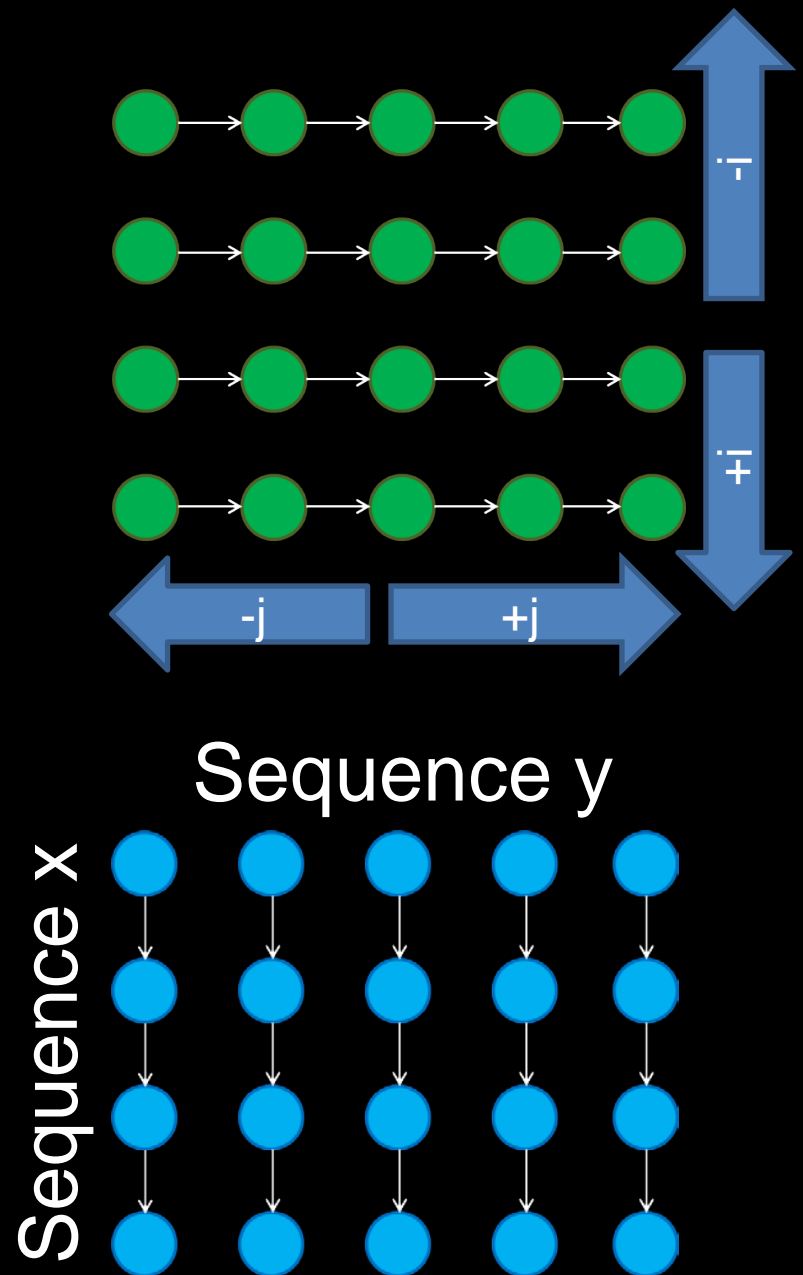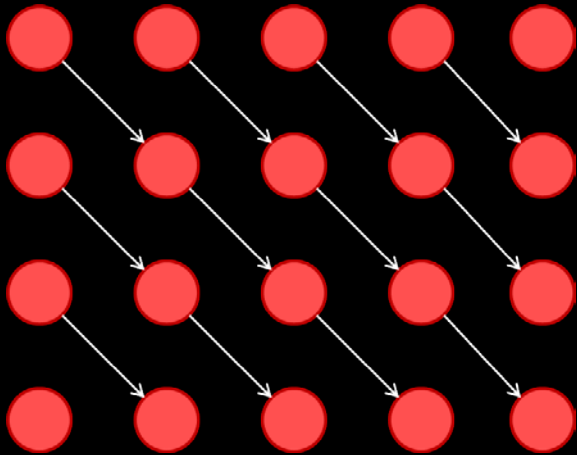# Backtracking on Affine Gap Tables

Brian Y. Chen

Affine Gap Tables

- Affine gap tables are subgraphs of the linear gap tables, connected to each other

