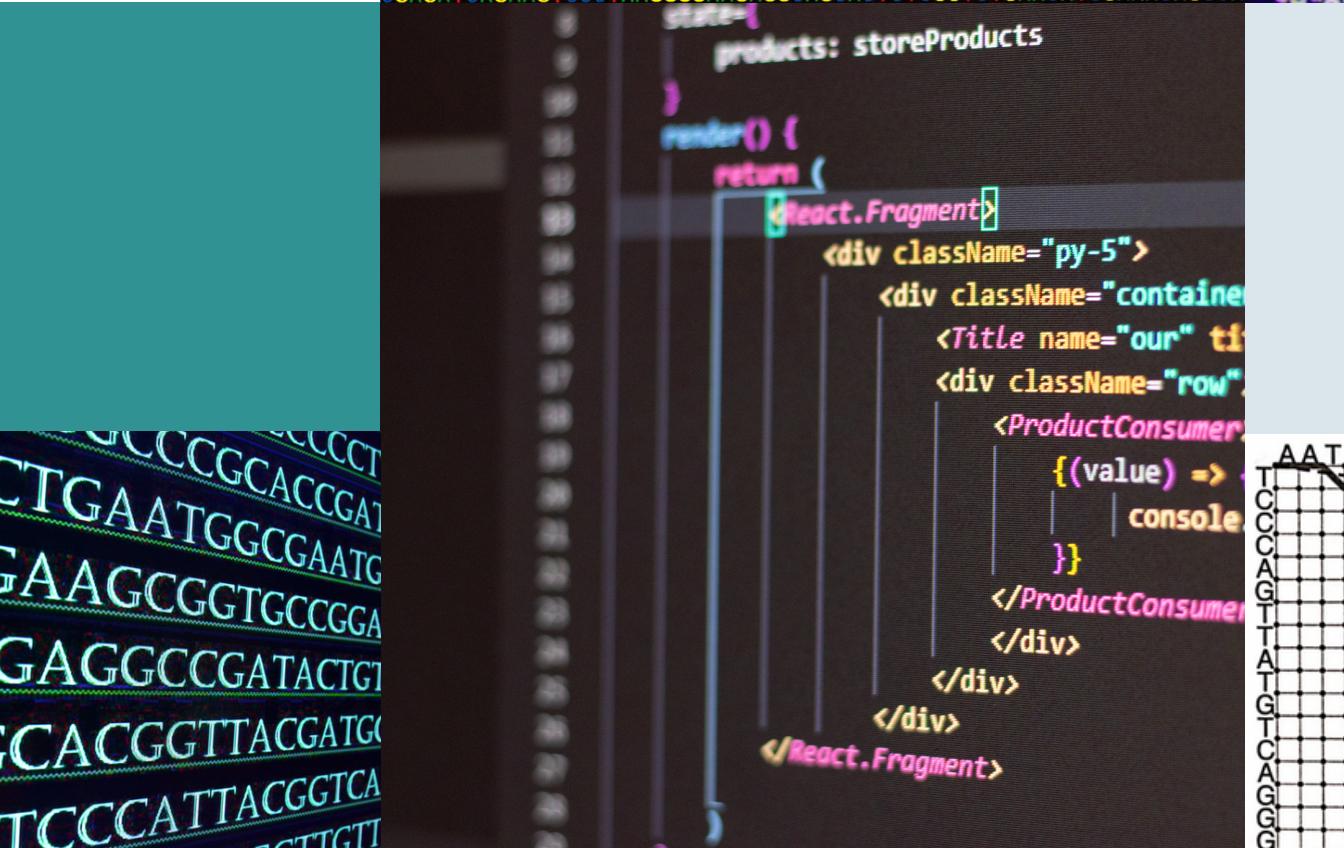


Algoritmo de Needleman-Wunsch

CCCGGCACCGAT
CTGAATGGCGAACATG
GAAGCGGTGCCGGA
GAGGCCGATACTGT
CACGGTTACGATG
TCCCATTACGGTCA
GGGGTTGT



```
state={ products: storeProducts }

}

render() {
  return (
    <React.Fragment>
      <div className="py-5">
        <div className="container">
          <Title name="our" title="Product Consumers" />
          <div className="row">
            <ProductConsumer value={this.state.products}>
              {(value) =>
                console.log(value)
              }
            </ProductConsumer>
            <div>
              <div>
                <div>
                  <h3>AAT</h3>
                  <table border="1">
                    <tr><td>T</td></tr>
                    <tr><td>C</td></tr>
                    <tr><td>C</td></tr>
                    <tr><td>A</td></tr>
                    <tr><td>G</td></tr>
                  </table>
                </div>
              </div>
            </div>
          </div>
        </div>
      </div>
    </React.Fragment>
  )
}
```



Algoritmo de Needleman-Wunsch

Proposto na década de 1970 por Saul Needleman e Christian Wunsch

- É um algoritmo baseado em programação dinâmica
- Realiza um alinhamento global entre um par de sequências



Needleman, S.B. e Wunsch, C.D.. "A general method applicable to the search for similarities in the amino acid sequence of two proteins." Journal of molecular biology 48.3 (1970): 443-453.

Programação dinâmica

Paradigma para o desenvolvimento de algoritmos

- **subestrutura ótima:** uma solução ótima para o problema contém em seu interior soluções ótimas para subproblemas
- **sobreposição de subproblemas:** um algoritmo recursivo reexamina o mesmo problema muitas vezes

Exemplo

Alinhamento de duas sequências **v** e **w**

ATCGTAC

ATGTTAT

Considerando que **v** tem n caracteres e **w**, tem m , o alinhamento poderá ter, no máximo $(n+m)$ posições

