Approximate matching

Read

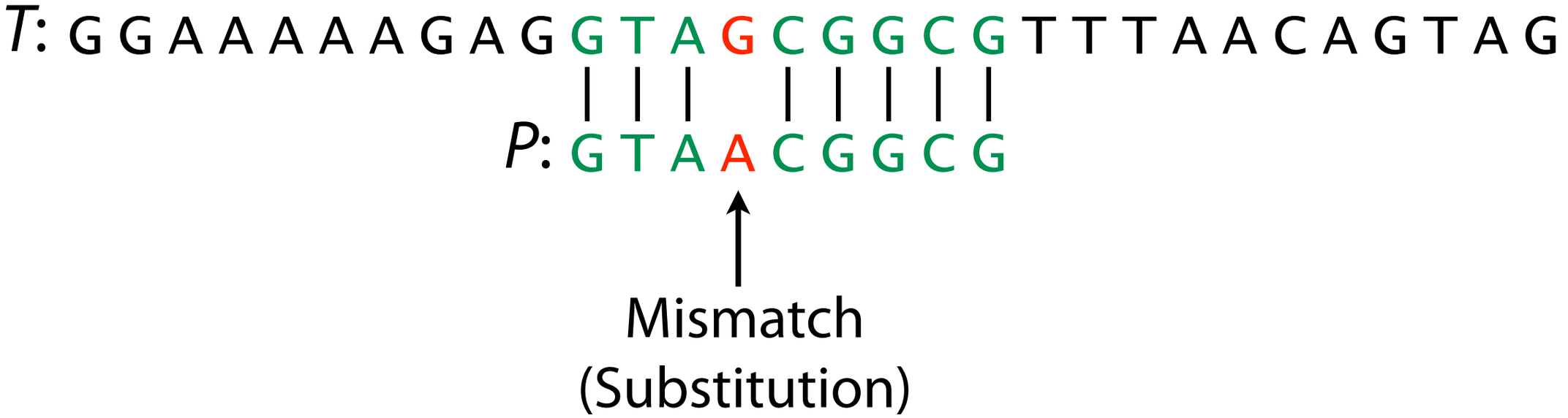
CTCAAACTCCTGACCTTTGGTGATCCACCCGCCTAGGCCTTC

Reference

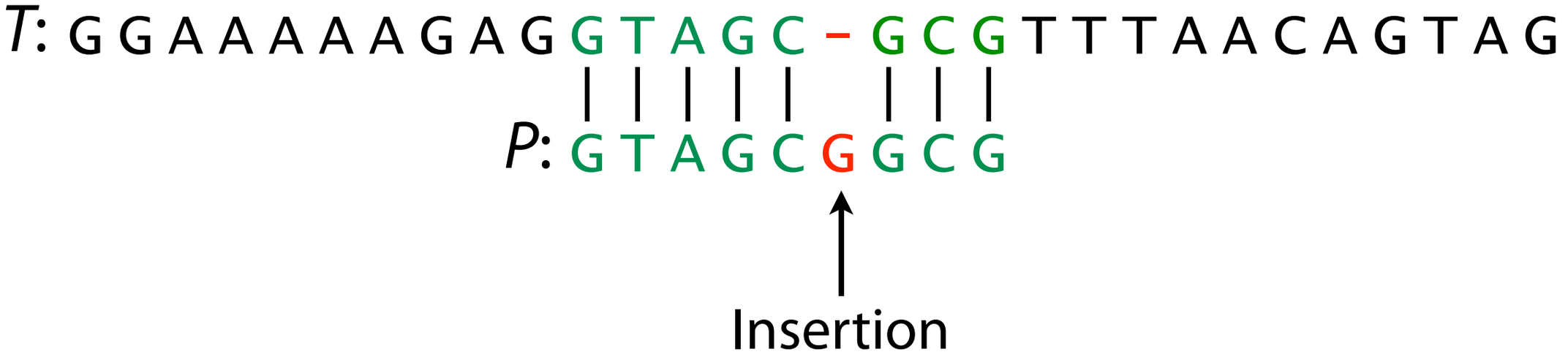
Differences between read and reference occur because of...

1. Sequencing error
2. Natural variation

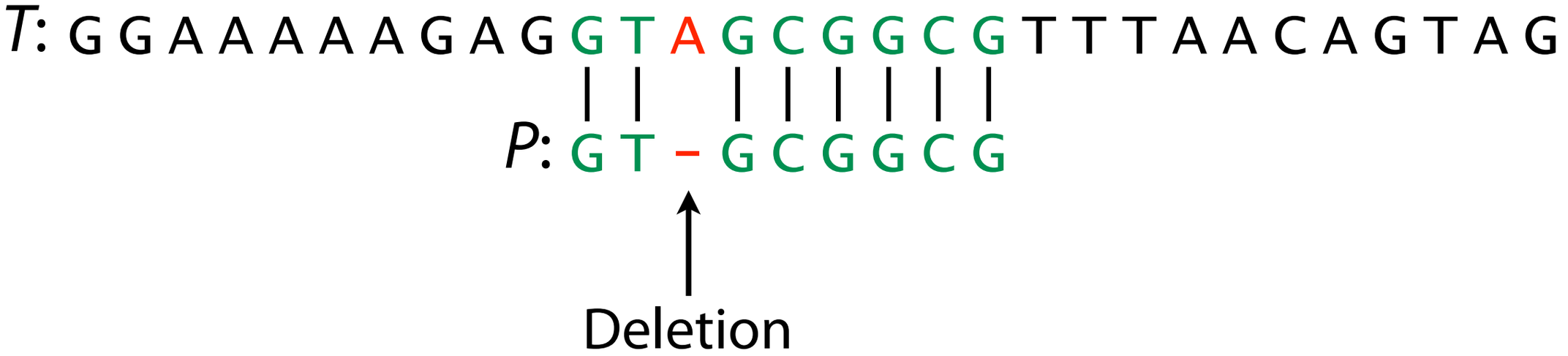
Approximate matching



Approximate matching



Approximate matching



Hamming distance

For X & Y where $|X| = |Y|$, *hamming distance* =
minimum # substitutions needed to turn one into the other

Hamming distance = 3

Edit distance

(AKA Levenshtein distance)

For *X* & *Y*, *edit distance* = minimum # edits (substitutions, insertions, deletions) needed to turn one into the other

```
X: T G G C C G C G C A A A A A C A G C
   | |   | | | | | | | | |   | | | |
Y: T G A C C G C G C A A A A - C A G C
```

Edit distance = 2

```
X: G C G T A T G C G G C T A - A C G C
   | |   | | | | | | | | |   | | |
Y: G C - T A T G C G G C T A T A C G C
```

Edit distance = 2

Edit distance

For X, Y where $|X| = |Y|$, *hamming distance* = minimum # substitutions needed to turn one into the other

For X, Y, *edit distance* = minimum # edits (substitutions, insertions, deletions) needed to turn one into the other

Edit distance

If $|X| = |Y|$ what can we say about the relationship between **`editDistance`**(X, Y) and **`hammingDistance`**(X, Y)?

$$\textbf{editDistance}(X, Y) \leq \textbf{hammingDistance}(X, Y)$$

```
X: G C G T A T G C G G C T A - A C G C
   | |   | | | | | | | | | |   | | |
Y: G C - T A T G C G G C T A T A C G C
```

Edit distance

If *x* and *y* are different lengths, what can we say about **`editDistance`**(*X*, *Y*)?

`editDistance`(*X*, *Y*) ≥ | | *X* | - | *Y* | |

X: ? ?

Y: ? ? ? ?

