



Algoritma Analizi

Dr. Öğr. Üyesi Emre DELİBAŞ

Dizi Karşılaştırma



Dizi Karşılaştırma

- Bioinformatikte yapılan çalışmaların en önemlilarından biri şüphesiz iki biyomoleküler dizinin karşılaştırılmasıdır.
- Çünkü bu karşılaştırma neticesinde ortaya çıkacak yüksek orandaki benzerlik, genellikle anlamlı yapısal veya fonksiyonel bir benzerliktir.



Dizi Karşılaştırma

- *Dizi karşılaştırması*, dizilerin hangi parçalarının benzer hangi parçalarının farklı olduğunu ortaya çıkarma problemi olarak tanımlanabilir.
- Bu işlem, çoklu dizi hizalama ve türler arasındaki evrimsel ilişkiyi açıklayan filogenetik ağaçlar gibi daha kompleks problemlerin temelini oluşturmaktadır.



Dizi Karşılaştırma

- İlgilenilen temel nokta iki dizi arasındaki benzerliklerin ve farkların ortaya konulmasıdır.
- Genellikle burada benzerliğin ölçüsü önemlidir.
- Bu problemi çözmek için en tipik yaklaşım, iki dizi arasında iyi ve anlamlı bir ***hızalama*** bulmaktır.
- Daha sonra bir ***skor şeması*** verilerek ***benzerlikleri (similarity)*** hesaplanabilir.



Dizi Hizalama ve Benzerlik

- İki dizinin hizalanması (diziler farklı uzunlukta olabilir), diziyi oluşturan harflerin arasına boşluklar koyarak ardışık sütunlarda yer alan aynı veya benzer harflerin bir hizada olmasını (alt alta gelmesini) sağlama prensibine dayanır.
- A= “ACAAGACAGCGT” ve
- B= “AGAACAAAGGCGT” dizileri hizalandığında;

D	D	V	Y	I	L	D	Q	A	E	E	A	G	H	D	L	P	Y	S	C	R	A	G	
D	N	V	Y	I	L	D	Q	A	E	E	A	G	H	D	L	P	Y	S	C	R	A	G	
D	D	V	Y	I	L	D	Q	A	E	E	E	G	H	D	L	P	Y	S	C	R	A	G	
P	D	V	Y	I	L	D	Q	A	E	E	E	E	G	H	D	L	P	Y	S	C	R	A	G
S	D	V	Y	I	L	D	H	A	E	E	V	G	I	D	L	P	Y	S	C	R	A	G	
P	D	V	Y	I	L	D	H	A	E	E	V	G	I	E	L	P	Y	S	C	R	A	G	
P	D	V	Y	I	L	D	H	A	E	E	E	A	G	I	D	L	P	Y	S	C	R	A	G
P	D	V	Y	I	L	D	H	A	E	E	E	E	G	I	D	L	P	Y	S	C	R	A	G
P	D	V	Y	V	L	D	A	A	E	E	A	G	I	D	L	P	Y	S	C	R	A	G	
P	D	V	Y	V	L	D	A	A	E	E	A	G	I	D	L	P	Y	S	C	R	A	G	
P	E	D	V	Y	V	L	D	A	A	E	E	A	G	I	D	L	P	Y	S	C	R	A	G
P	E	D	V	Y	V	L	D	A	A	E	E	A	G	I	D	L	P	Y	S	C	R	A	G
P	D	V	Y	I	L	D	Q	A	E	E	D	G	I	D	L	P	Y	S	C	R	A	G	

A = ACAAGACAG - CGT

| | | | | | | |

B = AGAACAA - AGGCCGT

Dizi Hizalama ve Benzerlik

A = ACAAGACAG - CGT

| | | | | | | | |

B = AGAACCA - AGGCCGT

- Burada amaç aynı alt dizileri olabildiğince eşleştirmektir.
- Örnekte dokuz eşleşme aralarındaki dikey barlarla işaretlenmiştir.
- Buna karşın, diziler aynı değilse eşleşmemeye durumu hizalandığı harflerdeki farklılıkla ortaya çıkar.



Karşılaştırmada Kullanılan Terimler

(Qry)	A C D E F G	A C D E F G	A C D E F G	A C – E F G
(Sbj)	A C D E F G	A C L E F G	A C – E F G	A C D E F G
Biological Event	Conservation	Substitution	Insertion	Deletion
Alignment Represent	Match	Mismatch	Gap	Gap

Edit-Distance

- Biyolojik uygulamalarda yaygın olarak kullanılan aşağıdaki tabloya göre sayısal bir skor üretilir

<i>operation</i>	<i>cost</i>
<i>insert a gap</i>	2
<i>align two characters that mismatch</i>	1
<i>align two characters that match</i>	0

- $x = \text{"AACAGTTACC"}$ ve
 $y = \text{"TAAGGTCA"}$ dizilerinin
iki olası hizalaması şöyledir:

x	y	cost	x	y	cost
A	T	1	A	T	1
A	A	0	A	A	0
C	A	1	C	-	2
A	G	1	A	A	0
G	G	0	G	G	0
T	T	0	T	G	1
T	C	1	T	T	0
A	A	0	A	-	2
C	-	2	C	C	0
C	-	2	C	A	1
		8			7

Skor Matrisi

	S	a	t	u	r	d	a	y	
s	0	1	2	3	4	5	6	7	8
u	1	0 ← 1 ← 2	3	4	5	6	7		
n	2	1	1	2	2	3	4	5	6
d	3	2	2	2	3	3	4	5	6
a	4	3	3	3	3	4	3	4	5
a	5	4	3	4	4	4	4	3	4
y	6	5	4	4	5	5	5	4	3

← OPERATION #1 - Delete "a"
← OPERATION #2 - Delete "t"
↑ OPERATION #3 - Transform "r" to "n"

← 3 operations in total

S - -UNDAY



SATURDAY

Hangi hizalama daha iyi?

