# 1. 基因突变问题

## (dna.pas/c/cpp)

### ★问题描述:

许多疾病往往是由基因突变引起的。这种基因突变是从一个正常的DNA序列通过几代人的遗传而产生的。换言之，一个正常的DNA序列经由不同的基因突变可产生不同的遗传基因序列。生物信息学研究的一个基本问题是，对于两个给定的DNA序列，判断其中的一个DNA序列是否从另一个DNA序列经由一系列基因突变而产生的。

设 $X=x_1x_2\cdots x_n$ 是一个给定DNA序列。其中，$x_i\in\{A,C,G,T\}$，$1\le i\le n$。 对于 $1\le i\le j\le n$， $X=x_1x_2\cdots x_n$ 的从 $i$ 到 $j$ 的子串 $x_{i,j}=x_ix_{i+1}\cdots x_j$ 称为 $X$ 的 $[i,j]$ 位段。

DNA序列 $X=x_1x_2\cdots x_n$ 中常见的两种基因突变是染色体反转 $\theta$ 和染色体易位 $\tau$，定义如下：

（1）对于单元素：$\theta(A)=T$， $\theta(T)=A$； $\theta(C)=G$， $\theta(G)=C$。

（2）对于子串 $x_{i,j}=x_ix_{i+1}\cdots x_j$ 的染色体反转 $\theta_{i,j}(X)$ 定义为

$\theta_{i,j}(X)=\theta(x_j)\theta(x_{j-1})\cdots\theta(x_i)$。

（3）对于子串 $x_{i,j}=x_ix_{i+1}\cdots x_j$ 和 $1\le i<k\le j\le n$，染色体易位 $\tau_{i,j,k}(X)$ 定义为

$\tau_{i,j,k}(X)=x_{k,j}x_{i,k-1}$。

当两个基因突变的位段不相交时，称这两个基因突变不相交。设 $\Theta$ 是一个不相交基因突变组成的集合， $X=x_1x_2\cdots x_n$ 是一个给定DNA序列，将 $\Theta$ 中基因突变依次作用于 $X$ 得到的DNA序列记为 $\Theta(X)$。例如，当 $X=TAGAC$， $\Theta=\{\tau_{1,3,2},\theta_{5,5}\}$ 时， $\Theta(X)$=AGTAG。

对于两个长度相同的DNA序列 $X=x_1x_2\cdots x_n$ 和 $Y=y_1y_2\cdots y_n$，如果存在不相交基因突变组成的集合 $\Theta$，使得 $\Theta(X)=Y$，则称 $X$ 可突变为 $Y$。$\Theta$ 的最小不相交突变个数称为 $X$ 和 $Y$ 之间的突变距离，并记为 $md(X,Y)$。当 $X$ 不可突变为 $Y$ 时， $md(X,Y)=\infty$。

例如，当 $X=TAGAC$， $Y=TAACG$ 时， $\Theta_1=\{\theta_{1,2},\tau_{3,5,4}\}$ 和 $\Theta_2=\{\tau_{3,5,4}\}$ 是仅有的两个不相交基因突变集合使得： $\Theta_1(X)=Y$， $\Theta_2(X)=Y$，因此， $md(X,Y)=|\theta_2|=1$。

基因突变问题是要求对于给定的两个长度相同的DNA序列 $X=x_1x_2\cdots x_n$ 和 $Y=y_1y_2\cdots y_n$，计算其突变距离 $md(X,Y)$。

### ★编程任务:

对于给定的两个长度相同的DNA序列 $X=x_1x_2\cdots x_n$ 和 $Y=y_1y_2\cdots y_n$，计算其突变距离 $md(X,Y)$。

### ★数据输入:

输入文件名为dna.in。

文件的第一行有1个正整数n，(1<=n<=23000)，表示DNA序列的长度。接下来的2行分别给出输入序列 $X=x_1x_2\cdots x_n$ 和 $Y=y_1y_2\cdots y_n$。其中序列中每个元素均为 $\{A,C,G,T\}$ 中字母。

### ★结果输出:

输出文件名为dna.out。

将计算出的突变距离 $md(X,Y)$ 输出到文件中。当 $md(X,Y)=\infty$ 时，输出-1。

| 输入示例 | 输出示例 |
| --- | --- |
| 5<br>TAGAC<br>TAACG | 1 |