X

G G C C G C G C A A A A A C A G C

Y

A T G C C G C G A A A A A C A T A

editDistance(X[:-1], Y[:-1]) = 147


G G C C G C G C A A A A A C A G C
α


A T G C C G C G A A A A A C A T A
β

α C

β A

$$\texttt{edist}(\alpha\mathsf{C}, \beta\mathsf{A}) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + 1 \\ \texttt{edist}(\alpha\mathsf{C}, \beta) + 1 \\ \texttt{edist}(\alpha, \beta\mathsf{A}) + 1 \end{cases}$$

αC

βA

$$\texttt{edist}(\alpha\mathsf{C}, \beta\mathsf{A}) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + 1 \\ \texttt{edist}(\alpha\mathsf{C}, \beta) + 1 \\ \texttt{edist}(\alpha, \beta\mathsf{A}) + 1 \end{cases}$$

α C

β A

$$\texttt{edist}(\alpha\mathsf{C}, \beta\mathsf{A}) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + 1 \\ \texttt{edist}(\alpha\mathsf{C}, \beta) + 1 \\ \texttt{edist}(\alpha, \beta\mathsf{A}) + 1 \end{cases}$$

αC

βA

$$\texttt{edist}(\alpha\mathsf{C}, \beta\mathsf{A}) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + 1 \\ \texttt{edist}(\alpha\mathsf{C}, \beta) + 1 \\ \texttt{edist}(\alpha, \beta\mathsf{A}) + 1 \end{cases}$$

α x

β y

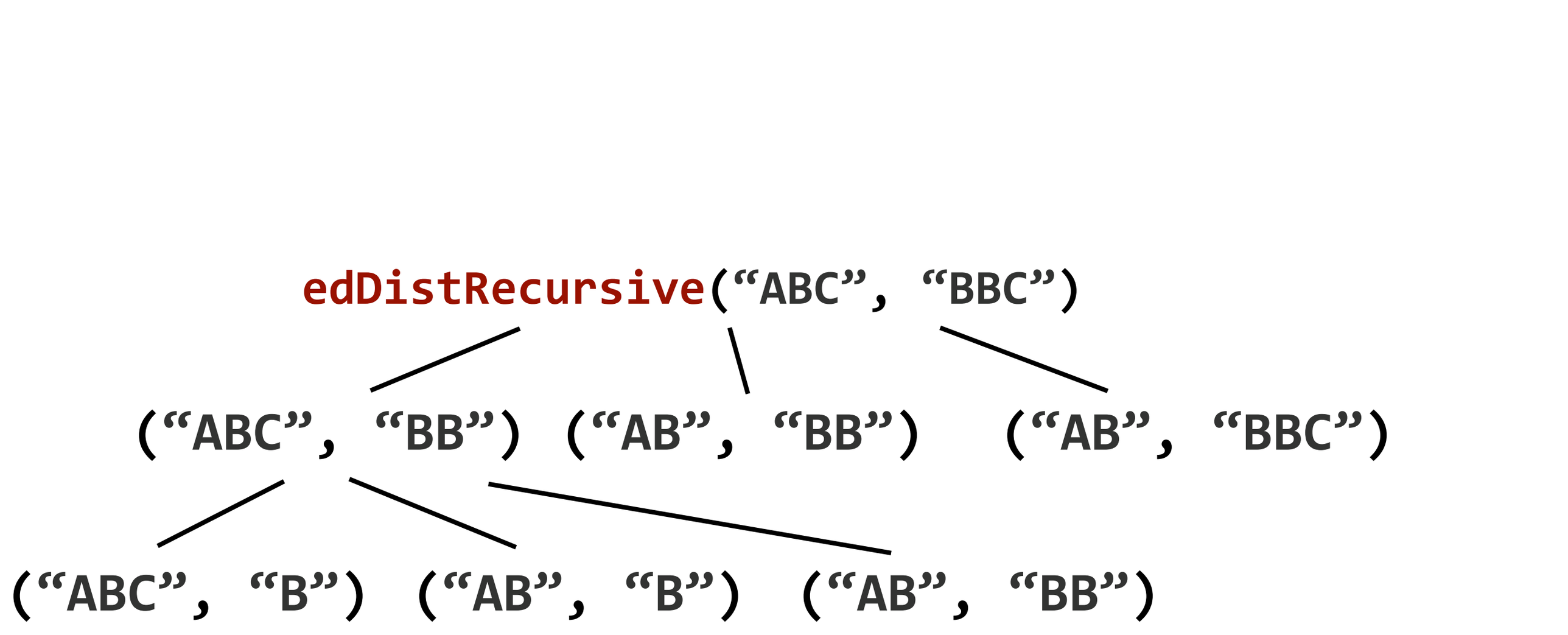
$$\texttt{edist}(\alpha x, \beta y) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + \delta(x, y) \\ \texttt{edist}(\alpha x, \beta) + 1 \\ \texttt{edist}(\alpha, \beta y) + 1 \end{cases}$$

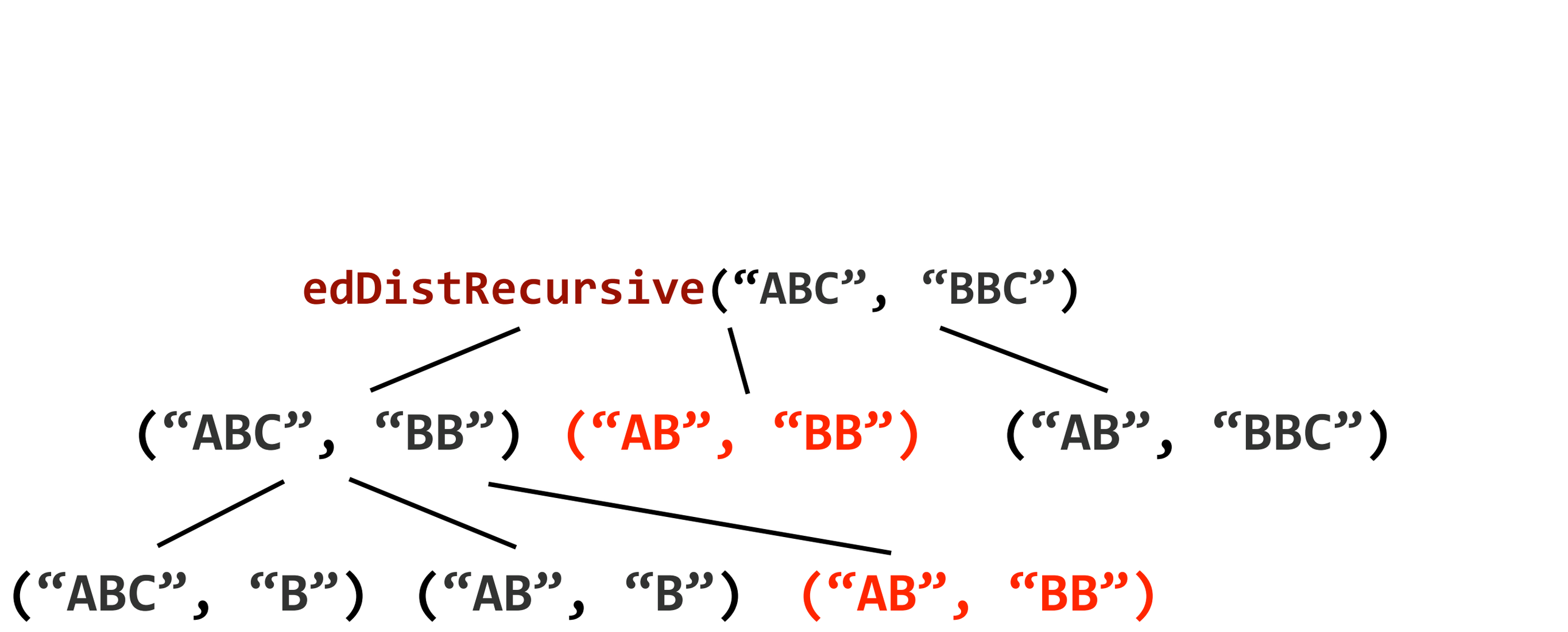
$\delta(x, y) = 0$ if $x = y$, or 1 otherwise