AGGCTACT~T~T~CA

GGCTACTATATCA

AGGCTACTTT~~CA

GGCTACTATATCA

Gap Penalty

- Boşluklar doğal değildir, boşluklar diziler arasındaki evrimsel ilişkileri anlamak için yaptığımız şeydir.
- Boşlukları dahil ettiğimizde dizideki her boşluk için boşluk cezamız var.
- Sıralamada boşluk eklerken çok dikkatli olmalıyız, gereksiz iki boşluk yerine bir substitution olması daha iyi sayılır.



Gap Penalty

- **Linear Gap Penalty:** Tüm dizinin boşluk değeri, boşluk karakterlerinin toplam sayısı ile bir sabitin çarpımıdır.
- **Affine Gap Penalty:** Biyolojik dizilerde, tek bir ekleme veya silme olayı nedeniyle bir dizide 10 uzunlığında büyük bir boşluğun meydana gelmesi, 1 uzunlığında 10 küçük aralığın oluşmasından çok daha olasıdır. Bu nedenle, afin boşluk cezaları uzunluğa bağımlıdır ve bir boşluk açma cezası, a ve bir boşluk uzatma cezası, e kullanır. L uzunlığundaki bir boşluğa $a + (L-1)*e$ cezası verilir.



Gap Penalty



1

AGGCTAC**T~T~T**CA
GGCTAC**TATAT**CA

-5 -5

Gap penalty: -10

2

AGGCTAC**TTT~~~**CA
GGCTAC**TATAT**CA

-5 -1

Gap penalty: -6

Dizi Hizalama Yöntemleri

- Çok kısa veya çok benzer diziler elle hizalanabilir. Ancak çoğu ilginç problem, insan eliyle yapılamayacak kadar uzun, karmaşık veya çok sayıda dizinin hizalanmasını gerektirir.
 - Dizi hizalaması için kullanılan hesaplamalı yöntemler genelde iki gruba ayrılır: global hizalama ve yerel hizalama.

Dizi Hizalama Yöntemleri

- ***Global hizalamanın*** bulunması bir global optimizasyon çeşididir ve elde edilecek hizalamayı, sorgulanen dizilerin tamamını kapsamaya "zorlar".
 - Buna karşın, ***yerel hizalamalar*** genelde birbirinden çok farklılık gösteren uzun dizilerde benzer bölgeleri tespit eder.

Dizi Hizalama Yöntemleri

- Dizi hizalama problemi için çeşitli algoritmalar uygulanmıştır.
 - Bunların bazıları, **dinamik programlama** gibi, yavaş ama formel olarak optimum yöntemlerdir, bazıları ise hızlı ama mutlaka mükemmel sonucu vermeyebilen, veri tabanı aramaları için tasarlanmış **sezgisel algoritmalar** veya **olasılıksal yöntemlerdir**.

İkili Hizalama Algoritmaları

- Global hizalamalarda her dizideki her harfin hizalanması amaçlanır. Sorgu kümesindeki diziler birbirine benzer ve yaklaşık aynı uzunlukta olursa global hizalamaları en yararlı olur.
 - Global hizalama tekniklerinden biri, dinamik programlamaya dayalı olan **Needleman-Wunsch Algoritmasıdır.**

İkili Hizalama Algoritmaları

- Birbirine benzemeyen ama benzer bölgeler içerdığı tahmin edilen diziler için lokal hizalamalar daha yararlıdır. Keza, benzer kısa dizi motiflerinin tespitinde lokal hizalamalar kullanılır.
 - *Smith-Waterman Algoritması* da dinamik programlamaya dayalı bir lokal hizalama yöntemidir.

Alignment

Global Alignment

Target Sequence

5' ACTACTAGATTACCTACGGATCAGGTACTTAGAGGCTTGCAACCA 3'

||||||| ||||| ||||| ||||| |||||

5' ACTACTAGATT----ACGGATC--GTACTTAGAGGCTAGCAACCA 3'

Query Sequence

Tries to align entire sequence

Align all letters from query and target

Suitable for closely related sequences

A general Global Alignment Method
is

NEEDLEMAN - WUNSCH

GLOBAL VS. LOCAL SEQUENCE ALIGNMENT

Local Alignment

Target Sequence

5' ACTACTAGATTACCTACGGATCAGGTACTTAGAGGCTTGCAACCA 3'

||||| ||||| ||||| |||||

Query Sequence 5' TACTCACGGATGAGGTACTTAGAGGC 3'

Align regions having highest similarities

Align substring of target with substring
of query

Suitable for more divergent sequences

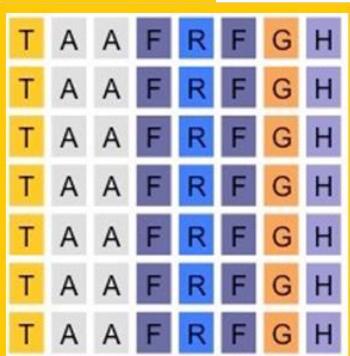
A general Local Alignment Method
is

SMITH - WATERMAN

İkili Hizalama Algoritmaları

- Needleman-Wunsch Algoritması X ve Y dizileri arasındaki en iyi global hizalama problemini çözen bir dinamik programlama algoritmasıdır. Algoritma ana hatlarıyla üç aşamadan oluşur:
 - Başlatma - Initialization,
 - Skorlama – Matrix Filling,
 - Başlangıç noktasına dönüş - Traceback (Hizalama).