Exemplo

Alinhamento de duas sequências v e w

AT_CGT-A-C

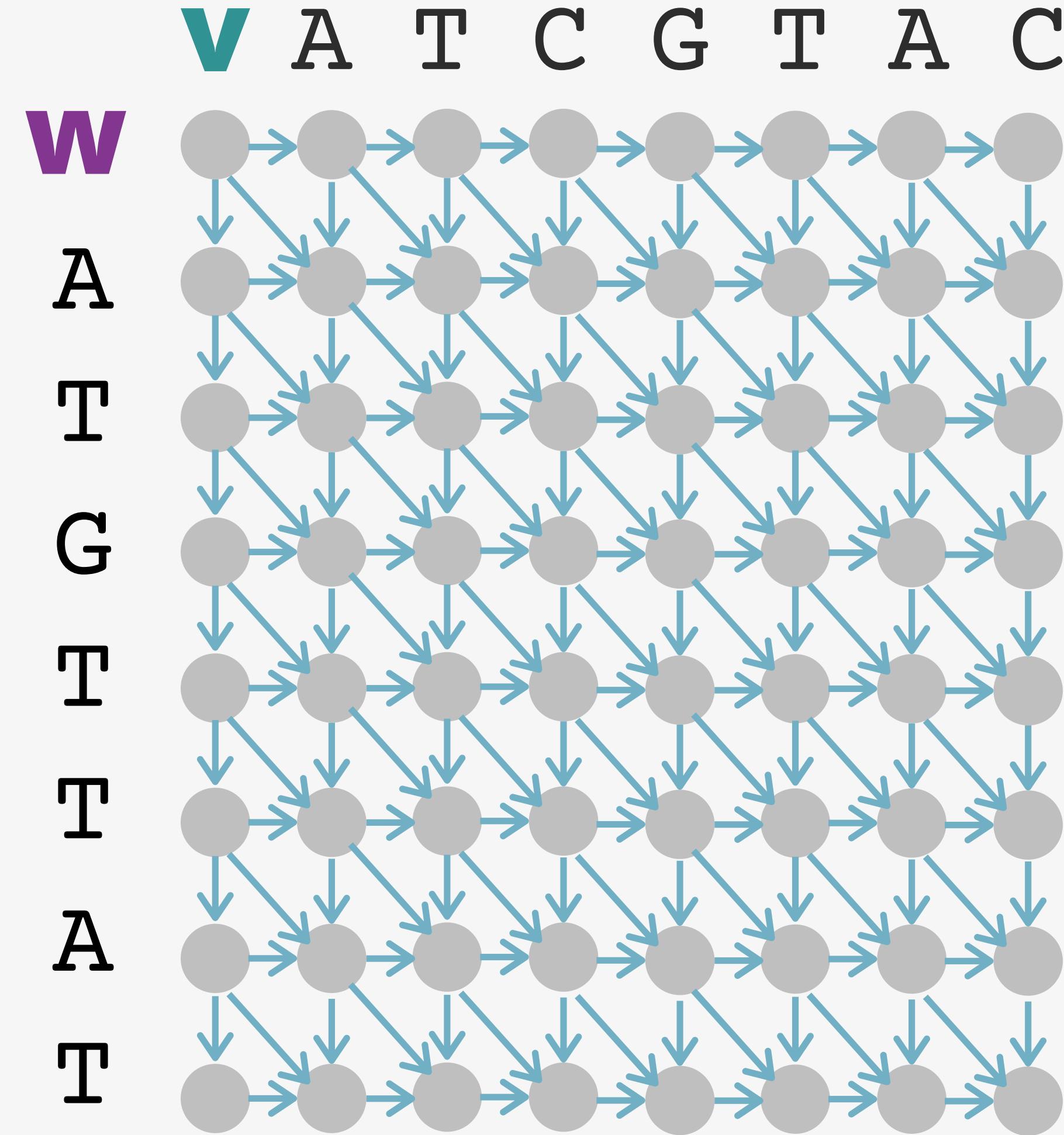
AT-GT_TA_T-

- As colunas que contêm o mesmo caracter são chamadas de *matches*
