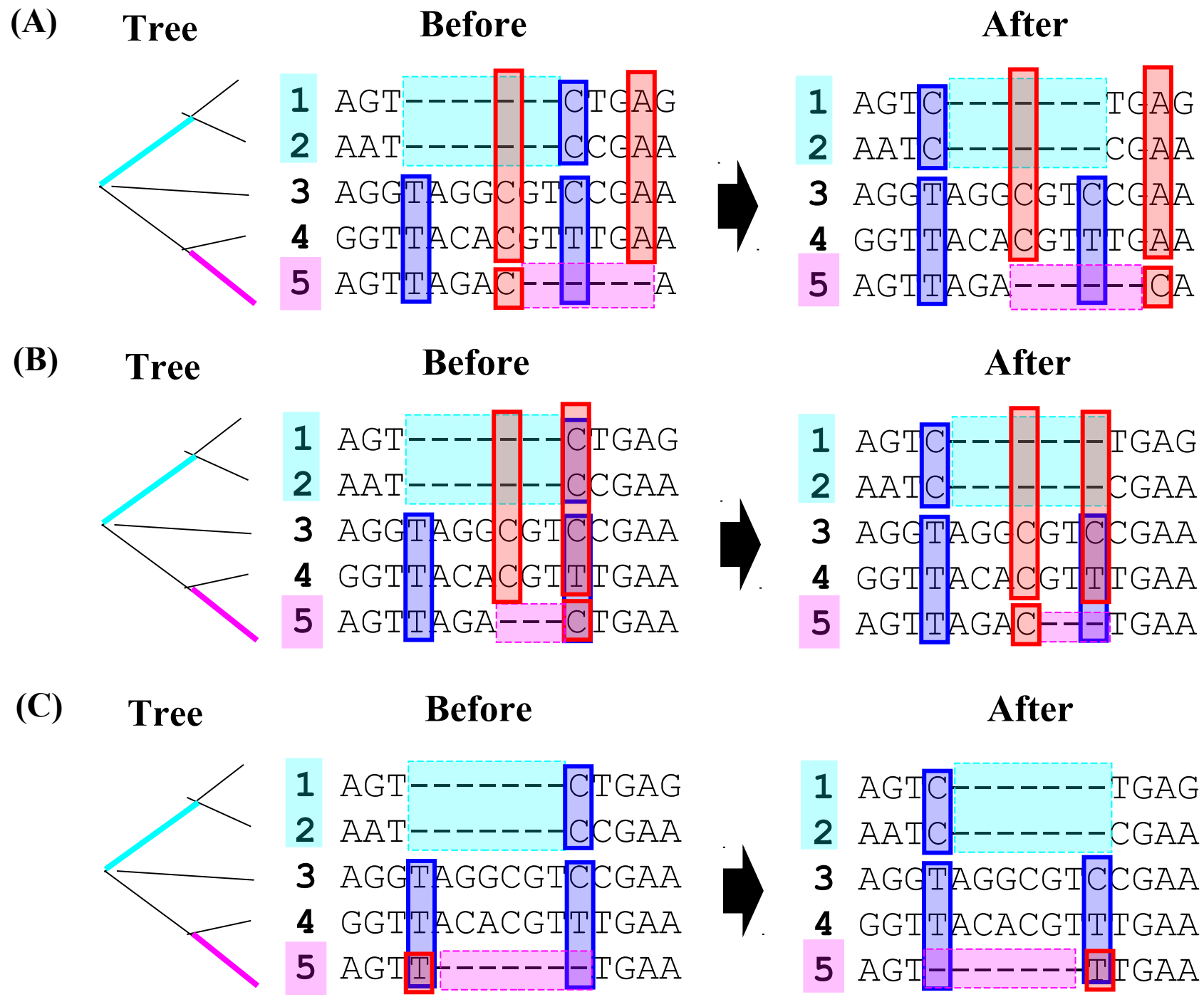


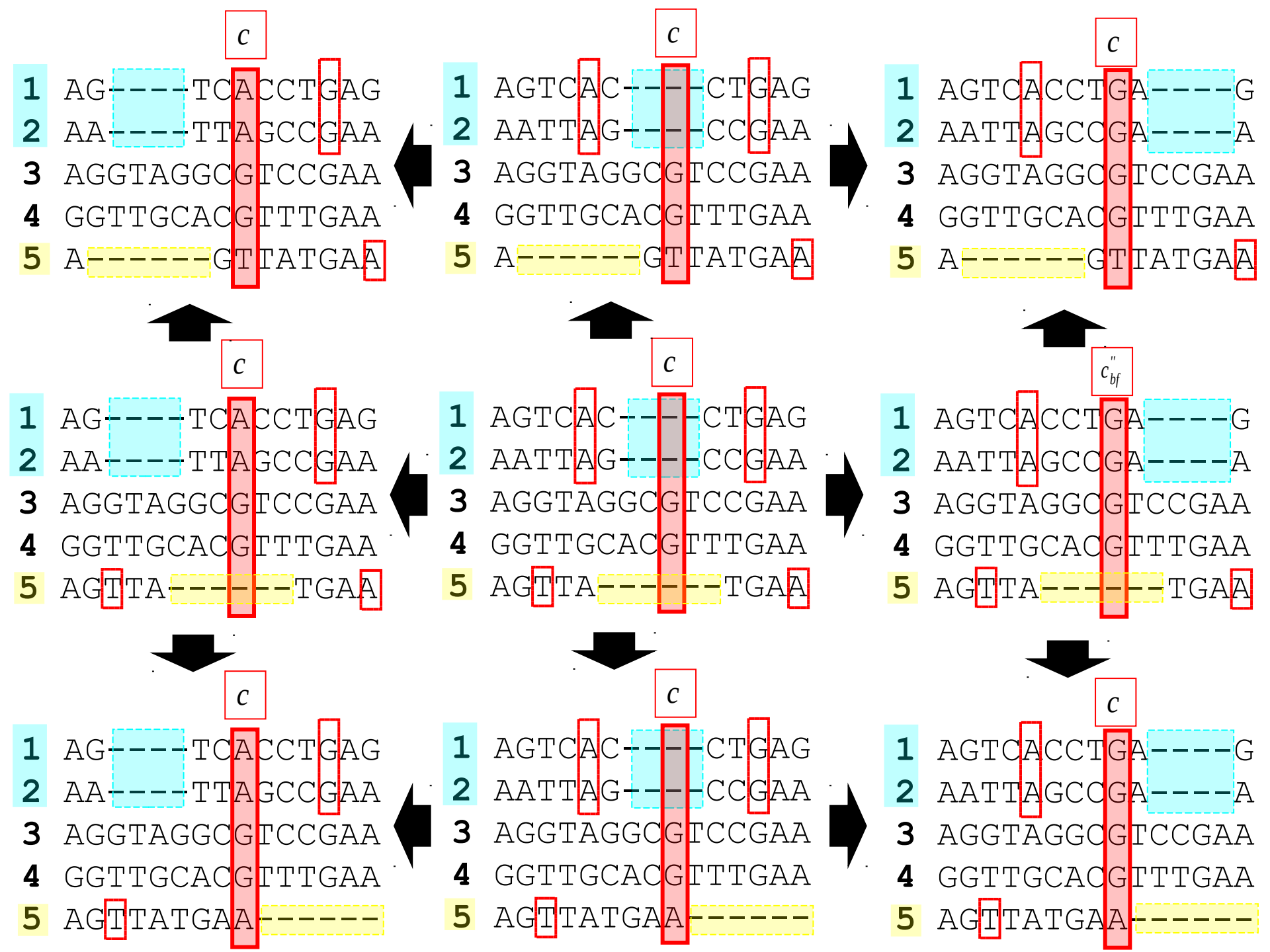

Category A
("Complex" error)
Category B
(Correct)
Category B
("Non-complex"
error)
Category B
(Correct)
1 .NN---NNNNN--NNNNN--NNNNNNNNNNNNNNNNNN.
2 .NNNNNNNNNNN-NNNNNNNNNNNNNNN--NN-NNNNNNN.
3 .NNNNNNNNNNN-NNNNNNNNNNNNNNNNNNNNNNNNN.
4 .NNNN----NNNNNNNNNNNNNNNNNNNNNNNNNNNN-NN.
5 .NN-N----NNNNNNNNNN-NNNNNNNNNNNNNNNNNNN.
A:A_ex
or
A:B_ex
A:A_in
A:B_ex
B:B_ex
B:B_in
B:B_ex
A:B_ex
or
B:B_ex