Needleman-Wunsch Algoritması



Başlatma:

$d > 0$ olmak şartı ile;

$$F(i, 0) := i.d, \quad i = 0, 1, 2, \dots, n$$

$$F(0,j) \coloneqq j \cdot d, \quad j = 0, 1, 2, \dots, m$$

Ana İterasyon:

$$F(i, j) := \max \begin{cases} F(i - 1, j - 1) + s(x_i, y_j) & Durum1 \\ F(i - 1, j) + d & Durum2 \\ F(i, j - 1) + d & Durum3 \end{cases}$$

Sonlandırma:

$$P_{tr}(i,j) := \max \begin{cases} \text{CAPR.} & \text{Durum1} \\ SOL & \text{Durum2} \\ \text{ÜST} & \text{Durum3} \end{cases}$$

		A	T	G	C	T
A						
G						
C						
T						

Rewards and Penalties

Match : 1 
 Mismatch : -1 
 GAP : -2 

Sequences

Seq 1 = A T G C T

Seq 2 = A G C T

		A	T	G	C	T
A						
G						
C						
T						

1 + M

Rewards and Penalties

Match	:	1	
Mismatch	:	-1	
GAP	:	-2	

Sequences

Seq 1 = A T G C T

M

Seq 2 = A G C T

Rewards and Penalties

Match	:	1	
Mismatch	:	-1	
GAP	:	-2	

		A	T	G	C	T
A						
G						
C						
T						

Sequences

Seq 1 = A T G C T

Seq 2 = A G C T

		A	T	G	C	T
	0					
A						
G						
C						
T						

Rewards and Penalties

Match : 1 
 Mismatch : -1 
 GAP : -2 

Sequences

Seq 1 = A T G C T

Seq 2 = A G C T

		A	T	G	C	T
	0	-2				
A						
G						
C						
T						

Rewards and Penalties

Match : 1 
 Mismatch : -1 
 GAP : -2 

Sequences

Seq 1 = A T G C T

Seq 2 = A G C T

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

Rewards and Penalties

Match : 1 
 Mismatch : -1 
 GAP : -2 

Sequences

Seq 1 = ATGCT

Seq 2 = AGCT

Rewards and Penalties

Match : 1 
Mismatch : -1 
GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A						
G						
C						
T						

Sequences

Seq 1 = A T G C T

Seq 2 = A G C T

Rewards and Penalties

Match : 1 
Mismatch : -1 
GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

Sequences

Seq 1 = A T G C T

Seq 2 = A G C T

Rewards and Penalties

Match : 1 
Mismatch : -1 
GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

Sequences

Seq 1 = A T G C T

Seq 2 = A G C T

A

0	-2
-2	-2

A

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

Match : 1 
Mismatch : -1 
GAP : -2 

Value form LEFT

-

Value form UP

-

Value form DIAGONAL

-

Match : 1 
Mismatch : -1 
GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

0	-2
-2	-2

 GAP

-2

A

Value form LEFT

-

Value form UP

-

Value form DIAGONAL

-

A

0	-2
-2	

A

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

Match : 1 
Mismatch : -1 
GAP : -2 

Value form LEFT

- **-4**

Value form UP

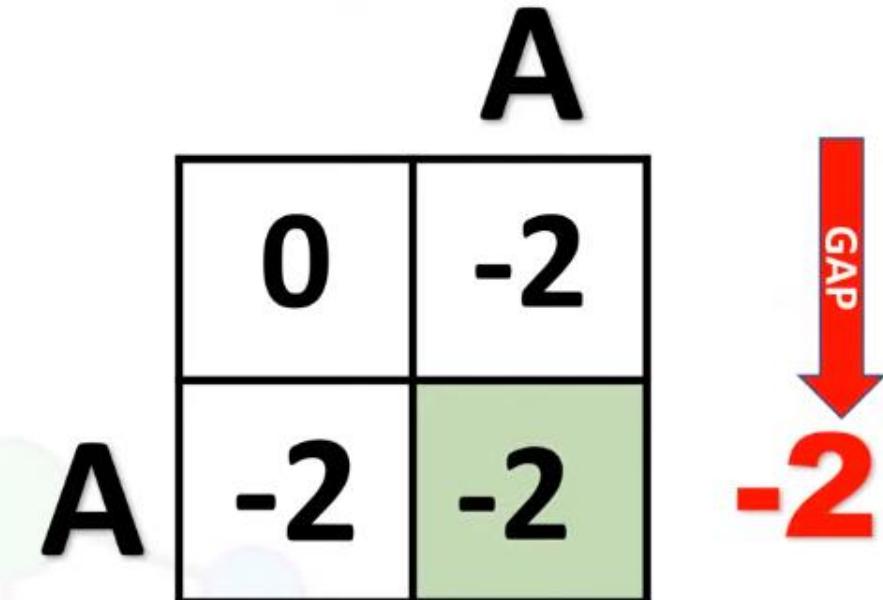
-

Value form DIAGONAL

-

Match : 1 
 Mismatch : -1 
 GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					