Sequence y

Sequence x

-i

+i

-j

+j

Begin with table values

- Once the table values are determined, we backtrack through the grid to find the optimal alignment

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-14.5	-15	-15.5	-16
A	-∞	-14.5	2	-8.5	-14	-9.5
T	-∞	-15	-13.5	-2	-7.5	-13

Matches, Mismatches only

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Horizontal gaps only

|	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-15	-20.5	-19.5	-20
T	-11.5	-10	-8.5	-19	-19.5	-20

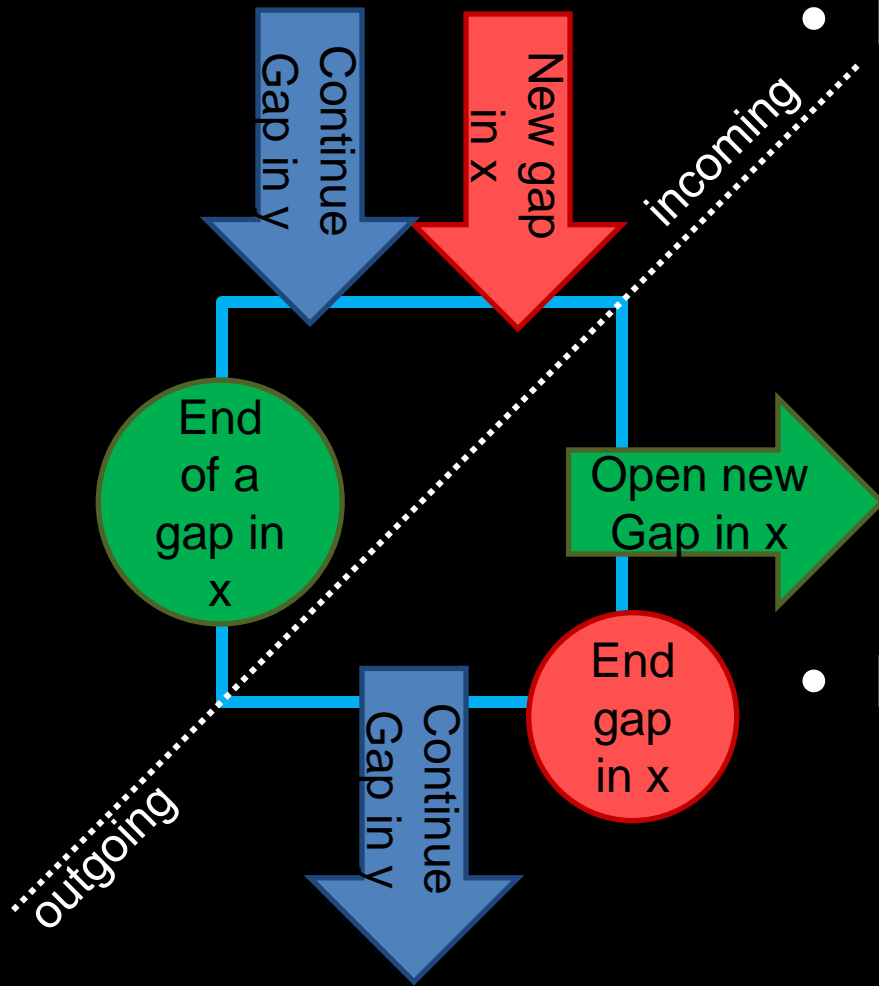Vertical gaps only

A node (i, j) in the M graph
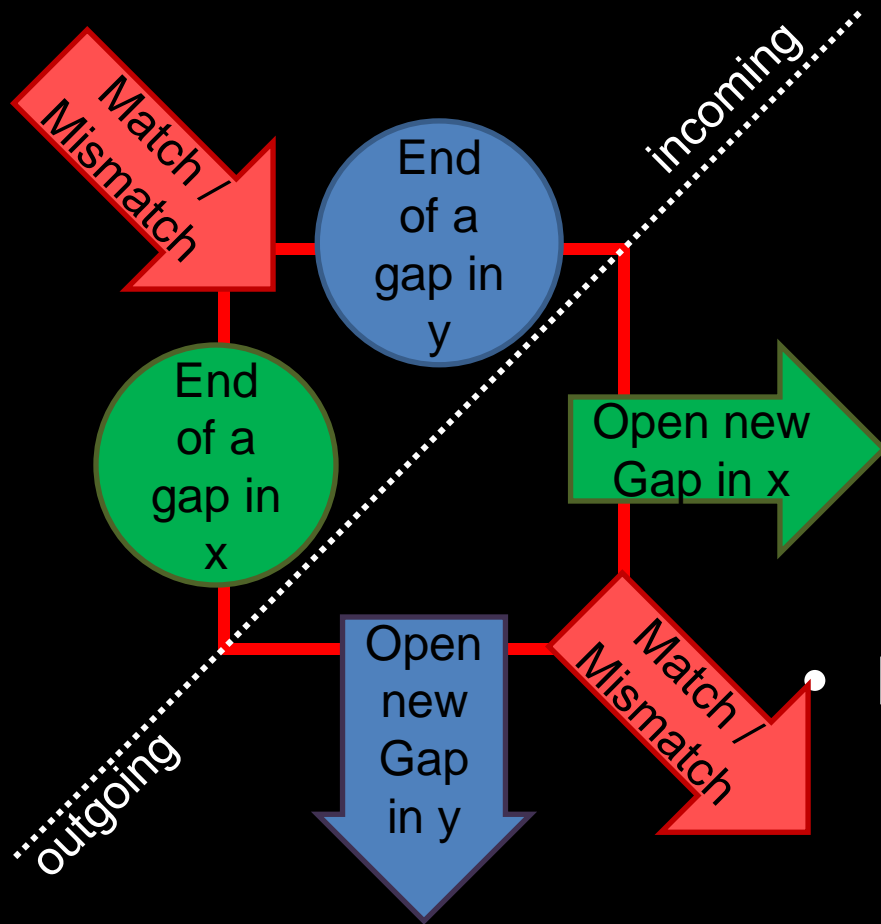
- Has edges coming from:
  - M[i-1, j-1] (cost +1 or -4)
    - This is either a match (+1) or a mismatch (-4)
  - X[i, j] (cost 0)
    - This corresponds to the end of a gap in sequence x (the sequence on the left)
  - Y[i, j] (cost 0)