- As colunas que contêm espaços são chamadas *indels*

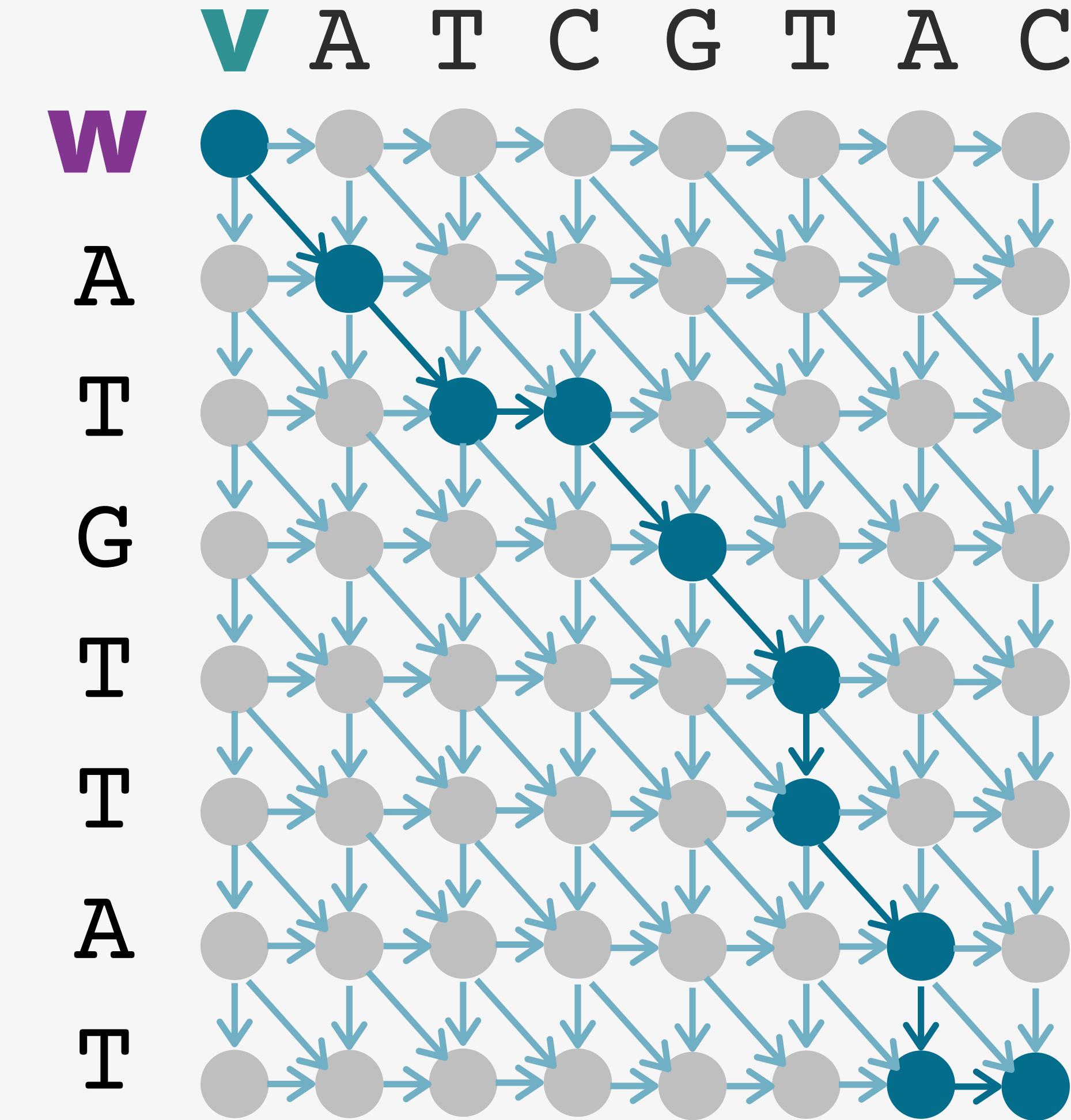
Como esse alinhamento pode ser calculado?