Value form LEFT

- **-4**

Value form UP

-

Value form DIAGONAL

-

Match : 1 
Mismatch : -1 
GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

A

0	-2
-2	0

DIAGONAL 

Value form LEFT

- **-4**
-4

Value form UP

- **-4**
-4

Value form DIAGONAL

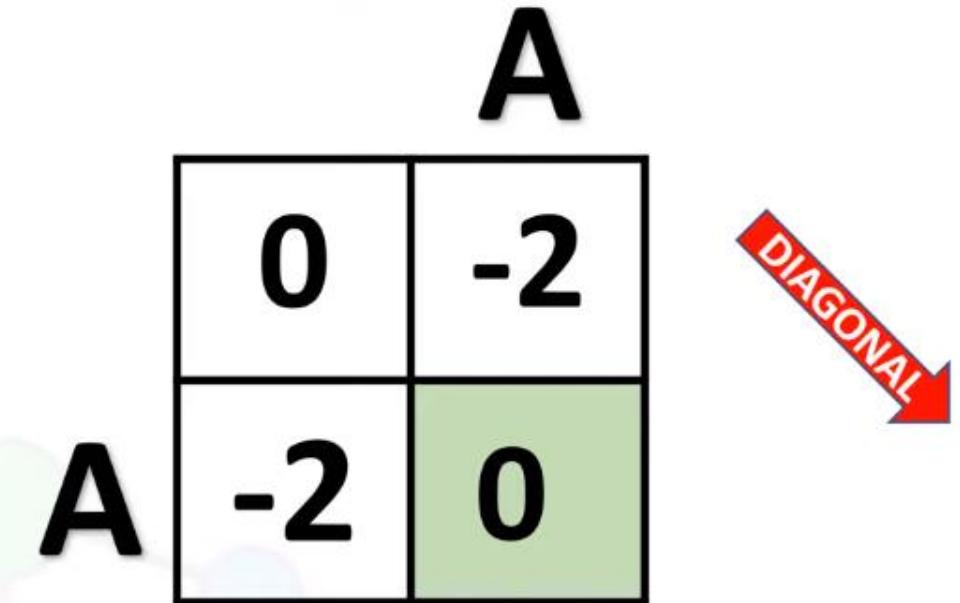
-

Match : 1 
 Mismatch : -1 
 GAP : -2 

A = A

MATCHED

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					



Value form LEFT

- **-4**
-4

Value form UP

- **-4**
-4

Value form DIAGONAL -

Match : 1 
 Mismatch : -1 
 GAP : -2 

A = A

MATCHED

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

A

0	-2
-2	01

DIAGONAL 

Value form LEFT

- **-4**
-4

Value form UP

- **-4**
-4

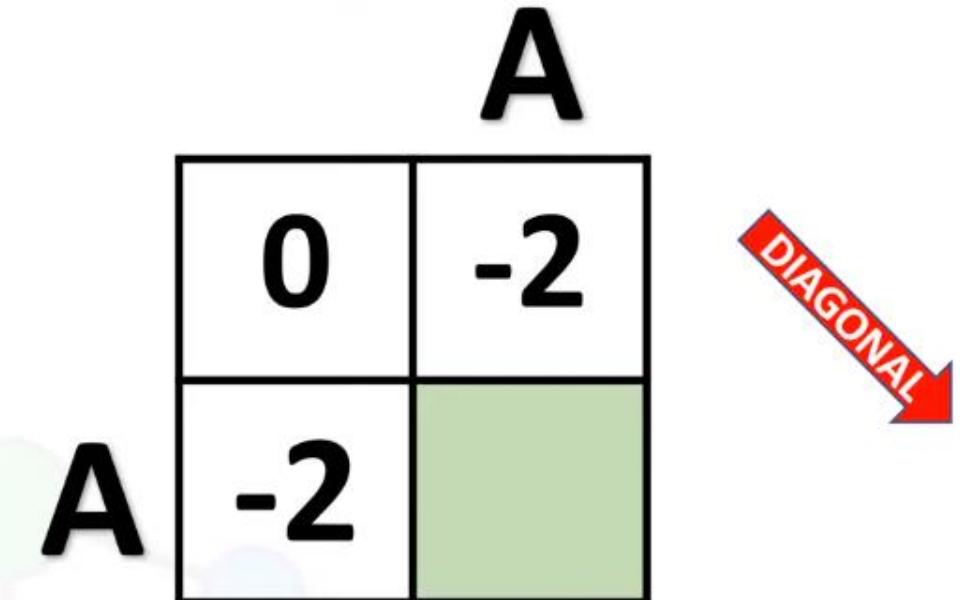
Value form DIAGONAL -

Match : 1 
 Mismatch : -1 
 GAP : -2 

A = A

MATCHED

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					



Value from LEFT

- **-4**

Value from UP

- **-4**

Value from DIAGONAL

- **1**

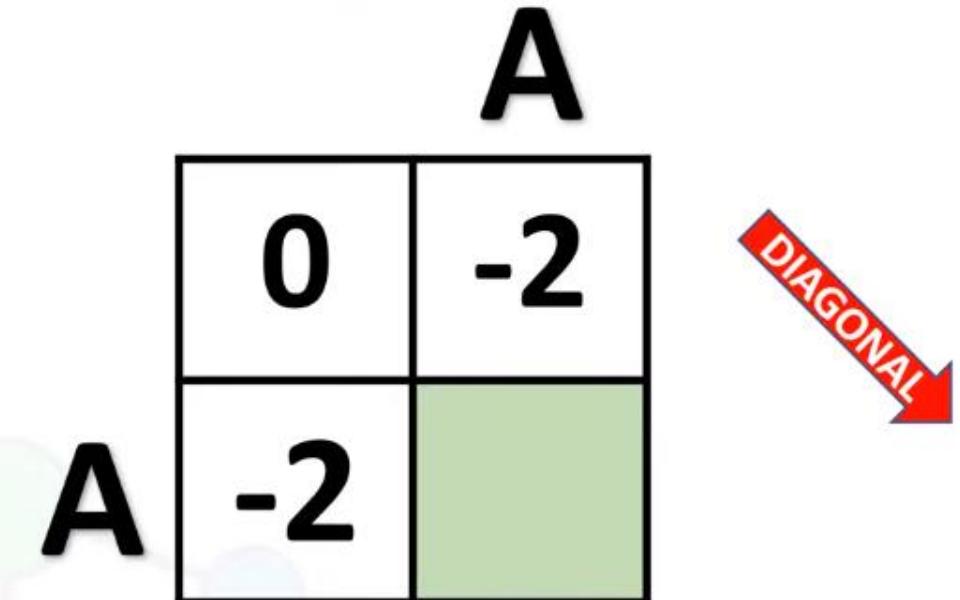
Maximum = **1**

Match : 1 
 Mismatch : -1 
 GAP : -2 

A = A

MATCHED

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1				
G	-4					
C	-6					
T	-8					



Value form LEFT

- **-4**

Value form UP

- **-4**

Value form DIAGONAL

- **1**