    - This corresponds to the end of a gap in sequence y (the sequence on top)
- Has edges going to:
  - M[i+1, j+1] (cost +1 or -4)
    - Also a match or a mismatch
  - X[i, j+1] (cost S (-10) + E (-.5))
    - Open a new gap in sequence x
  - Y[i+1, j] (cost S (-10) + E (-.5))
    - Open a new gap in sequence y

A node (i, j) in the X graph
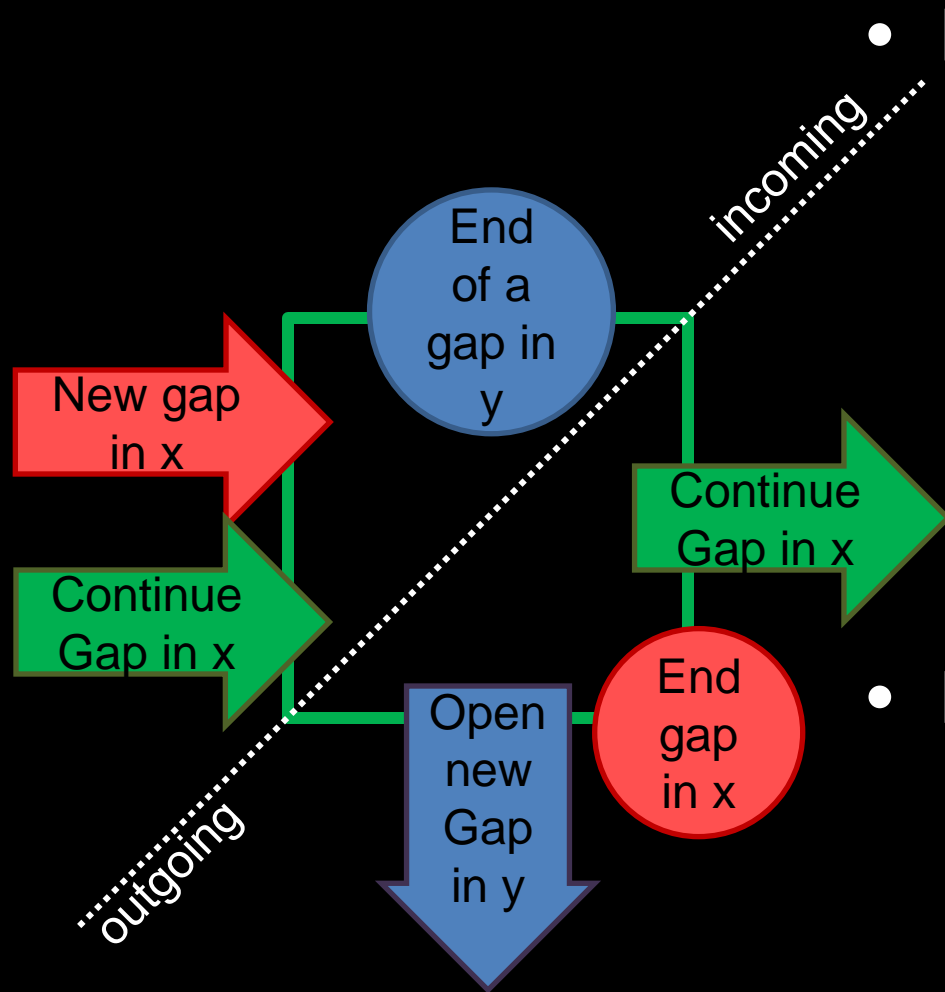
- Has edges coming from:
  - M[i, j-1] (cost S + E)
    - This is for a new gap coming from a match or mismatch
  - X[i, j-1] (cost E (-.5))
    - This extends a gap in sequence x (the sequence on the left)
  - Y[i, j-1] (cost S (-10) + E (-.5))
    - This is for a new gap coming from a gap in sequence y (top)
- Has edges going to:
  - M[i, j] (cost 0)
    - This is for ending the gap
  - X[i, j+1] (cost E (-.5))
    - Extend the gap in sequence x
  - Y[i, j+1] (cost S + E)
    - Open a new gap in sequence y