A. Input
1 .NNNNNNNNN-NNNNNNNNNNNNNNNNNN-NNNN--N--N-N-NNNN.
2 .NNNNNNNNNNN-N-NNNNNNNNNNNNNNN-NNNNNNNNNNN-N-NNNN.
3 .NNNNNNNNNNNNNNNNN-NNNNNNN--NNNNNNNNNNNN--N-N-N--N.
4 .N-NN--NNNNNNNNNNN-NN---NNNNNNNNNNNNNNN--NNN-NNNN.
5 .N-NN--NNNNNNNNNNN-NN---NNNNNNNNNNNNNNN--N-NNNNNN.
B. Method I
1 .NNNNNNNNN-NNNNNNNNNNNNNNNNNN-NNNN--N--N-N-NNNN.
2 .NNNNNNNNNNN-N-NNNNNNNNNNNNNNN-NNNNNNNNNNN-N-NNNN.
3 .NNNNNNNNNNNNNNNNN-NNNNNNN--NNNNNNNNNNNN--N-N-N--N.
4 .N-NN--NNNNNNNNNNN-NN---NNNNNNNNNNNNNNN--NNN-NNNN.
5 .N-NN--NNNNNNNNNNN-NN---NNNNNNNNNNNNNNN--N-NNNNNN.
C. Method III
1 .NNNNNNNNN-NNNNNNNNNNNNNNNNNN-NNNN--N--N-N-NNNN.
2 .NNNNNNNNNNN-N-NNNNNNNNNNNNNNN-NNNNNNNNNNN-N-NNNN.
3 .NNNNNNNNNNNNNNNNN-NNNNNNN--NNNNNNNNNNNN--N-N-N--N.
4 .N-NN--NNNNNNNNNNN-NN---NNNNNNNNNNNNNNN--NNN-NNNN.
5 .N-NN--NNNNNNNNNNN-NN---NNNNNNNNNNNNNNN--N-NNNNNN.
D. Method II
1 .NNNNNNNNN-NNNNNNNNNNNNNNNNNN-NNNN--N--N-N-NNNN.
2 .NNNNNNNNNNN-N-NNNNNNNNNNNNNNN-NNNNNNNNNNN-N-NNNN.
3 .NNNNNNNNNNNNNNNNN-NNNNNNN--NNNNNNNNNNNN--N-N-N--N.
4 .N-NN--NNNNNNNNNNN-NN---NNNNNNNNNNNNNNN--NNN-NNNN.
5 .N-NN--NNNNNNNNNNN-NN---NNNNNNNNNNNNNNN--N-NNNNNN.


Tree
Before
After
c'
1 AGT-------CTGAG
2 AAT-------CCGAA
3 AGGTAGGCGTCCGAA
4 GGTTACACGTTTGAA
5 AGTTAGACGTCTGAA
c''af
c''bf
1 AGTC-------TGAG
2 AATC-------CGAA
3 AGGTAGGCGTCCGAA
4 GGTTACACGTTTGAA
5 AGTTAGACGTCTGAA
T


c
1 AG----TCACCTGAG
2 AA----TTAGCCGAA
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTAGACGTCTGAA
c
1 AGTCAC----CTGAG
2 AATTAG----CCGAA
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTAGACGTCTGAA
c
1 AGTCACCTGA----G
2 AATTAGCCGA----A
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTAGACGTCTGAA


(A)
Tree
Before
After
1 AGT----CTGAG
2 AAT----CCGAA
3 AGGTAGGCGTCCGAA
4 GGTTACACGTTTGAA
5 AGTTAGAC-----A
1 AGTC-----TGAG
2 AATC-----CGAA
3 AGGTAGGCGTCCGAA
4 GGTTACACGTTTGAA
5 AGTTAGA-----CA
(B)
Tree
Before
After
1 AGT----CTGAG
2 AAT----CCGAA
3 AGGTAGGCGTCCGAA
4 GGTTACACGTTTGAA
5 AGTTAGA---CTGAA
1 AGTC-----TGAG
2 AATC-----CGAA
3 AGGTAGGCGTCCGAA
4 GGTTACACGTTTGAA
5 AGTTAGAC---TGAA
(C)
Tree
Before
After
1 AGT------CTGAG
2 AAT------CCGAA
3 AGGTAGGCGTCCGAA
4 GGTTACACGTTTGAA
5 AGTT------TGAA
1 AGTC-------TGAG
2 AATC-------CGAA
3 AGGTAGGCGTCCGAA
4 GGTTACACGTTTGAA
5 AGT-------TTGAA


c
1 AG----TCACCTGAG
2 AA----TTAGCCGAA
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 A------GTTATGAA
c
1 AGTCAC----CTGAG
2 AATTAG----CCGAA
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 A------GTTATGAA
c
1 AGTCACCTGA----G
2 AATTAGCCGA----A
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 A------GTTATGAA
c
1 AG----TCACCTGAG
2 AA----TTAGCCGAA
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTA------TGAA
c
1 AGTCAC----CTGAG
2 AATTAG----CCGAA
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTA------TGAA
c''_{bf}
1 AGTCACCTGA----G
2 AATTAGCCGA----A
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTA------TGAA
c
1 AG----TCACCTGAG
2 AA----TTAGCCGAA
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTATGAA------
c
1 AGTCAC----CTGAG
2 AATTAG----CCGAA
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTATGAA------
c
1 AGTCACCTGA----G
2 AATTAGCCGA----A
3 AGGTAGGCGTCCGAA
4 GGTTGCACGTTTGAA
5 AGTTATGAA------