Maximum = **1**

Match : 1 
Mismatch : -1 
GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1				
G	-4					
C	-6					
T	-8					

-2	-4
1	

A

Value form LEFT -
Value form UP -
Value form DIAGONAL -

T

Match : 1 
 Mismatch : -1 
 GAP : -2 

		A	T	G	C	T	
	A	0	-2	-4	-6	-8	-10
A	-2	1					
G	-4						
C	-6						
T	-8						

-2	-4
1	1

 GAP

-2

A

-

Value form LEFT

-

Value form UP

-

Value form DIAGONAL

T

-2	-4
1	

A

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1				
G	-4					
C	-6					
T	-8					

Match : 1 
Mismatch : -1 
GAP : -2 

Value form LEFT - -1
Value form UP -
Value form DIAGONAL -

Match : 1 
 Mismatch : -1 
 GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1				
G	-4					
C	-6					
T	-8					

-2	-4
1	-4

 GAP
-2

A

Value form LEFT

- -1

Value form UP

-

Value form DIAGONAL

-

Match : 1 
 Mismatch : -1 
 GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1				
G	-4					
C	-6					
T	-8					

-2	-4
1	-2

 DIAGONAL

A

Value form LEFT

- **-1**

Value form UP

- **-6**

Value form DIAGONAL

-

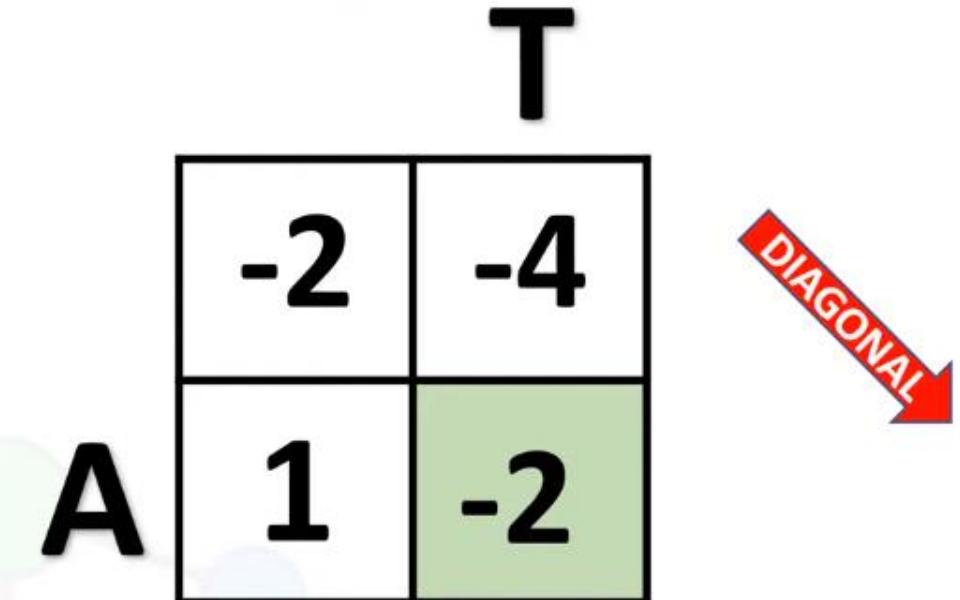
-6

Match : 1 
 Mismatch : -1 
 GAP : -2 

A ≠ T

MISMATCHED

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1				
G	-4					
C	-6					
T	-8					