Esse alinhamento poderia ser representado
como um grid ou matriz na qual a
primeira sequência é representada nas colunas
e a segunda nas linhas

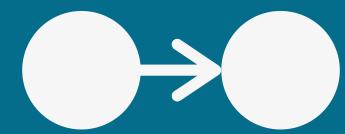
Matriz



Matriz



ATCGT-A-C
AT-GTTAT-



AT**C**GT-A-**C**

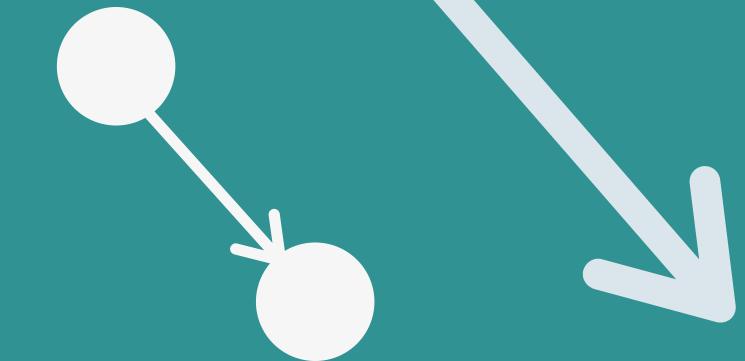
AT-**G**TAT-

Horizontais

significam **deleções** na primeira sequência **v**, ou seja, um caracter na primeira sequência é impresso e um *gap* é impresso na segunda sequência **w**

Verticais

significam **inserções** na primeira sequência **v**, ou seja, um *gap* é impresso na primeira sequência e um caracter é impresso na segunda sequência **w**



AT**C**GT-A-**C**

AT-**G**T**T**A**T**-



ATCGT-A-C
AT-GT**T**A**T**-

Diagonais

significam um **match** ou **mismatch**, ou seja, um caracter é impresso em cada uma das sequências **v** e **w** podendo ser iguais ou diferentes

Esquema de pontuação

- Falta um esquema de pontuação que seja capaz de julgar o mérito dos diversos possíveis alinhamentos
- Custo de admitir ***matches*** e ***mismatches*** bem como dos ***indels***
 - Por simplicidade, podemos pontuar “**+1**” para um ***match*** e “**0**”, caso contrário
- Não contemplamos um ***mismatch***, visto que o esquema de pontuação permite caminhamento em diagonal apenas no caso de igualdade entre os caracteres
 - Não há perda de generalidade