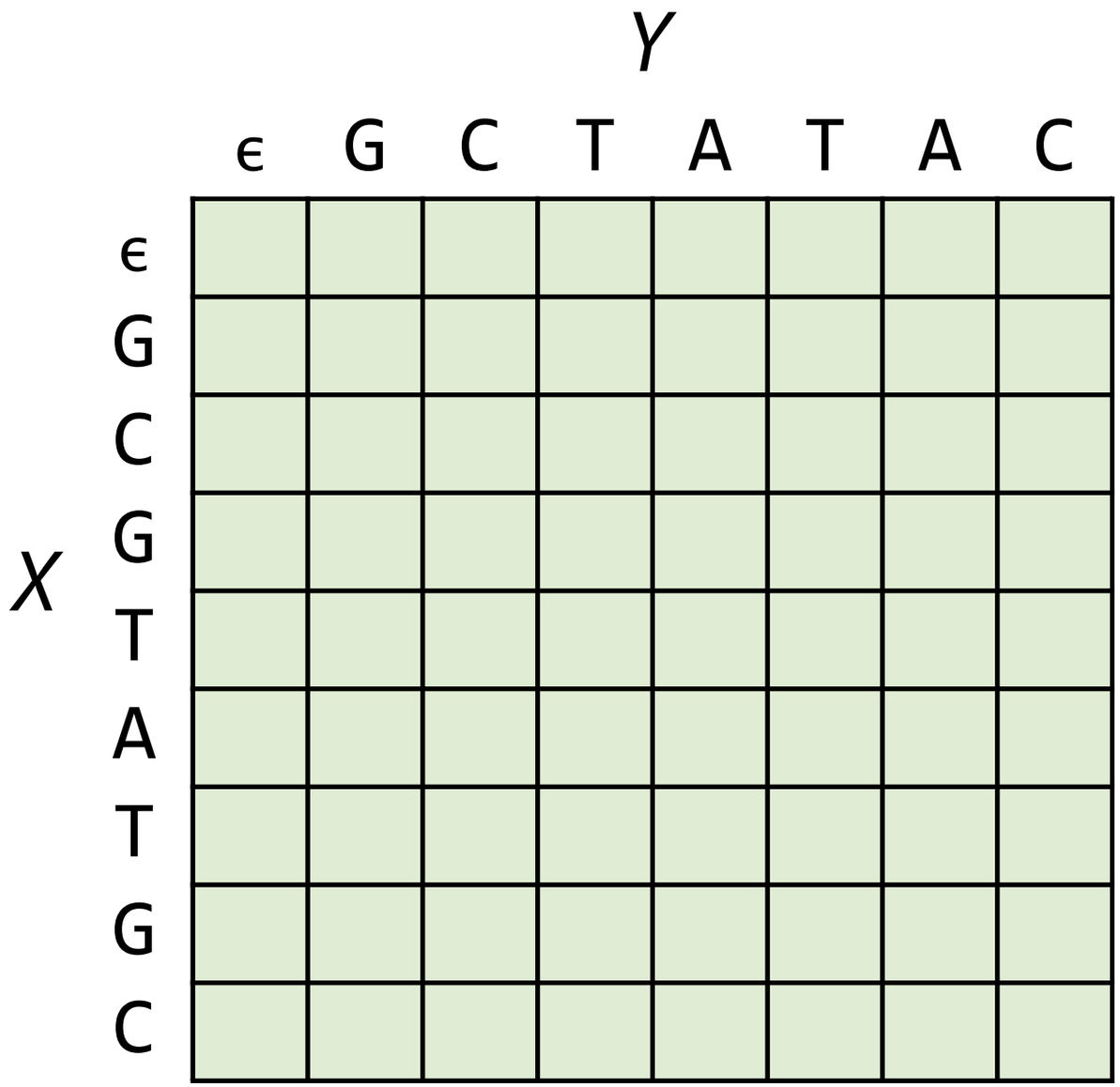
```
delt = 1 if a[-1] != b[-1] else 0
return min(edDistRecursive(a[:-1], b[:-1]) + delt,
           edDistRecursive(a, b[:-1]) + 1,
           edDistRecursive(a[:-1], b) + 1)
```

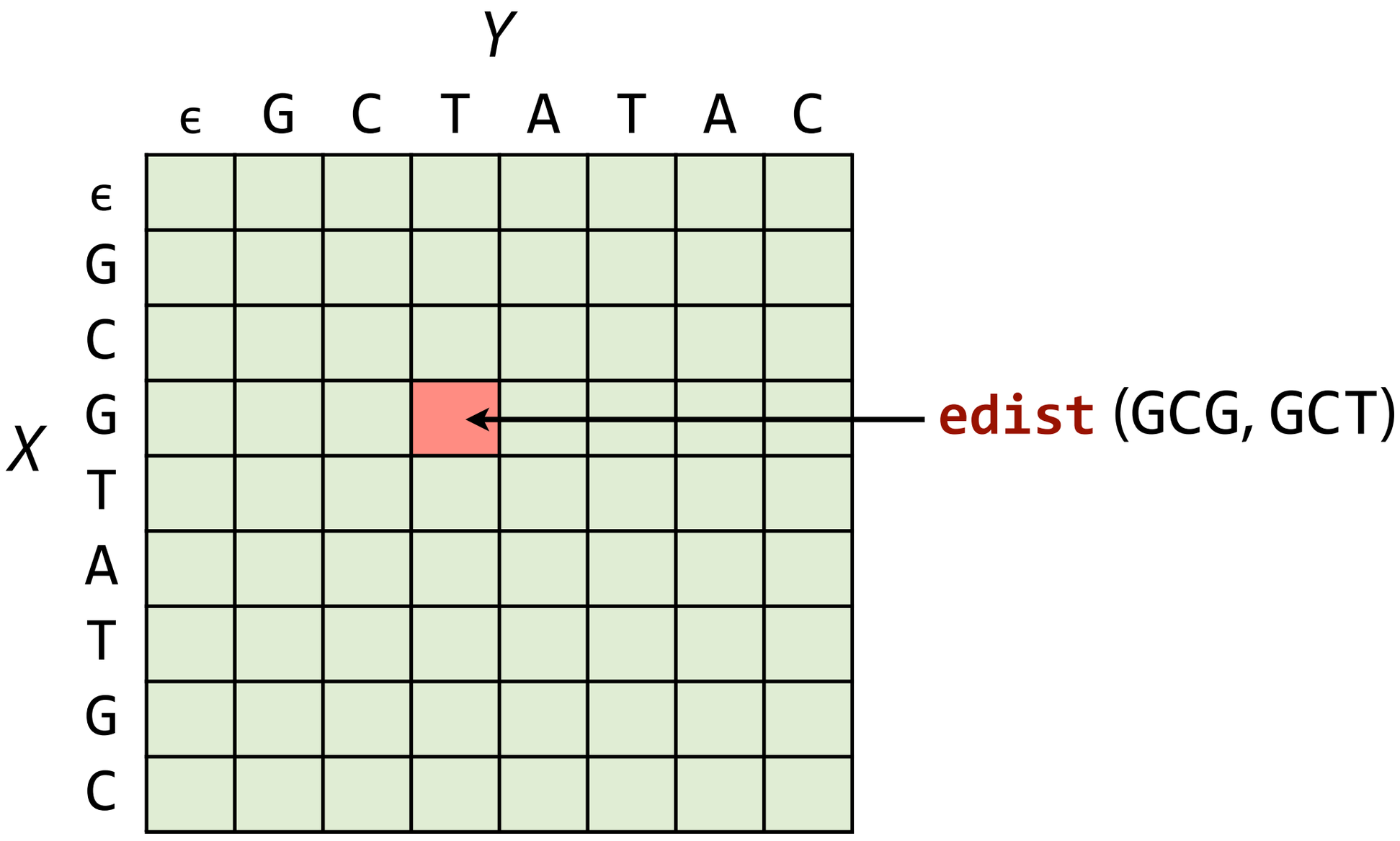
$$\texttt{edist}(\alpha x, \beta y) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + \delta(x, y) \\ \texttt{edist}(\alpha x, \beta) + 1 \\ \texttt{edist}(\alpha, \beta y) + 1 \end{cases}$$