Value form LEFT

- **-1**

Value form UP

- **-6**

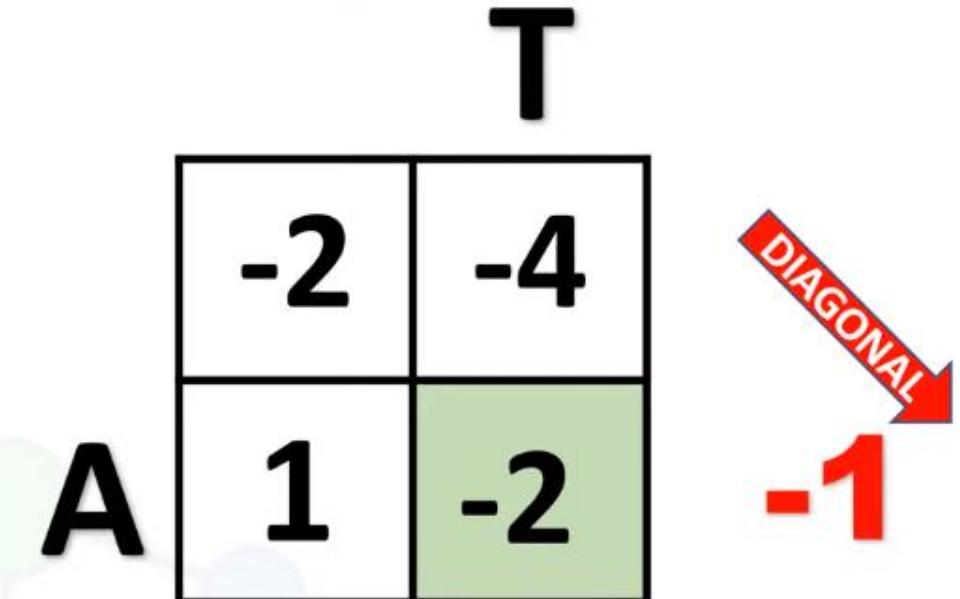
Value form DIAGONAL -

Match : 1 
 Mismatch : -1 
 GAP : -2 

A ≠ T

MISMATCHED

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1				
G	-4					
C	-6					
T	-8					



Value form LEFT

- **1**

Value form UP

- **6**

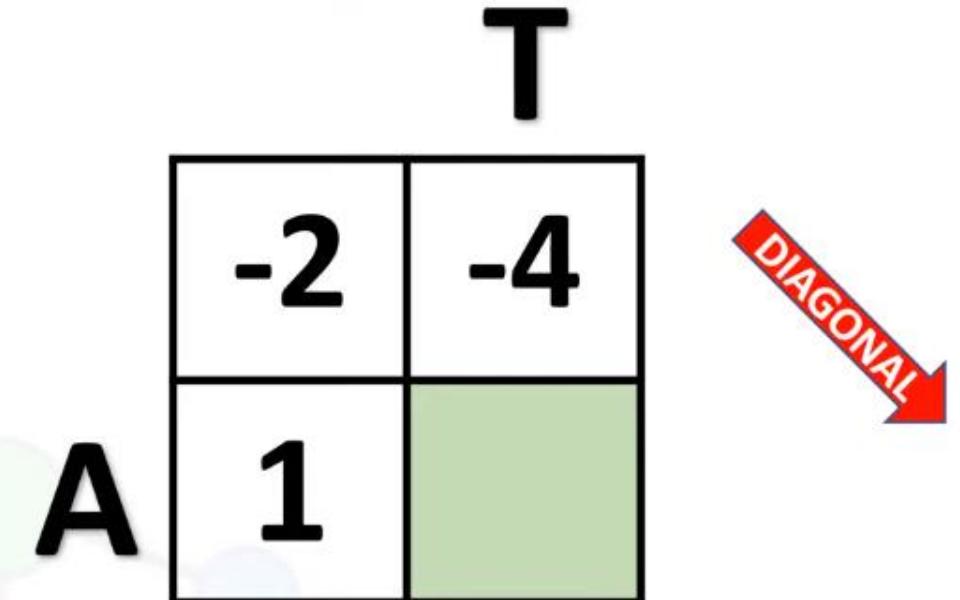
Value form DIAGONAL -

Match : 1 
 Mismatch : -1 
 GAP : -2 

A ≠ T

MISMATCHED

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1				
G	-4					
C	-6					
T	-8					



Value form LEFT

- **-1**

Value form UP

- **-6**

Value form DIAGONAL

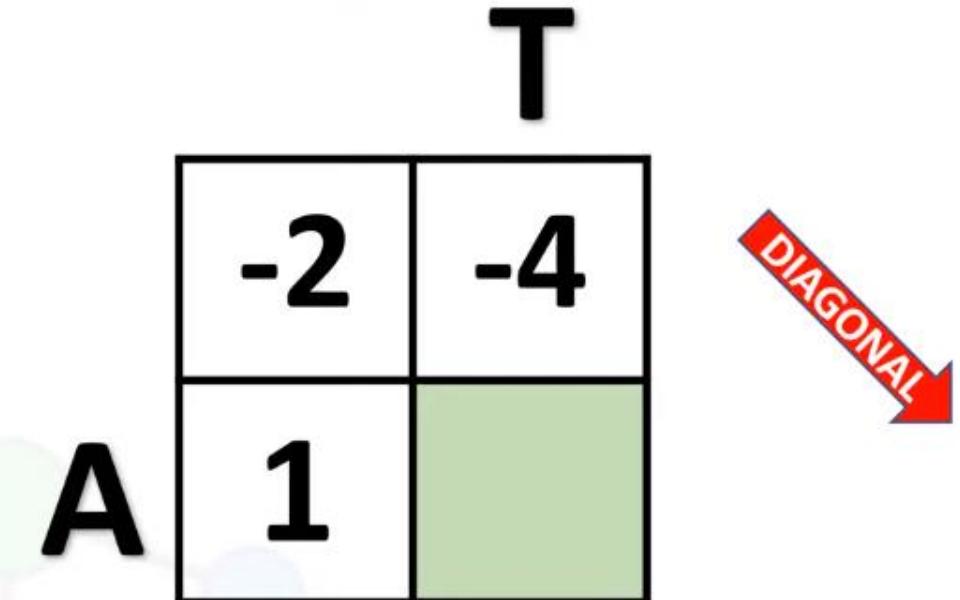
- **-3**

Match : 1 
 Mismatch : -1 
 GAP : -2 

A ≠ T

MISMATCHED

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1			
G	-4					
C	-6					
T	-8					