Esquema de pontuação

$$S_{i,j} = \max \left\{ \begin{array}{l} \bullet S_{i-1,j} \\ \bullet S_{i,j-1} \\ \bullet S_{i-1,j-1} + 1, \text{ se } v[i] = w[j] \end{array} \right\}$$

Pontuação



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0							
T	0							
G	0							
T	0							
T	0							
A	0							
T	0							

Pontuação



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑						
T	0							
G	0							
T	0							
T	0							
A	0							
T	0							

Pontuação



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←					
T	0							
G	0							
T	0							
T	0							
A	0							
T	0							

Pontuação



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↖	1←	1←	1←	1←	1↖	1←
T	0							
G	0							
T	0							
T	0							
A	0							
T	0							

Pontuação



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↖	2←	2←	2↖	2←	2←
G	0	1↑	2↑	2←	3↖	3←	3←	3←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
A	0	1↖	2↑	2←	3↑	4↑	5↖	5←
T	0	1↑	2↖	2←	3↑	4↖	5↑	5←

Reconstrução do alinhamento



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↖	2←	2←	2↖	2←	2←
G	0	1↑	2↑	2←	3↖	3←	3←	3←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
A	0	1↖	2↑	2←	3↑	4↑	5↖	5←
T	0	1↑	2↖	2←	3↑	4↖	5↑	5←

C
—

5←

Reconstrução do alinhamento



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↖	2←	2←	2↖	2←	2←
G	0	1↑	2↑	2←	3↖	3←	3←	3←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
A	0	1↖	2↑	2←	3↑	4↑	5↖	5←
T	0	1↑	2↖	2←	3↑	4↖	5↑	5↖

-C
T-

Reconstrução do alinhamento



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↖	2←	2←	2↖	2←	2←
G	0	1↑	2↑	2←	3↖	3←	3←	3←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
A	0	1↖	2↑	2←	3↑	4↑	5↖	5←
T	0	1↑	2↖	2←	3↑	4↖	5↑	5←

A-C
AT-

Reconstrução do alinhamento



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↖	2←	2←	2↖	2←	2←
G	0	1↑	2↑	2←	3↖	3←	3←	3←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
T	0	1↑	2↖	2←	3↑	4↖	4←	4←
A	0	1↖	2↑	2←	3↑	4↑	5↖	5←
T	0	1↑	2↖	2←	3↑	4↖	5↑	5←

TA-C
TAT-

Reconstrução do alinhamento



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↖	2↖	2↖	2↖	2↖	2↖
G	0	1↑	2↑	2↖	3↖	3↖	3↖	3↖
T	0	1↑	2↖	2↖	3↑	4↖	4↖	4↖
T	0	1↑	2↖	2↖	3↑	4↖	4↖	4↖
A	0	1↖	2↑	2↖	3↑	4↑	5↖	5↖
T	0	1↑	2↖	2↖	3↑	4↖	5↑	5↖

-TA-C
TTAT-

Reconstrução do alinhamento



v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0
A	0	1↖	1↖	1↖	1↖	1↖	1↖
T	0	1↑	2↖	2↖	2↖	2↖	2↖
G	0	1↑	2↑	2↖	3↖	3↖	3↖
T	0	1↑	2↖	2↖	3↑	4↖	4↖
T	0	1↑	2↖	2↖	3↑	4↖	4↖
A	0	1↖	2↑	2↖	3↑	4↑	5↖
T	0	1↑	2↖	2↖	3↑	4↖	5↑

G-TA-C
GTTAT-

Reconstrução do alinhamento



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↖	2↖	2↖	2↖	2↖	2↖
G	0	1↑	2↑	2↖	3↖	3↖	3↖	3↖
T	0	1↑	2↖	2↖	3↑	4↖	4↖	4↖
T	0	1↑	2↖	2↖	3↑	4↖	4↖	4↖
A	0	1↖	2↑	2↖	3↑	4↑	5↖	5↖
T	0	1↑	2↖	2↖	3↑	4↖	5↑	5↖