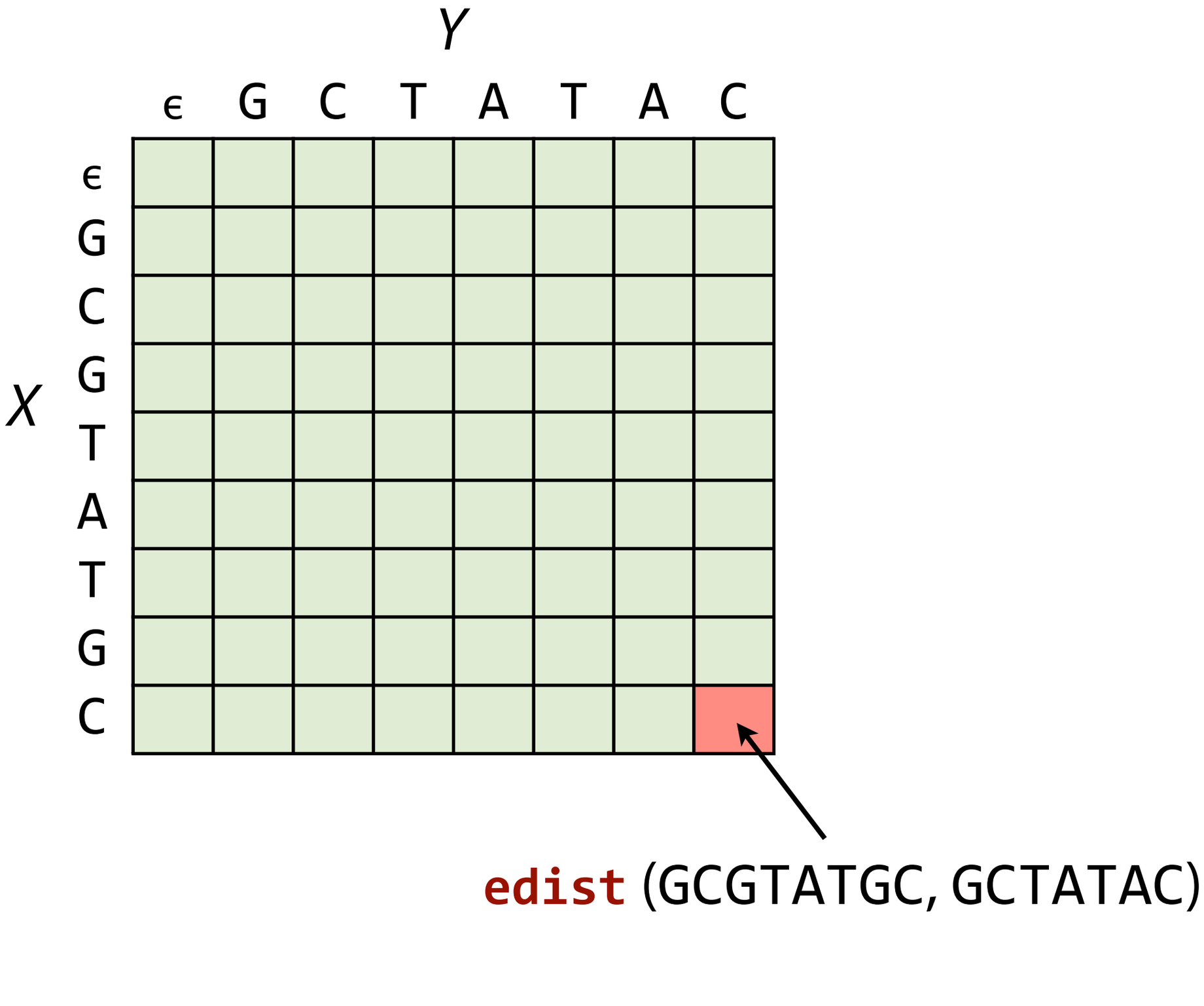
$$\delta(x, y) = 0 \text{ if } x = y, \text{ or } 1 \text{ otherwise}$$

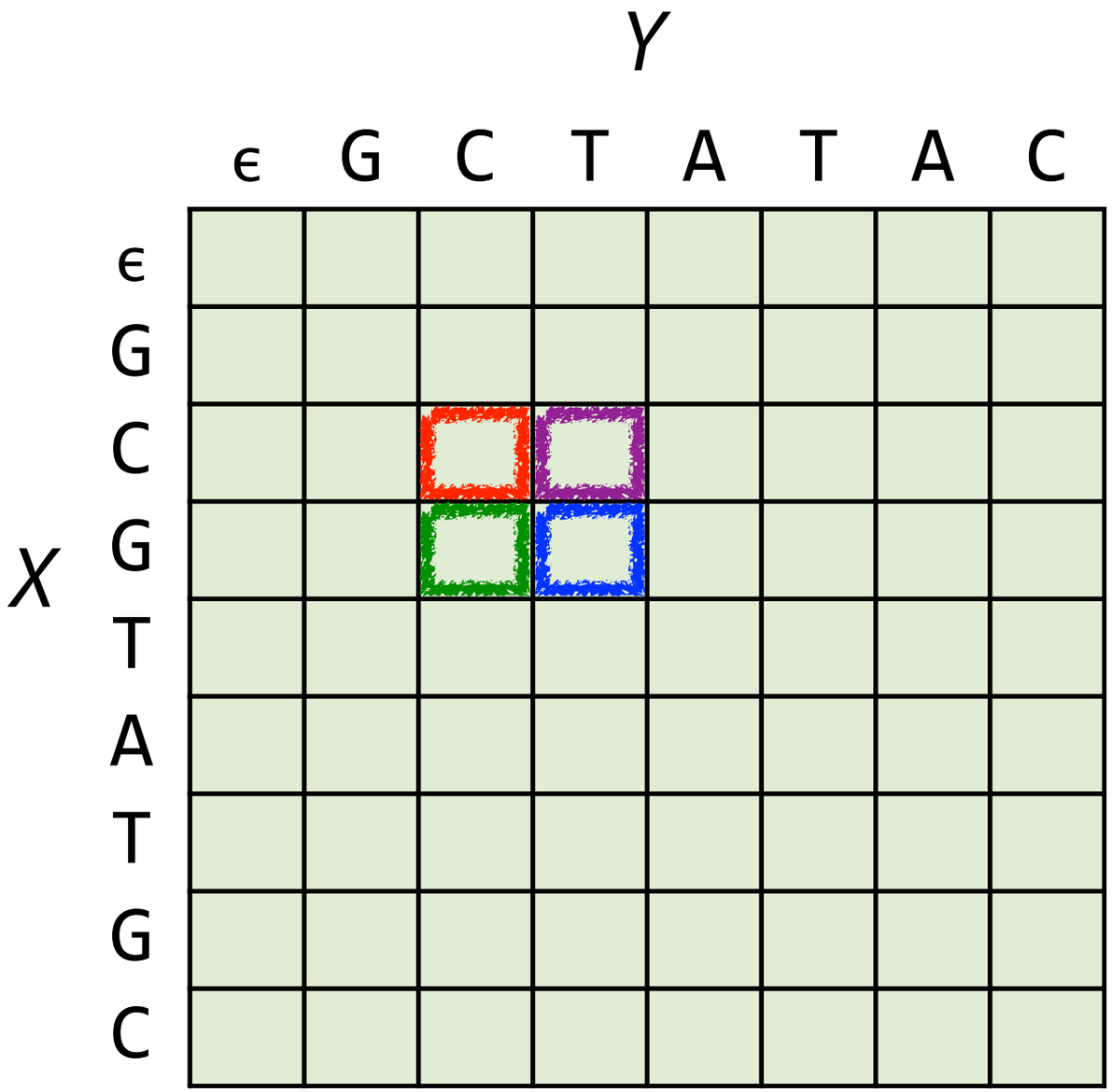

editDistRecursive("ABC", "BBC")
("ABC", "BB") ("AB", "BB") ("AB", "BBC")
("ABC", "B") ("AB", "B") ("AB", "BB")


edDistRecursive("ABC", "BBC")
("ABC", "BB") ("AB", "BB") ("AB", "BBC")
("ABC", "B") ("AB", "B") ("AB", "BB")

