

Universidade Federal do Rio Grande do Sul
Instituto de Informática
Prof. Dr. Márcio Dorn
Lista 2 de Biología Computacional

Aluno: Vicente Merlo
Matrícula: 244950

O algoritmo foi implementado em Ruby, tentando maximizar o uso de programação funcional, que acolhe bem problemas de programação dinâmica. Ele segue os passos descritos pelo professor nos slides e encontrados complementarmente em:

<http://vlab.amrita.edu/?sub=3&brch=274&sim=1431&cnt=1>

Especialmente para a etapa de traceback, usei como referência a implementação em:

<https://github.com/alevchuk/pairwise-alignment-in-python>

Foram criadas as funções:

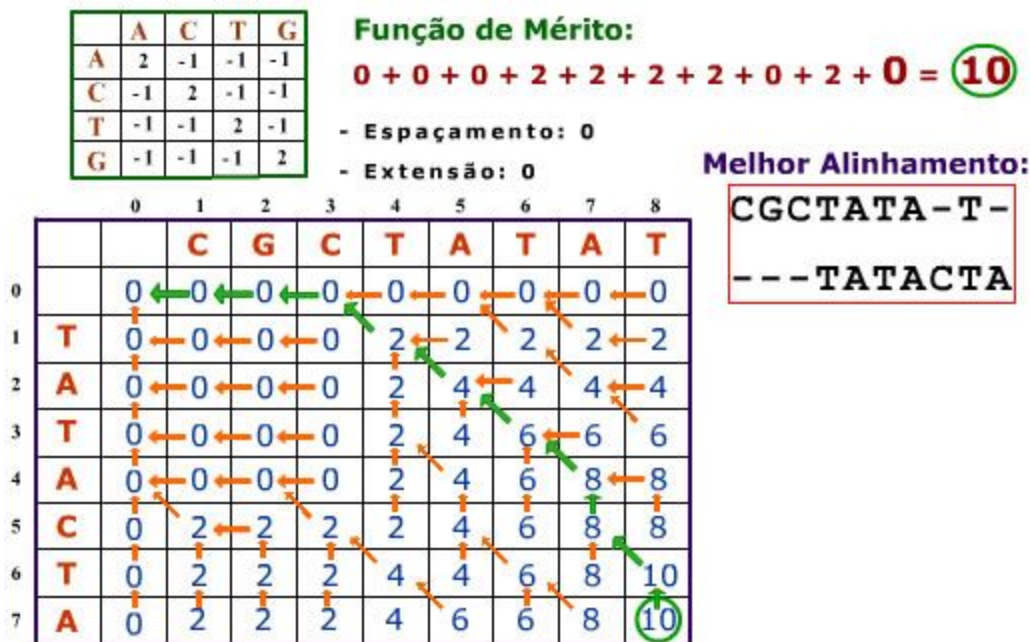
- $s(s1, s2)$: para calcular uma pontuação entre duas letras, retorna um valor em caso de *match* entre $s1$ e $s2$, e outro em caso de *mismatch*.
- $distance(s1, s2, w)$: função para verificar a distância entre duas subsequências e suas respectivas pontuações de *score* e de *identidade*.
- $traceback(..)$: função para abusar da programação funcional que ruby oferece, fazendo o traceback para obter a solução optimal da tabela de alinhamento, ao mesmo tempo que faz o alinhamento entre as duas sequências e identifica para qual linha/coluna a função deverá ser chamada no próximo passo de traceback.

Assim, seguem-se os passos:

- Inicialização de uma matriz S com $\text{len}(s1)+1$ linhas e $\text{len}(s2)+1$ colunas
- Preenchimento da primeira coluna e da primeira linha da matriz criada com suas respectivas penalidades ($w * i$)
- Preenchimento da tabela usando as regras:

$$S_{i,j} = \text{MAX} \left[\begin{array}{l} S_{i-1,j-1} + s(a_i, b_j)(\text{match/mismatch}) \\ S_{i,j-1} + w(\text{gap} - \text{seq1}) \\ S_{i-1,j} + w(\text{gap} - \text{seq2}) \end{array} \right]$$

- Chamada da função *traceback* para fazer o alinhamento usando a matriz S , começando pela borda inferior de S , da direita para a esquerda. Ou seja, começando de $S[n][m]$, onde n = número de linhas e m = número de colunas de S . Conforme a imagem, os dois alinhamentos são construídos:



- A função *traceback* retorna duas strings, que correspondem às sequências alinhadas. Então, é chamada a função *distance* nessas duas strings, que retorna dois valores: a identidade e o score entre os dois alinhamentos.