CG-TA-C
-GTTAT-

Reconstrução do alinhamento



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↑	2←	2←	2↑	2←	2←
G	0	1↑	2↑	2←	3↑	3←	3←	3←
T	0	1↑	2↑	2←	3↑	4↑	4←	4←
T	0	1↑	2↑	2←	3↑	4↑	4←	4←
A	0	1↑	2↑	2←	3↑	4↑	5↑	5←
T	0	1↑	2↑	2←	3↑	4↑	5↑	5←

TCG-TA-C
T-GTTAT-

Reconstrução do alinhamento



	v	A	T	C	G	T	A	C
w	0	0	0	0	0	0	0	0
A	0	1↑	1←	1←	1←	1←	1↑	1←
T	0	1↑	2↑	2←	2←	2↑	2←	2←
G	0	1↑	2↑	2←	3↑	3←	3←	3←
T	0	1↑	2↑	2←	3↑	4↑	4←	4←
T	0	1↑	2↑	2←	3↑	4↑	4←	4←
A	0	1↑	2↑	2←	3↑	4↑	5↑	5←
T	0	1↑	2↑	2←	3↑	4↑	5↑	5←

ATCG-TA-C
AT-GTTAT-

Esquema de pontuação

$$S_{i,j} = \max \left\{ \begin{array}{l} \bullet S_{i-1,j} \\ \bullet S_{i,j-1} \\ \bullet S_{i-1,j-1} + 1, \text{ se } v[i] = w[j] \end{array} \right\}$$

Caso exista mais de uma opção com o mesmo valor, o alinhamento é **ambíguo**, ou seja,
pode existir mais de um alinhamento de pontuação máxima

Algoritmo de Needleman-Wunsch

Algoritmo calcula o alinhamento global par-a-par ótimo

- Útil quando a similaridade entre sequências se estender por toda sua extensão
 - Exemplo: **proteínas de uma mesma** família que, normalmente, são conservadas, tem **comprimentos próximos** mesmo em organismos tão diversos quanto moscas e seres humanos

Qual a ordem de complexidade?

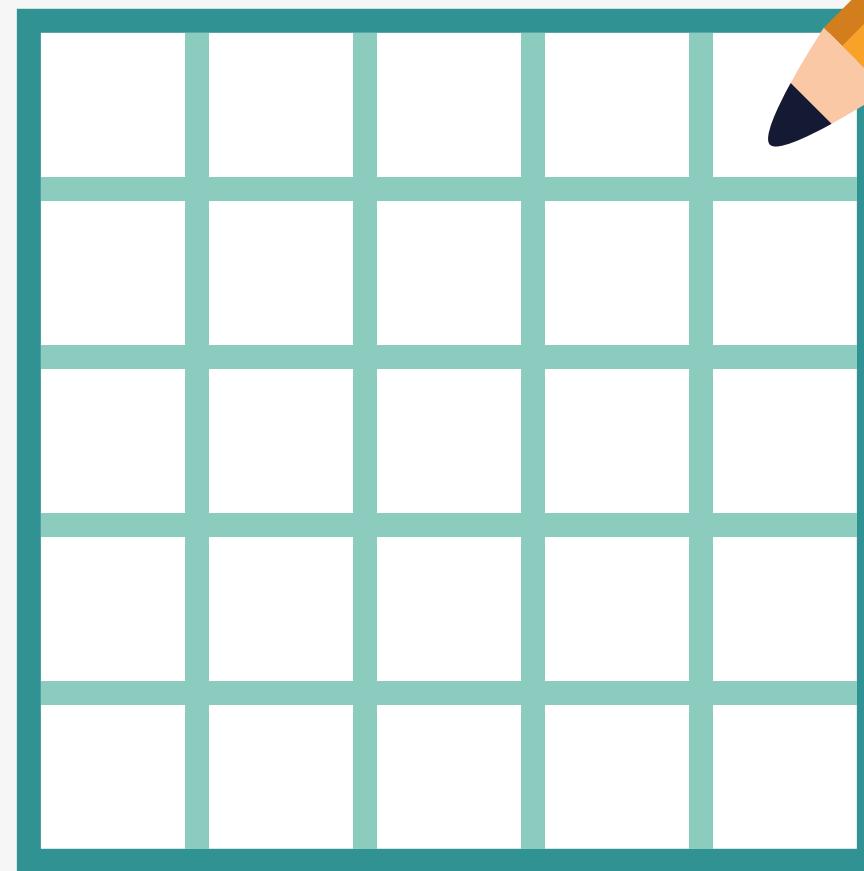
O(n^2)