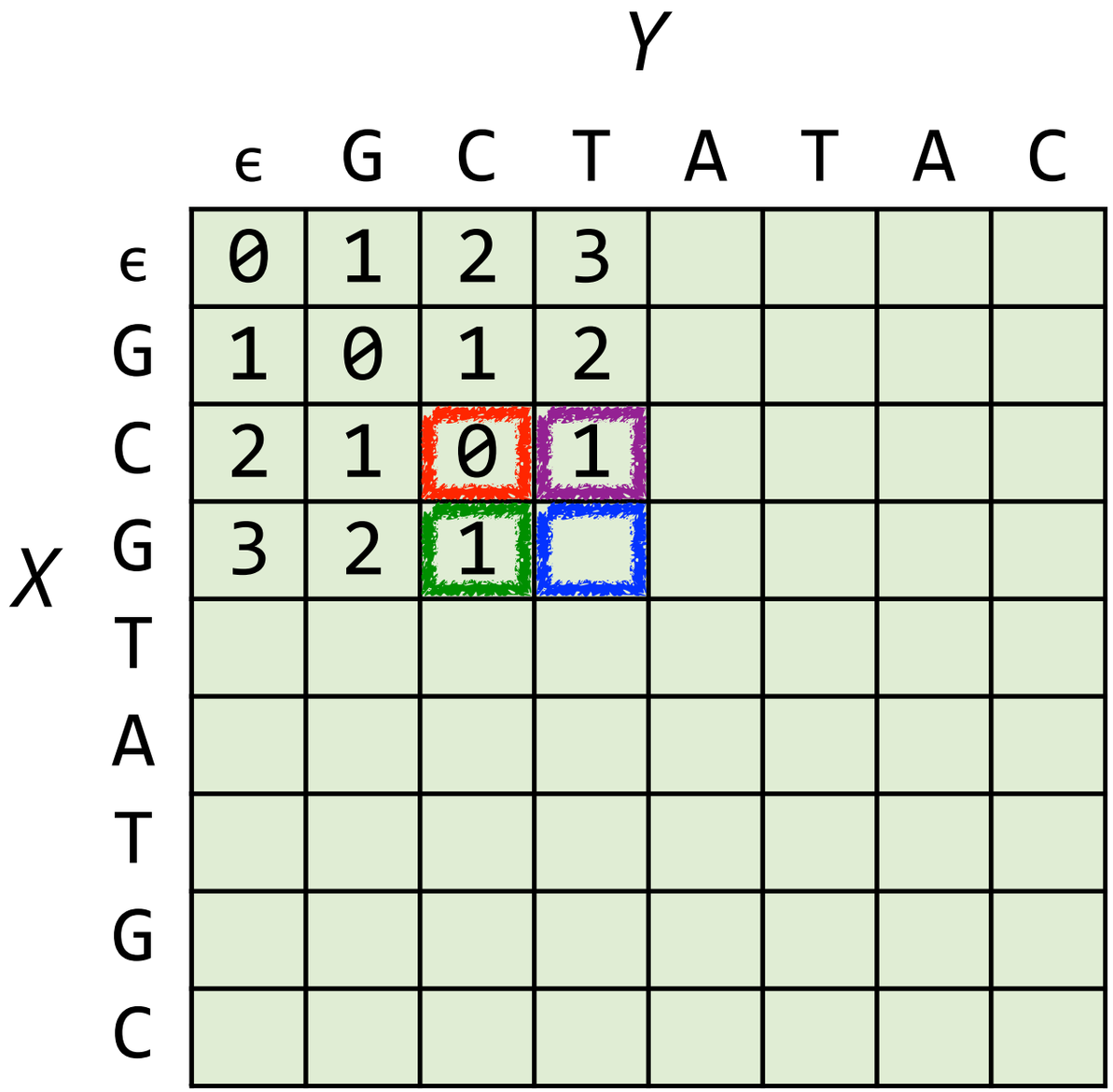

Y
ϵ G C T A T A C
X
ϵ
G
C
G
T
A
T
G
C


Y
ϵ G C T A T A C
X
ϵ
G
C
G
T
A
T
G
C
edist (GCG, GCT)


Y
ϵ G C T A T A C
X
ϵ
G
C
G
T
A
T
G
C
edist (GCGTATGC, GCTATAC)



$$\texttt{edist}(\alpha x, \beta y) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + \delta(x, y) \\ \texttt{edist}(\alpha x, \beta) + 1 \\ \texttt{edist}(\alpha, \beta y) + 1 \end{cases}$$



$$\mathtt{edist}(\alpha x, \beta y) = \min \begin{cases} \mathtt{edist}(\alpha, \beta) + \delta(x, y) & = 0 + 1 = 1 \\ \mathtt{edist}(\alpha x, \beta) + 1 & = 1 + 1 = 2 \\ \mathtt{edist}(\alpha, \beta y) + 1 & = 1 + 1 = 2 \end{cases}$$

		Y							
		ϵ	G	C	T	A	T	A	C
	ϵ	0	1	2	3				
	G	1	0	1	2				
	C	2	1	0	1				
X	G	3	2	1	1				
	T								
	A								
	T								
	G								
	C								

$$\texttt{edist}(\alpha x, \beta y) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + \delta(x, y) & = 0 + 1 = 1 \\ \texttt{edist}(\alpha x, \beta) + 1 & = 1 + 1 = 2 \\ \texttt{edist}(\alpha, \beta y) + 1 & = 1 + 1 = 2 \end{cases}$$

		Y						
		ϵ	G	C	T	A	T	A	C
	ϵ	0	1	2	3	4	5	6	7
	G	1	0	1	2	3	4	5	6
	C	2	1	0	1	2	3	4	5
X	G	3	2	1	1	2	3	4	5
	T	4	3	2	1	2	2	3	4
	A	5	4	3	2	1	2	2	3
	T	6	5	4	3	2	1	2	3
	G	7	6	5	4	3	2	2	3
	C	8	7	6	5	4	3	3	2