A node (i, j) in the Y graph

Continue Gap in y

New gap in x

incoming

End of a gap in x

Open new Gap in x

outgoing

End gap in x

Continue Gap in y

- Has edges coming from:
  - M[i-1, j] (cost S + E)
    - This is for a new gap coming from a match or mismatch
  - X[i-1, j] (cost S + E)
    - This is for a new gap coming from a gap in sequence x (left)
  - Y[i-1, j] (cost S)
    - This extends a gap in sequence y (the sequence on the top)
- Has edges going to:
  - M[i, j] (cost 0)
    - This is for ending the gap
  - X[i+1, j] (cost S + E)
    - Open a new gap in sequence x
  - Y[i+1, j] (cost E)
    - Extend the gap in sequence y

Start Here, M[3,5]

- Note that this is the same as the lower right corner in M, X, or Y, because of edge to get here are zero.

Matches, Mismatches only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	**-17**

Horizontal gaps only

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Vertical gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

End of a gap in y — Match / Mismatch — End of a gap in x — Open new Gap in x — Open new Gap in y — Match / Mismatch — incoming — outgoing

- The best transition: X[3,5]
  - Note that this is the same as the lower right corner in M, X, or Y, because of edge to get here are zero.

Matches, Mismatches only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C -∞	1	-13.5	-14	-14.5	-15
A -∞	-13.5	2	-12.5	-13	-8.5
T -∞	-14	-12.5	-2	-11.5	-17

Horizontal gaps only

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C -∞	-21	-9.5	-10	-10.5	-11
A -∞	-21.5	-20	-8.5	-9	-9.5
T -∞	-22	-20.5	-19	-12.5	-13

Vertical gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C -10.5	-21	-21.5	-22	-22.5	-23
A -11	-9.5	-20	-20.5	-21	-21.5
T -11.5	-10	-8.5	-19	-19.5	-19

Possible backtracks:

M[2,4]:   -13 - 4

X[3,5]:   -13 + 0

Y[3,5]:   -19 + 0

Match / Mismatch

End of a gap in y

incoming

End of a gap in x

Open new Gap in x

Open new Gap in y

Match / Mismatch

outgoing

The best transition: X[3,5]

- Note that this is the same as the lower right corner in M, X, or Y, because of edge to get here are zero.

