

## 序列两两比较之序列比对法：双序列局部比对

在看局部比对的算法之前，先看看序列 a 和序列 b 这两条序列，谁和序列 c 更相似一些（图 1）。AC 做全局比对的得分是 -46 分。BC 做全局比对的得分是 8 分。单从分数上看，b 跟 c 更相似，因为得分高。但是如果只看红框里这部分，a 和 c 显然比 b 和 c 对得要好。如果只比较红框里这部分的话，a 和 c 的得分是 24 分，比 b 和 c 的得分要高。这个例子告诉我们，对于像这样一长一短的两条序列，比较局部比比较全长更有意义。这就是为什么除了全局比对，还有局部比对。

全局比对	全局比对	局部比对
序列a: ASTDT <b>PYMNVI</b> PPCDEEFV	序列b: <b>ATPY-ELFFV</b>	序列a: PYMNVI
序列c: ----- <b>PYINVF</b> -----	序列c: -- <b>PYINVF</b> --	序列c: PYINVF
比对得分: -46	比对得分: 8	比对得分: 24

图 1. 序列 a 和序列 b 与序列 c 的比较

局部比对的算法和全局比对很相似，只是在选最大值时通过增加了第四个元素“0”，来达到比对局部的效果。序列 p 和序列 q，一长一短，其他输入值跟全局比对的一样（图 2）。

序列p: ACGTC		A	G	C	T	-
序列q: CG	A	10	-1	-3	-4	
	G	-1	7	-5	-3	
	C	-3	-5	9	0	
	T	-4	-3	0	8	
	-		-5			

替换记分矩阵

图 2. 局部比对输入值：序列 p 和序列 q，替换记分矩阵，空位罚分

局部比对的计算公式在全局比对的基础上增加了第四个元素“0”。得分矩阵初始值仍是 0，但第一行和第一列与全局比对不同，全是 0（图 3）。

**A**

$$s(0,0) = 0$$

$$s(0,j) = 0, 1 \leq j \leq m$$

$$s(i,0) = 0, 1 \leq i \leq n$$

$$s(i,j) = \max \begin{cases} 0 \\ s(i-1,j-1) + w(i,j) \\ s(i-1,j) + \text{gap} \\ s(i,j-1) + \text{gap} \end{cases}$$

**B**

		0	1	2	3	4	5	序列 p
			A	C	G	T	C	
0		0	0	0	0	0	0	
1	C	0						
2	G	0						

序列 q

得分矩阵

图 3. 局部比对计算公式及得分矩阵

从  $s(1,1)$  开始要选择四个值中的最大值。除了上面格  $s(0,1)+gap=0+-5=-5$ ，左边格  $s(1,0)+gap=0+-5=-5$ ，斜上格  $s(0,0)+w(1,1)=0+-3=-3$ ，还有一个 0。 $\max(-5, -5, -3, 0)=0$ 。并且这个 0 既不是从上面格，也不是从左边格，以及斜上格三个方向来的，而是来自于公式里增加的“0”，所以不用画箭头（图 4）。