← Final result

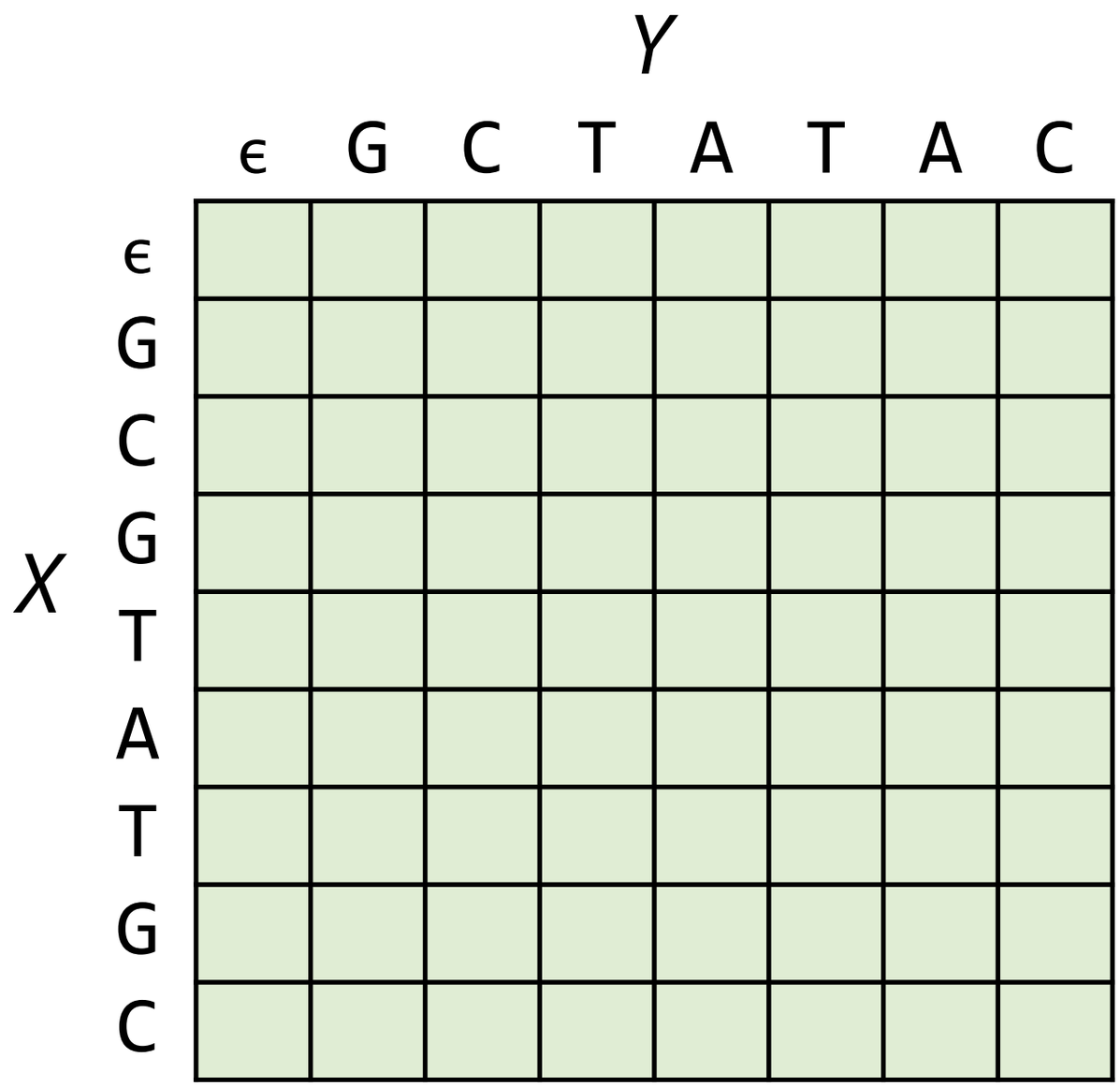
$$\mathbf{edist}(\alpha x, \beta y) = \min \begin{cases} \mathbf{edist}(\alpha, \beta) + \delta(x, y) \\ \mathbf{edist}(\alpha x, \beta) + 1 \\ \mathbf{edist}(\alpha, \beta y) + 1 \end{cases}$$

		Y							
		ϵ	G	C	T	A	T	A	C
X	ϵ	0	1	2	3	4	5	6	7
	G	1	0	1	2	3	4	5	6
	C	2	1	0	1	2	3	4	5
	G	3	2	1	1	2	3	4	5
	T	4	3	2	1	2	2	3	4
	A	5	4	3	2	1	2	2	3
	T	6	5	4	3	2	1	2	3
	G	7	6	5	4	3	2	2	3
	C	8	7	6	5	4	3	3	2

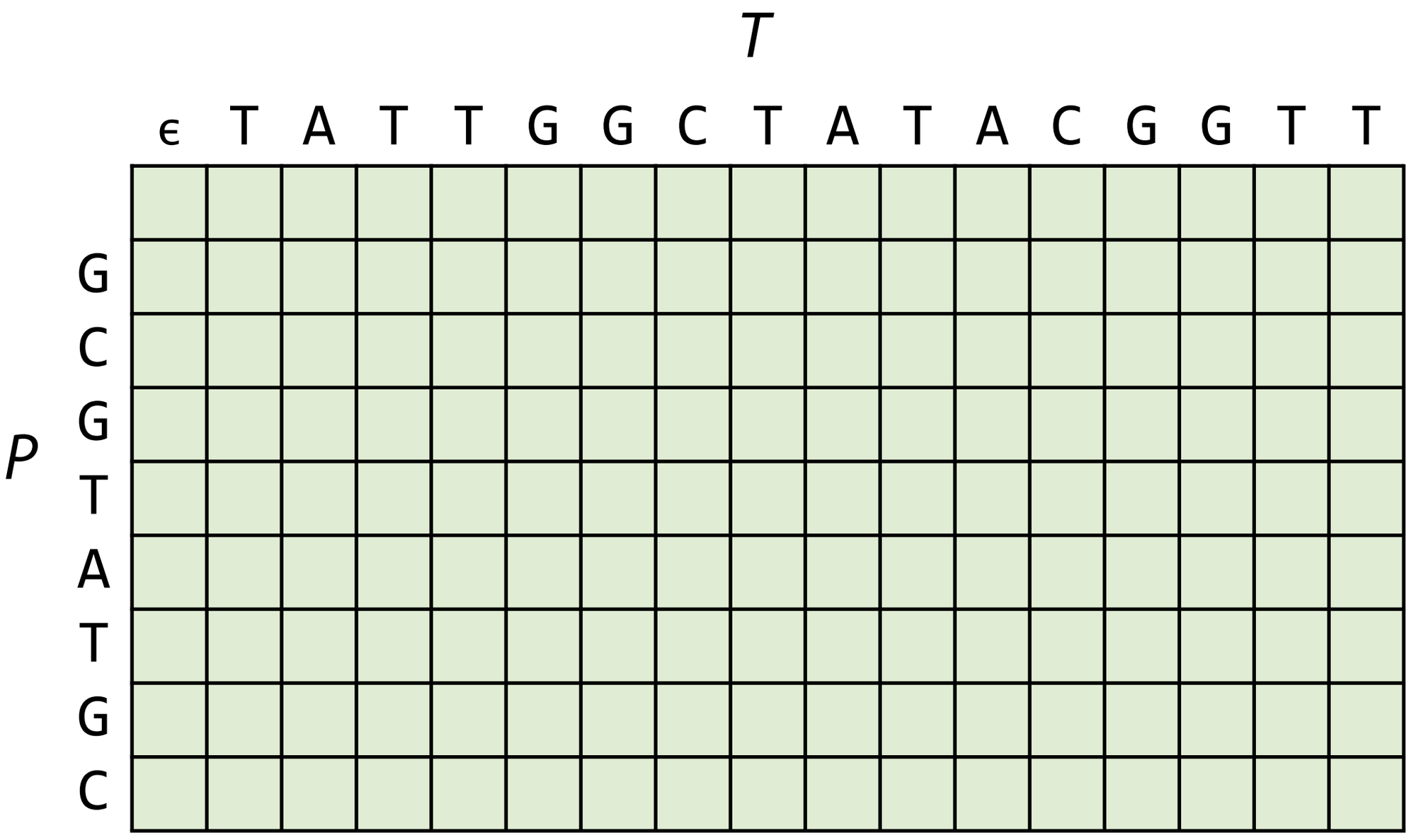
For any pair of prefixes from *X* & *Y*, edit distance is calculated *once*

Dynamic programming

Edit distance



Edit distance & approximate matching



			Y					
	ϵ	G	C	T	A	T	A	C
ϵ	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
G	**1**							
C	**2**							
G	**3**							
T	**4**							
A	**5**							
T	**6**							
G	**7**							
C	**8**							

X

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
offset →		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1																
	C	2																
	G	3																
P	T	4																
	A	5																
	T	6																
	G	7																
	C	8																

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

P occurs in *T* with 2 edits

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

How did I get here?