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C -∞	1	-13.5	-14	-14.5	-15
A -∞	-13.5	2	-12.5	-13	-8.5
T -∞	-14	-12.5	-2	-11.5	-17

Matches, Mismatches only

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C -∞	-21	-9.5	-10	-10.5	-11
A -∞	-21.5	-20	-8.5	-9	-9.5
T -∞	-22	-20.5	-19	-9.5	-13

Horizontal gaps only

First transition: M[3,5] to X[3,5]

Possible backtracks:

M[2,4]:   -13 - 4
X[3,5]:   -13 + 0
Y[3,5]:   -19 + 0

A
–

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C -10.5	-21	-21.5	-22	-22.5	-23
A -11	-9.5	-20	-20.5	-21	-21.5
T -11.5	-10	-8.5	-19	-19.5	-19

Vertical gaps only

End of a gap in y

incoming

Match / Mismatch

End of a gap in x

Open new Gap in x

Open new Gap in y

Match / Mismatch

outgoing

Now we are here: X[3,5]

- Based on the X node transition diagram, we have 3 backtrack possibilities

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	-17

Matches, Mismatches only

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Horizontal gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

Vertical gaps only

A
–

- # Now we are here: X[3,5]
  - – Based on the X node transition diagram, we have 3 backtrack possibilities

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	-17

Matches, Mismatches only

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Horizontal gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

Vertical gaps only

Possible backtracks:

M[3,4]: -11.5 - S - E

X[3,4]: -12.5 - E

Y[3,4]: -19.5 - S - E

A
–

End of a gap in y
New gap in x
Continue Gap in x
Continue Gap in x
Open new Gap in y
End gap in x
incoming
outgoing

Best Transition: X[3,4]

- Best Transition: X[3,4]
  - Based on the X node transition diagram, we have 3 backtrack possibilities

Matches, Mismatches only

	C	A	G	T	A
	0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	-17
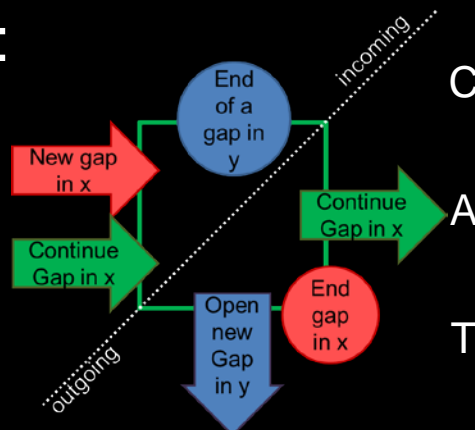
Horizontal gaps only

	C	A	G	T	A
	0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Second Transition: X[3,5] to X[3,4]

Vertical gaps only

	C	A	G	T	A
	0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

Possible backtracks:

M[3,4]: -11.5 - S - E

X[3,4]: -12.5 - E

Y[3,4]: -19.5 - S - E

TA
- -

New gap in x

Continue Gap in x

End of a gap in y

Continue Gap in x

Continue Gap in x

Continue Gap in x

Open new Gap in y

End gap in x

incoming

outgoing

- Now we are here: X[3,4]

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Horizontal gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	-17

Matches, Mismatches only

Possible backtracks:

M[3,3]: -2 - S - E

X[3,3]: -19 - E

Y[3,3]: -19 - S - E

TA
--

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

Vertical gaps only

End of a gap in y

New gap in x

incoming

Continue Gap in y

Continue Gap in x

Open new Gap in y

End gap in x

outgoing

- The Best Transition: X[3,3]
  - M[3,3] scores a lot higher than the other two options
  - Mismatch was chosen over the dual gap, because of gap opening costs

Matches, Mismatches only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.	-17
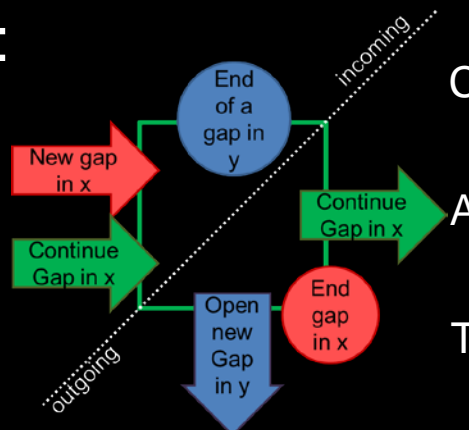
	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20	-12.5	-13

Next Transition: X[3,4] to M[3,3]

Possible backtracks:

M[3,3]: -2 - S - E

X[3,3]: -19 - E

Y[3,3]: -19 - S - E

```
GTA
T--
```

New gap in x

End of a gap in y

incoming

Continue Gap in x

Continue Gap in x

Open new Gap in y

End gap in x

outgoing

Vertical gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

- Now we are here: X[3,3]

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C -∞	-21	-9.5	-10	-10.5	-11
A -∞	-21.5	-20	-8.5	-9	-9.5
T -∞	-22	-20.5	(-19)	(-12.5)	(-13)

Horizontal gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C -∞	1	-13.5	-14	-14.5	-15
A -∞	-13.5	(2)	-12.5	-13	-8.5
T -∞	-14	-12.5	(-2)	-11.5	(-17)

Matches, Mismatches only

Possible backtracks:

M[2,2]: 2 + -4

X[3,3]: -19 - 0

Y[3,3]: -19 - 0

```
GTA
T--
```


	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C -10.5	-21	-21.5	-22	-22.5	-23
A -11	-9.5	-20	-20.5	-21	-21.5
T -11.5	-10	-8.5	(-19)	-19.5	-19

Vertical gaps only

The best Transition: X[2,2]

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Horizontal gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	-17

Matches, Mismatches only

Possible backtracks:

M[2,2]: 2 + -4

X[3,3]: -19 - 0

Y[3,3]: -19 - 0

AGTA
AT--

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

Vertical gaps only

End of a gap in y

incoming

Match / Mismatch

End of a gap in x

Open new Gap in x

Open new Gap in y

Match / Mismatch

outgoing

- Now we are here: X[2,2]

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C -∞	-21	-9.5	-10	-10.5	-11
A -∞	-21.5	-20	-8.5	-9	-9.5
T -∞	-22	-20.5	-19	-12.5	-13

Horizontal gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C -∞	1	-13.5	-14	-14.5	-15
A -∞	-13.5	2	-12.5	-13	-8.5
T -∞	-14	-12.5	-2	-11.5	-17

Matches, Mismatches only

Possible backtracks:

M[1,1]: 1 + 1

X[2,2]: -20 - 0

Y[2,2]: -20 - 0

AGTA
AT--
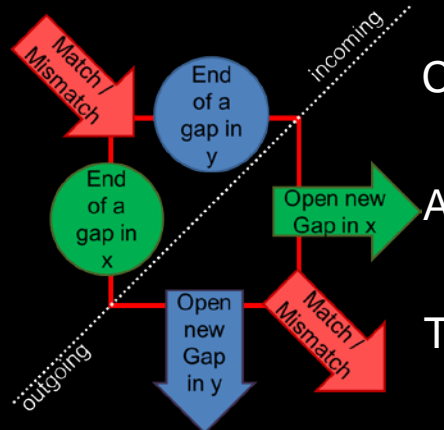
	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C -10.5	-21	-21.5	-22	-22.5	-23
A -11	-9.5	-20	-20.5	-21	-21.5
T -11.5	-10	-8.5	-19	-19.5	-19

Vertical gaps only

The best transition: X[1,1]

	C	A	G	T	A
0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Horizontal gaps only

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	-17

Matches, Mismatches only

Possible backtracks:

M[1,1]: 1 + 1

X[2,2]: -20 - 0

Y[2,2]: -20 - 0

CAGTA
CAT--

	C	A	G	T	A
0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

Vertical gaps only

End of a gap in y

incoming

Match / Mismatch

End of a gap in x

Open new Gap in x

Open new Gap in y

Match / Mismatch

outgoing

- # Now we are here: M[1,1]
  - This is the final step, and it is kind of trivial, but important to be aware of:

Matches, Mismatches only (red grid)

	C	A	G	T	A
	0	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	-17

Matches, Mismatches only

Horizontal gaps only (green grid)

	C	A	G	T	A
	0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Horizontal gaps only

Vertical gaps only (blue grid)

	C	A	G	T	A
	0	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

Vertical gaps only

Possible backtracks:

M[0,0]: 0 + 1

X[1,1]: -21 - 0

Y[1,1]: -21 - 0

CAGTA
CAT--

(diagram) Match / Mismatch — End of a gap in y — incoming — Open new Gap in x — End of a gap in x — Open new Gap in y — Match / Mismatch — outgoing

- Final Transition: M[0,0]