- A complexidade é **quadrática** ou $O(mn)$, onde m é o comprimento de uma das sequências e n é o comprimento da outra



DESAFIO #1

Tente realizar alinhamentos globais usando o algoritmo de Needleman-Waterman manualmente



>sequencia1

DRQTAKAAAGTD

>sequencia2

ERQLAKAAAGTD



DESAFIO #2

Tente realizar alinhamentos globais usando a ferramenta do NCBI*

National Center for Biotechnology Information

Sign in to NCBI

U.S. National Library of Medicine

BLAST®

Needleman-Wunsch Global Align Nucleotide Sequences

Home Recent Results Saved Strategies

Query Sequence

Enter accession number, gi, or FASTA sequence Clear

Input limited to 100,000 letters for either input sequence. The total length of both query and subject may not exceed 150,000 letters.

Query subrange From To

New columns added to the Description Table

Click 'Select Columns' or 'Manage Columns'.

Or, upload file Escolher arquivo Nenhum arquivo selecionado

Job Title

Enter a descriptive title for your BLAST search

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence Clear

Input limited to 100,000 letters for either input sequence. The total length of both query and subject may not exceed 150,000 letters.

Subject subrange From To

Or, upload file Escolher arquivo Nenhum arquivo selecionado

Align Show results in a new window

+ Algorithm parameters

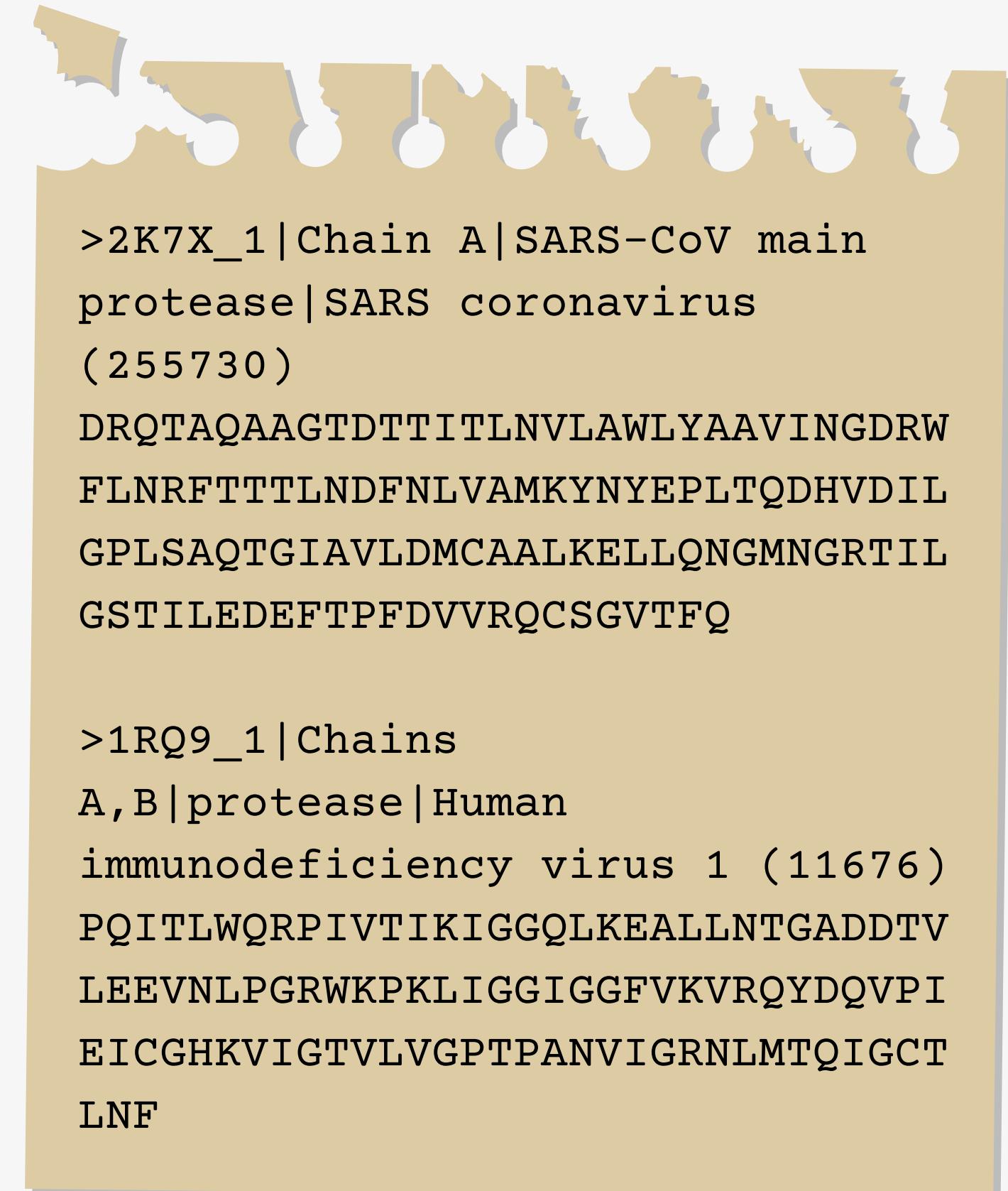
BLAST is a registered trademark of the National Library of Medicine