Value form LEFT

- **1**

Value form UP

- **6**

Value form DIAGONAL

- **3**

Maximum = **-1**

Match : 1 
Mismatch : -1 
GAP : -2 

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1				
C	-6	-3				
T	-8	-5				

Match : 1 
Mismatch : -1 
GAP : -2 

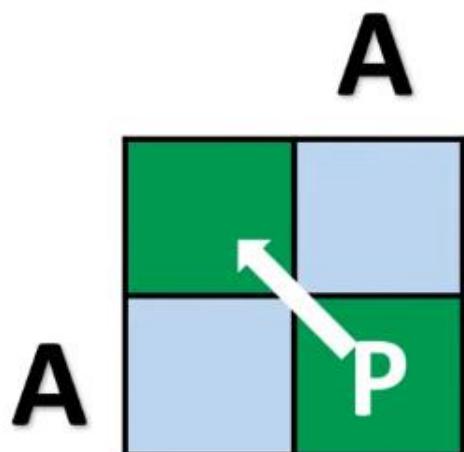
		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	

Traceback

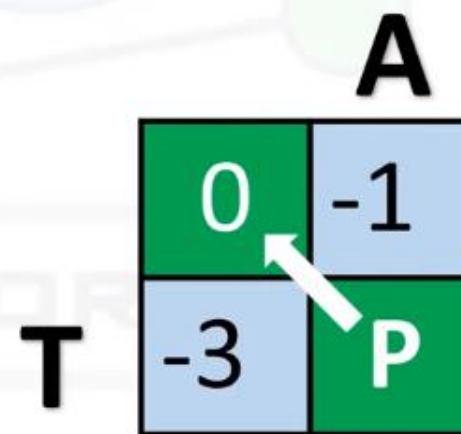
Beginning

From the right - bottom in the matrix where the highest score is present, continuing upto the upper left corner

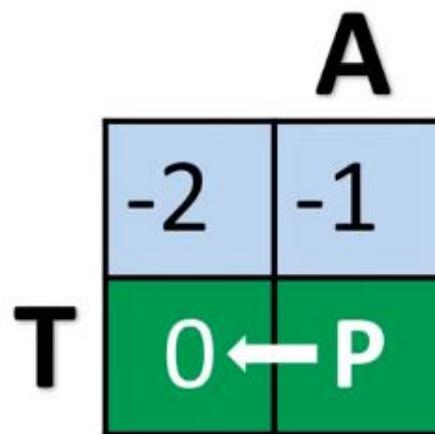
You can trace the arrows back that lead to the starting point



MATCHED



NOT MATCHED



		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

Starting from highest score i.e in the right most bottom

$$T = T$$

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

$C = C$

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

$G = G$

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

$A \neq T$

-2	-4
1	-1

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

$A \neq T$

	-1

-2 1 -4

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

$A \neq T$

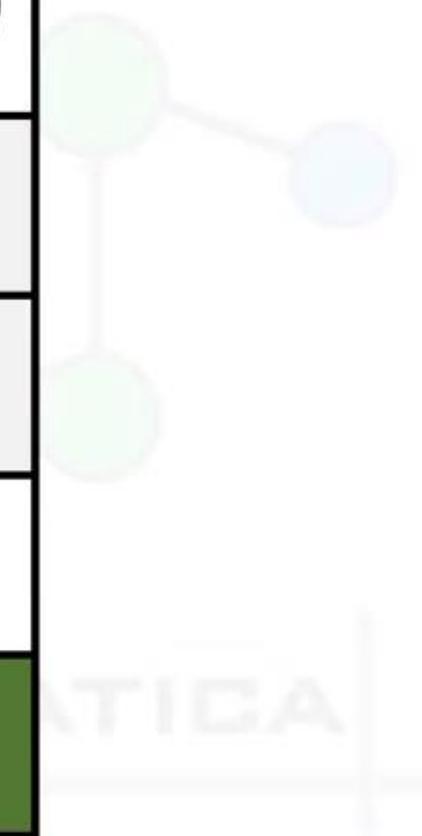
	-1

1

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1 ← -1	-3	-5	-7	
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

A = A

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1 ← -1	-3	-5	-7	
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2



smaller

4	0
-3	5

larger

MATCH

larger

3	0
-3	2

smaller

MISMATCH

-3	0
3	5

GAP

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2	1	-1	-3	-5	-7
G	-4	-1	0	0	-2	-4
C	-6	-3	-2	-1	1	-1
T	-8	-5	-2	-3	-1	2