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

How did I get here?

T

P	ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
P	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

T

		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

Match occurs at offset 5

T

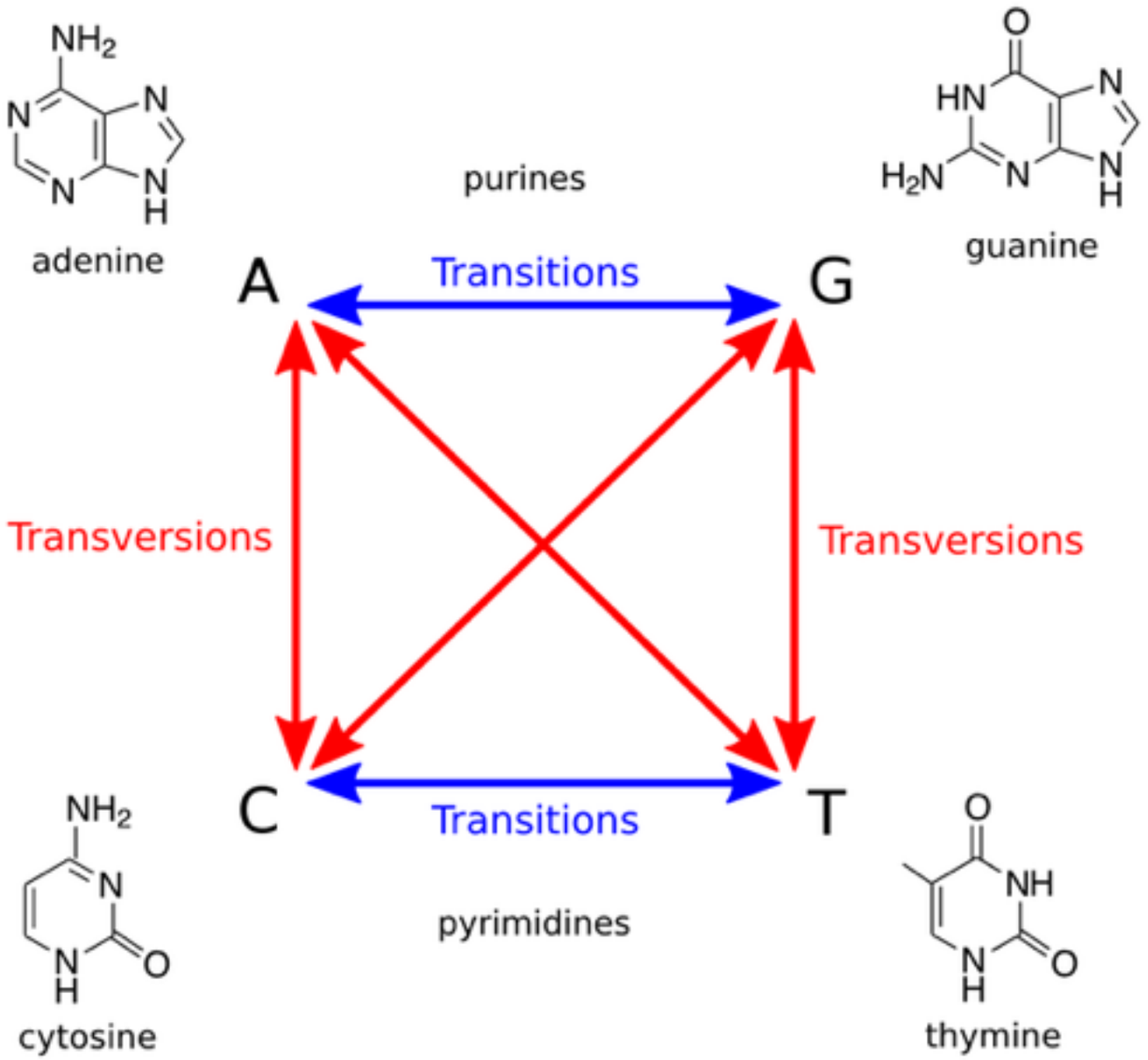
		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
offset →		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
P	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

```
T: G  C  -  T  A  T  A  C
   |  |     |  |  |     |
P: G  C  G  T  A  T  G  C
```

T

P		ϵ	T	A	T	T	G	G	C	T	A	T	A	C	G	G	T	T
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1
	C	2	2	2	2	2	1	1	0	1	2	2	2	1	1	1	1	2
	G	3	3	3	3	3	2	1	1	1	2	3	3	2	1	1	2	2
	T	4	3	4	3	3	3	2	2	1	2	2	3	3	2	2	1	2
	A	5	4	3	4	4	4	3	3	2	1	2	2	3	3	3	2	2
	T	6	5	4	3	4	5	4	4	3	2	1	2	3	4	4	3	2
	G	7	6	5	4	4	4	5	5	4	3	2	2	3	3	4	4	3
	C	8	7	6	5	5	5	5	5	5	4	3	3	2	3	4	5	4

		Y							
		ϵ	G	C	T	A	T	A	C
X	ϵ	0	1	2	3	4	5	6	7
	G	1	0	1	2	3	4	5	6
	C	2	1	0	1	2	3	4	5
	G	3	2	1	1	2	3	4	5
	T	4	3	2	1	2	2	3	4
	A	5	4	3	2	1	2	2	3
	T	6	5	4	3	2	1	2	3
	G	7	6	5	4	3	2	2	3
	C	8	7	6	5	4	3	3	2



Human *transition to transversion ratio* (AKA *ti/tv*) is ~2.1

G G G T A G C G G G T T T A A C
| | | | | | | | | | | | | | |
G G G T A A C G G G T T T A A C

Human substitution rate ≈ 1 in 1,000

G G G T A G C G G G T T T A A C
| | | | | | | | | | | | | | |
G G G T A – – G G G T T T A A C

Small-gap rate is ≈ 1 in 3,000

Penalty matrix

	A	C	G	T	-
A	0	4	2	4	8
C	4	0	4	2	8
G	2	4	0	4	8
T	4	2	4	0	8
-	8	8	8	8	

2 *Transitions* (A↔G, C↔T)

4 *Transversions*

8 *Gaps*

$$\texttt{edist}(\alpha x, \beta y) = \min \begin{cases} \texttt{edist}(\alpha, \beta) + \delta(x, y) \\ \texttt{edist}(\alpha x, \beta) + 1 \\ \texttt{edist}(\alpha, \beta y) + 1 \end{cases}$$

$$\texttt{galign}(\alpha x, \beta y) = \min \begin{cases} \texttt{galign}(\alpha, \beta) + p(x, y) \\ \texttt{galign}(\alpha x, \beta) + p(x, -) \\ \texttt{galign}(\alpha, \beta y) + p(-, y) \end{cases}$$

Use penalty matrix

Global alignment

	ϵ	T	A	T	G	T	C	A	T	G	C
ϵ	0	8	16	24	32	40	48	56	64	72	80
T	8	0	8	16	24	32	40	48	56	64	72
A	16	8	0	8	16	24	32	40	48	56	64
C	24	16	8	2	10	18	24	32	40	48	56
G	32	24	16	10	2	10	18	26	34	40	48
T	40	32	24	16	10	2	10	18	26	34	42
C	48	40	32	24	18	10	2	10	18	26	34
A	56	48	40	32	26	18	10	2	10	18	26
G	64	56	48	40	32	26	18	10	6	10	18
C	72	64	56	48	40	34	26	18	12	10	10