Matches, Mismatches only

	C	A	G	T	A
	0	-∞	-∞	-∞	-∞
C	-∞	1	-13.5	-14	-14.5	-15
A	-∞	-13.5	2	-12.5	-13	-8.5
T	-∞	-14	-12.5	-2	-11.5	-17

Horizontal gaps only

	C	A	G	T	A
	0	-10.5	-11	-11.5	-12	-12.5
C	-∞	-21	-9.5	-10	-10.5	-11
A	-∞	-21.5	-20	-8.5	-9	-9.5
T	-∞	-22	-20.5	-19	-12.5	-13

Vertical gaps only

	C	A	G	T	A
	0	-∞	-∞	-∞	-∞	-∞
C	-10.5	-21	-21.5	-22	-22.5	-23
A	-11	-9.5	-20	-20.5	-21	-21.5
T	-11.5	-10	-8.5	-19	-19.5	-19

Possible backtracks:

M[0,0]: 0 + 1

X[1,1]: -21 - 0

Y[1,1]: -21 - 0

CAGTA
CAT--

End of a gap in y

incoming

End of a gap in x

Open new Gap in x

Match / Mismatch

Match / Mismatch

Open new Gap in y

outgoing

Some hints on the applications project

- Gene predictions from Glimmer and Genemark will be frequently incorrect and contradictory
  - Rather than assuming they are always correct, assume they are <u>wrong</u>.

Wrong boundries   Missing genes
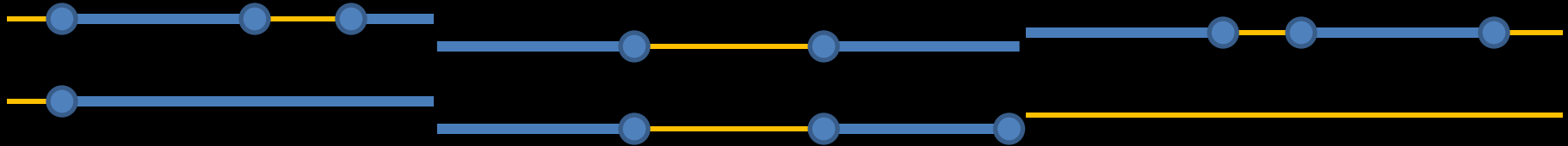
- What you need to do is piece together information that indicates the real position of a gene
  - Blast the gene positions that are predicted
  - Use predictions where Glimmer and Genemark agree
  - Put contigs together and predict on the assembly

Blasting predicted Genes

- Suppose Glimmer and Genemark predict the following gene, A:

A

- Go to the contig that is predicted, and cut and and paste out the part of the contig that is predicted

A → BLAST

- When you search for this in megaBLAST, you will find a number of things:
  - Genomes
    - This indicate what virus you have, but don't tell you what gene you have
  - Genes (suppose you find a gene B)
    - These tell you what gene you might have
    - Use ClustalW to align these genes to your gene, A:

The fact that B is longer, but still aligns, indicates that Genemark was off by some nucleotides

A

B
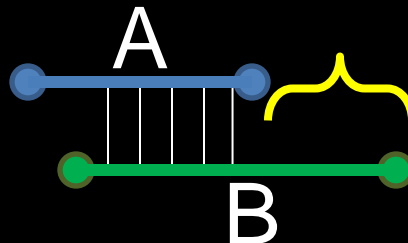
Now check to rest of the contig to see if it is the same as the rest of B

How to really use BLAST

What you give to BLAST
What you get from BLAST

A

B

?

- By aligning BLAST genes to your input, you can see where the ends line up.

  - If B runs off the end of your contig, perhaps <u>another</u> contig has the same nucleotides as the one from blast – so check it out in the fasta file.

- The BLAST result is an actual gene

  - So if it aligns partially, use the blast result to define the actual end points of your gene, for correct results.

Using Genemark and Glimmer together

Genemark Prediction ———————————

Glimmer Prediction ———————————

- Genemark and Glimmer do not always agree, but where they do, their predictions will overlap.

- If you want to be more certain of the predictions, try blasting for regions where predictions overlap.

Predicting on the assembly

Genemark Prediction 1 ●━━━━━●

Genemark Prediction 2 ●━━━━━━━━━━●

- Genemark and Glimmer change their predictions based on what you give them.

  – If you give them a longer contig, their prediction about the contig may be different, depending on what they see.

- Assembling contigs together can lead to more accurate predictions because of shortening

  – Especially when you know there is a longer gene that extends onto the other contig, you can try assembling them

  – Look at the sequence and see if the letters match up against BLAST results