Seq-1

A T G C T

Seq-2

A - G C T



SMITH WATERMAN

NEGATIVE VALUES

BECOMES

0

Rewards and Penalties

Mismatch

-1

Match

1

Gap

-2

INITIALIZATION

		A	T	G	C	T
A						
G						
C						
T						

Seq 1

Seq 1

ATGCT
AGCT

INITIALIZATION

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

Seq 1

Seq 1

ATGCT
AGCT

		A	T	G	C	T
	0	-2	-4	-6	-8	-10
A	-2					
G	-4					
C	-6					
T	-8					

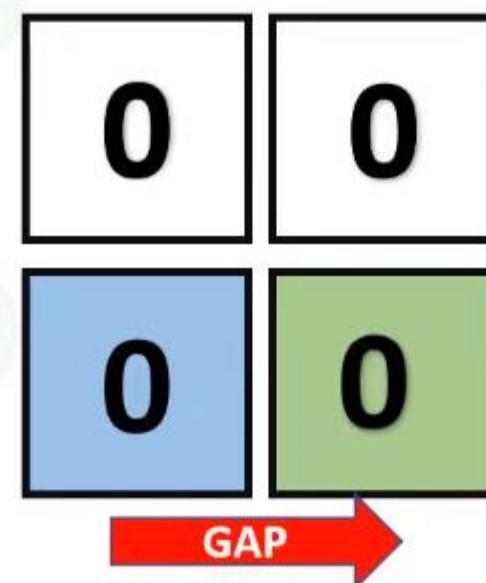
Negative → 0

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					

BIOINFORMATICA

MATRIX FILLING

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					

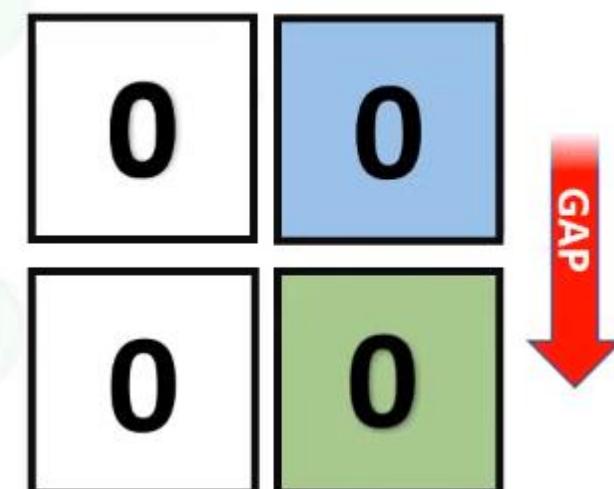


MATRIX FILLING

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					

0	0
0	-2
0	-2

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					



-2

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					

0	0
0	-2

-2

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					

0	0
0	0

-2 -2

A = A

MATCH

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					

0	0
0	1

-2

-2

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					

Negative



0

-2

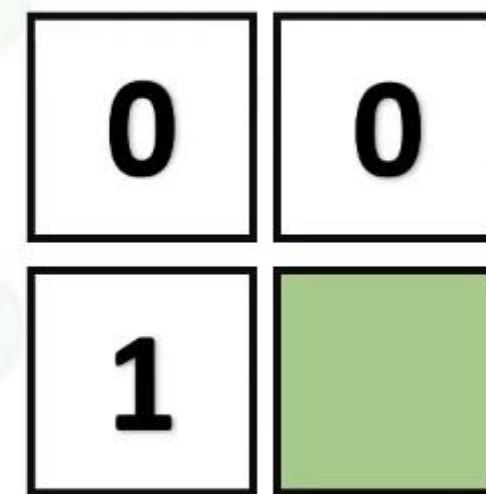
-2

1

		A	T	G	C	T
	0	0	0	0	0	0
A	0					
G	0					
C	0					
T	0					

0 0 1

		A	T	G	C	T
	0	0	0	0	0	0
A	0	1				
G	0					
C	0					
T	0					



A ≠ T MISMATCH

		A	T	G	C	T
	0	0	0	0	0	0
A	0	1				
G	0					
C	0					
T	0					

0	0
1	0

-1	-2
----	----

A ≠ T

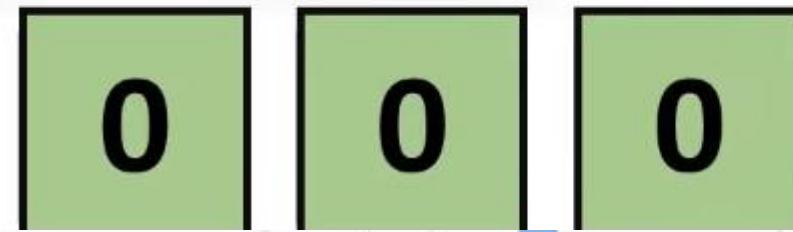
MISMATCH

		A	T	G	C	T
	0	0	0	0	0	0
A	0	1				
G	0					
C	0					
T	0					

0	0
1	-1

-1	-2
----	----

		A	T	G	C	T
	0	0	0	0	0	0
A	0	1				
G	0					
C	0					
T	0					



		A	T	G	C	T
	0	0	0	0	0	0
A	0	1	0	0	0	0
G	0	0	0	1	0	0
C	0	0	0	0	2	0
T	0	0	0	0	0	3

TRACE BACK

Strat from
Highest Value

An arrow
for
a base

0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	2	0	1	0	0	0
0	0	0	0	0	2	0	0
0	0	0	0	0	0	3	0
0	0	0	0	0	0	0	4
0	0	0	0	0	0	0	0

4 3 2 1 0

&

2 1 0

Proceed towards
Zeros

Multiple local
alignments
possible

		A	T	G	C	T
	0	0	0	0	0	0
A	0	1	0	0	0	0
G	0	0	0	1	0	0
C	0	0	0	0	2	0
T	0	0	0	0	0	3

GCT



GCT

Locally Aligned Sequences

Sequence 1

A T G C T

|||

A G C T

Sequence 2