Support center Mailing list YouTube

NCBI National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA

Policies and Guidelines | Contact

NATIONAL LIBRARY OF MEDICINE NIH USA.gov



*National Center for Biotechnology Information



DESAFIO #3

Tente realizar alinhamentos globais usando a ferramenta do NCBI*

National Center for Biotechnology Information

Sign in to NCBI

BLAST®

Needleman-Wunsch Global Align Nucleotide Sequences

Home Recent Results Saved Strategies

Query Sequence

Enter accession number, gi, or FASTA sequence Clear
Input limited to 100,000 letters for either input sequence. The total length of both query and subject may not exceed 150,000 letters.

Query subrange From To
New columns added to the Description Table
Click 'Select Columns' or 'Manage Columns'.

Or, upload file Nenhum arquivo selecionado
Job Title
Enter a descriptive title for your BLAST search

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence Clear
Input limited to 100,000 letters for either input sequence. The total length of both query and subject may not exceed 150,000 letters.

Subject subrange From To

Or, upload file Nenhum arquivo selecionado

Align Show results in a new window

Algorithm parameters

BLAST is a registered trademark of the National Library of Medicine

Support center Mailing list YouTube

NCBI National Center for Biotechnology Information, U.S. National Library of Medicine 8600 Rockville Pike, Bethesda MD, 20894 USA

Policies and Guidelines | Contact

NIH USA.gov

>2K7X_1 | Chain A | SARS-CoV main protease | SARS coronavirus (255730)
DRQTAQAAGTDTTITLNVLAWLYAAVINGDRWFLNRFTT
TLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLD
CAALKELLQNGMNGRTILGSTILEDEFTPFDVVRQCSGV
TFQ

>6XCH_1 | Chain A | 3C-like proteinase | Severe acute respiratory syndrome coronavirus 2 (2697049)
SGFRKMAFPSGKVEGCMVQVTCGTTLNGLWLDDVVYCP
RHVICTSEDMLNPNEYEDLLIRKSNNHNFLVQAGNVQLRVI
GHSMQNCVLKLKVDTANPKTPKYKFVRIQPGQTFSVLAC
YNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDC
VSFCYMHMELPTGVHAGTDLEGNFYGPFDVRQTAQAAG
TDTTITVNVLAWLYAAVINGDRWFLNRFTTLNDFNLVA
MKYNYEPLTQDHVDILGPLSAQTGIAVLDMCASLKELLO
NGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQ

*National Center for Biotechnology Information