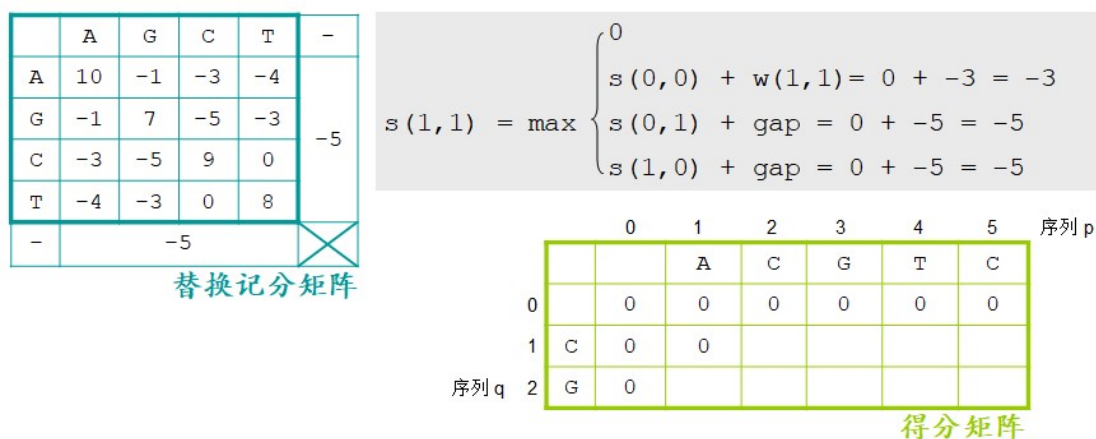


图 4. 得分矩阵中的  $s(1,1)$

$s(1,4)$  的计算：上面格  $s(0,4)+gap=0+-5=-5$ ，左边格  $s(1,3)+gap=4+-5=-1$ ，斜上格  $s(0,3)+w(1,4)=0+0=0$ ，还有一个 0。 $\max(-5, -1, 0, 0)=0$ 。这个 0 和  $s(1,1)$  的 0 是不一样的。这个 0 应该画上斜上的箭头（图 5），因为它可以来自公式中的 0，也可以来自斜上格。而  $s(1,1)$  的 0 没有箭头因为它只来自公式中的 0。两种情况虽然都是 0，但来源不同，一定要通过箭头标识清楚。

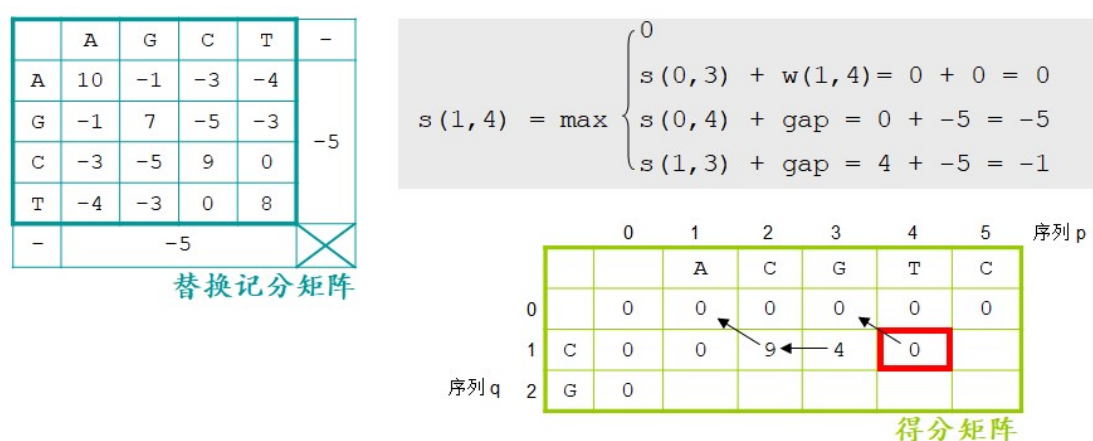


图 5. 得分矩阵中的  $s(1,4)$

按照公式，填充满整个得分矩阵（图 6）。与全局比对不同，局部比对的得分不是在右下角，而是在整个矩阵中找最大值。这个最大值才是局部比对的最终得分，他可能出现在任

何一个位置。这次箭头追溯也不是从右下角到左上角，而是从刚刚找到的最大值开始追溯到没有箭头为止。追溯箭头终止的位置也可以是得分矩阵中的任何一个位置。

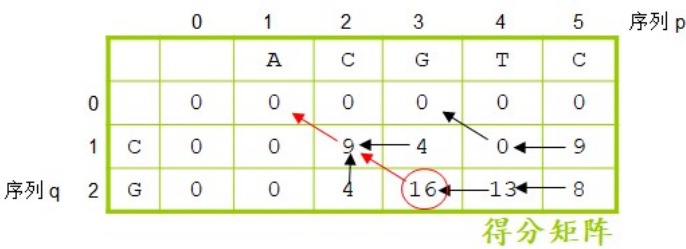


图 6. 填满分值和箭头的得分矩阵

最后根据标记好的箭头写出比对结果（图 7）。从左上到右下标记的红色箭头依次是：斜箭头字母对字母，C 对 C；斜箭头字母对字母，G 对 G。相比这两条序列的全局比对结果，两端的空位在局部比对中就全部被忽略掉了。

序列p:	C	G	序列p:	A	C	G	T	C
序列q:	C	G	序列q:	-	C	G	-	-
局部序列比对结果:	16	全局序列比对结果:	1					

图 7. 序列 p 和序列 q 的局部比对与全局比对的比较