	A	C	G	T	-
A	0	4	2	4	8
C	4	0	4	2	8
G	2	4	0	4	8
T	4	2	4	0	8
-	8	8	8	8	

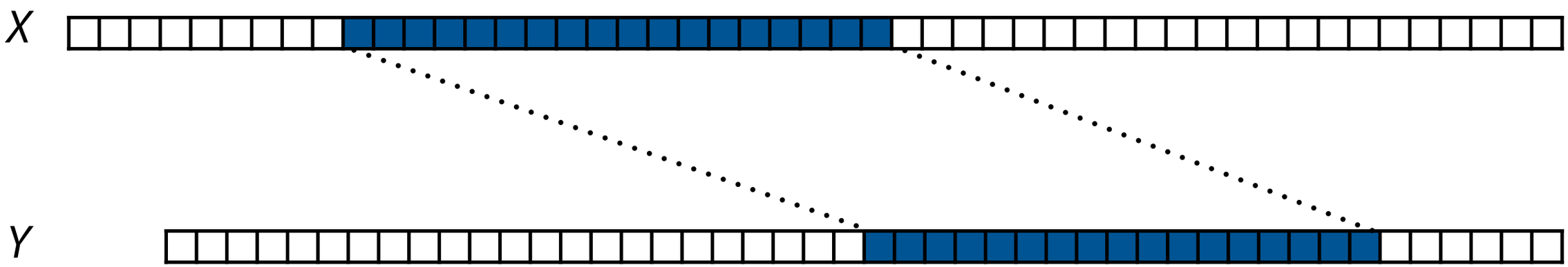
Global alignment

	ϵ	T	A	T	G	T	C	A	T	G	C
ϵ	0	8	16	24	32	40	48	56	64	72	80
T	8	0	8	16	24	32	40	48	56	64	72
A	16	8	0	8	16	24	32	40	48	56	64
C	24	16	8	2	10	18	24	32	40	48	56
G	32	24	16	10	2	10	18	26	34	40	48
T	40	32	24	16	10	2	10	18	26	34	42
C	48	40	32	24	18	10	2	10	18	26	34
A	56	48	40	32	26	18	10	2	10	18	26
G	64	56	48	40	32	26	18	10	6	10	18
C	72	64	56	48	40	34	26	18	12	10	8

	A	C	G	T	-
A	0	4	2	4	8
C	4	0	4	2	8
G	2	4	0	4	8
T	4	2	4	0	8
-	8	8	8	8	

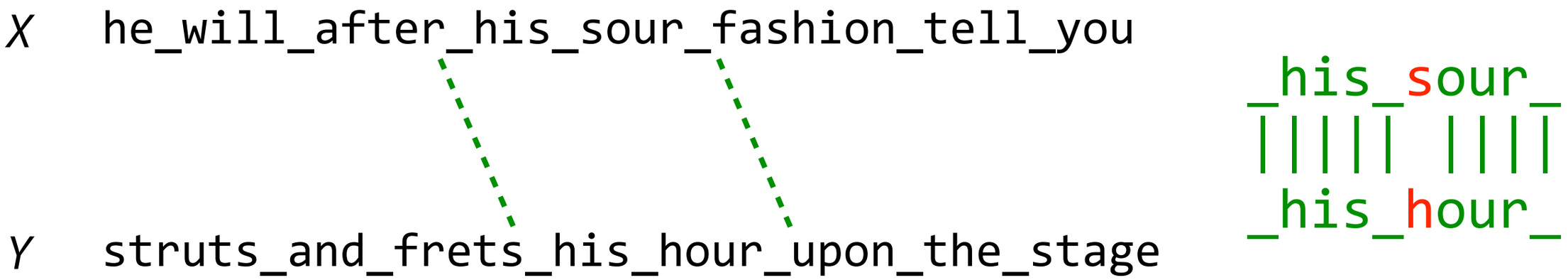
Local alignment

Find the most similar *pair of substrings* from X and Y



Local alignment

Find the most similar *pair of substrings* from X and Y



$$\texttt{lalign}(\alpha \mathrm{x}, \beta \mathrm{y}) = \max \begin{cases} \texttt{lalign}(\alpha, \beta) + s(\mathrm{x}, \mathrm{y}) \\ \texttt{lalign}(\alpha \mathrm{x}, \beta) + s(\mathrm{x}, \text{-}) \\ \texttt{lalign}(\alpha, \beta \mathrm{y}) + s(\text{-}, \mathrm{y}) \\ 0 \end{cases}$$

Scoring matrix: matches are positive, differences negative

	A	C	G	T	-
A	2	-4	-4	-4	-6
C	-4	2	-4	-4	-6
G	-4	-4	2	-4	-6
T	-4	-4	-4	2	-6
-	-6	-6	-6	-6	

Y

X		ϵ	T	A	T	A	T	G	C	G	G	C	G	T	T	T
	ϵ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0	0	0	2	0	2	2	0	2	0	0	0
	G	0	0	0	0	0	0	2	0	2	4	0	2	0	0	0
	T	0	2	0	2	0	2	0	0	0	0	0	0	4	2	2
	A	0	0	4	0	4	0	0	0	0	0	0	0	0	0	0
	T	0	2	0	6	0	6	0	0	0	0	0	0	2	2	2
	G	0	0	0	0	2	0	8	2	2	2	0	2	0	0	0
	C	0	0	0	0	0	0	2	10	4	0	4	0	0	0	0
	T	0	2	0	2	0	2	0	4	6	0	0	0	2	2	2
	G	0	0	0	0	0	0	4	0	6	8	2	2	0	0	0
	G	0	0	0	0	0	0	2	0	2	8	4	4	0	0	0
	C	0	0	0	0	0	0	0	4	0	2	10	4	0	0	0
	G	0	0	0	0	0	0	2	0	6	2	4	12	6	0	0
	C	0	0	0	0	0	0	0	4	0	2	4	6	8	2	0
	T	0	2	0	2	0	2	0	0	0	0	0	0	8	10	4
	A	0	0	4	0	4	0	0	0	0	0	0	0	2	4	6

		Y													
		ϵ	T	A	T	A	T	G	C	G	G	C	G	T	T	T
	ϵ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	G	0	0	0	0	0	0	2	0	2	2	0	2	0	0	0
	G	0	0	0	0	0	0	2	0	2	4	0	2	0	0	0
	T	0	2	0	2	0	2	0	0	0	0	0	0	4	2	2
	A	0	0	4	0	4	0	0	0	0	0	0	0	0	0	0
	T	0	2	0	6	0	4	0	0	0	0	0	0	2	2	2
	G	0	0	0	0	2	0	4	2	2	2	0	2	0	0	0
	C	0	0	0	0	0	0	2	6	4	0	4	0	0	0	0
X	T	0	2	0	2	0	2	0	4	6	0	0	0	2	2	2
	G	0	0	0	0	0	0	4	0	6	8	2	2	0	0	0
	G	0	0	0	0	0	0	2	0	2	4	4	4	0	0	0
	C	0	0	0	0	0	0	0	4	0	2	6	4	0	0	0
	G	0	0	0	0	0	0	2	0	6	4	4	6	0	0	0
	C	0	0	0	0	0	0	0	4	0	2	4	6	8	2	0
	T	0	2	0	2	0	2	0	0	0	0	0	0	8	10	4
	A	0	0	4	0	4	0	0	0	0	0	0	0	2	4	6

```
Y  T A T G C - G G C G
   | | | | |   | | | |
X  T A T G C T G G C G
```