Resultados:

Human (Homo sapiens) x Horse (Equus caballus)

Cadeias alinhadas:

```
VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKV
ADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLD
KFLASVSTVLTSKY-
```

```
VLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHFDLSHGSAQVKAHGKKV
GDALTAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVHLPNDFTPAVHASLD
KFLSSVSTVLTSKYR
```

Identidade: 560

Distancia: 0.8723404255319149

Human (Homo sapiens) x Deer (Odocoileus virginianus)

Cadeias alinhadas:

```
VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKV
ADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLD
KFLASVSTVLTSKY-
```

VLSAANKSNVKAAWGKVGGNAPAYGAQALQRMFLSFPTTKTYFPHFDLSHGSAQQKAHGQK
VANALTKAQGHLNDLPGTSLNLSNLHAHKLRVNPVNFKLLSHSLLVTLASHLPTNFTP AVHANL
NKFLANDSTVLT SKYR

Identidade: 440

Distancia: 0.7659574468085106

Human (*Homo sapiens*) x Cow (*Bos taurus*)

Cadeias alinhadas:

VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKV
ADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLD
KFLASVSTVLT SKY-

VLSAADKGNVKAAWGKVGGHAAEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGAKV
AAALT KAVEHLDDLPGALSELSDLHAHKLRVDPVNFKLLSHSLLVTLASHLPSDFTPAVHASLDK
FLANVSTVLT SKYR

Identidade: 560

Distancia: 0.8723404255319149

Human (*Homo sapiens*) x Pig (*Sus scrofa*)

Cadeias alinhadas:

VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKV
ADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLD
KFLASVSTVLT SKY-

VLSAADKANVKAAWGKVGQAGAHGAEALERMFLGFPTTKTYFPHFNL SHGSDQVKAHGQK
VADALT KAVGHLLDLPGALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHHPDDFNPSVHASL
DKFLANVSTVLT SKYR

Identidade: 520

Distancia: 0.8368794326241135

Human (*Homo sapiens*) x Maned wolf (*Chrysocyon brachyurus*)

Cadeias alinhadas:

VLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKV
ADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLD
K-FLASVSTVLT SKY-

VLSPADKTNIKSTWDKIGGHAGDYGGEALDRTFQSFPPTTKTYFPHFDLSPGSAQVKAHGKKVA
DALTTAVAHLLDLPGALSALSDLHAYKLRVDPVNFKLLSHCLLVTLACHHPTEFTPAVHASLDKF
FTA-VSTVLT SKYR

Identidade: 507

Distancia: 0.823943661971831

Human (*Homo sapiens*) x Chicken (*Gallus gallus*)

Cadeias alinhadas:

```
VLSPA-DKTNVKAAWGKVGA-HAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKK  
VADALTNAVAHVDDMPNALS---A-LSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHA  
SLDKFL-ASVSTVLTSKY-  
ML-TAEDKKLIQQAWEK-AASHQEEFGAEALTRMFTTYPQTKTYFPHFDLSPGSDQVRGHGKK  
VLGALGNAVKNV-D--N-LSQAMAELSNLHAYNLRVDPVNFKLLSQCIVVLAVHMGKDYTPEVH  
AAFDKFLSA-VSAVLAEKYR
```

Identidade: 253

Distancia: 0.6013513513513513

Human (*Homo sapiens*) x Trout (*Oncorhynchus mykiss*)

Cadeias alinhadas:

```
-VLSPA-DKTNVKAAWGK-VGAHAGEYGAEALERMFL-SFPTTKTYFPH-FDLSHGSAQVKGHGK  
KVADALTNAVAHVDDMPNALSALSDDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVH-AS  
LDKFLASVSTVLTSKY-  
XSL-TAKDKSVVKAFWVGKISG-KADVVGAEALGRM-LTAYPQTKTYFSHWADLSPGSGPVKKHG  
GIIMGAIGKAVGLMDDL VGGM SALSDDLHAFKLRVDPGNFKILSHNILVT LAIHFP SDFTP EVHIA-V  
DKFLAAVSAALADKYR
```

Identidade: 220

Distancia: 0.5714285714285714

Assim, as espécies que mais se aproximam da humana são os **cavalos (*Equus caballus*)** e as **vacas (*Bos taurus*)**, tanto por distância quanto por aproximação (ambas espécies possuem os mesmos valores).