

# 一种 DNA 的距离定义及其应用

张 春 霆

(天津大学物理系)

在 Watson-Crick 的双螺旋模型中,螺旋参数是不变的。例如对 B-DNA,螺旋扭角总是  $36^\circ$ 。近年来,对 DNA 寡聚体单晶的 X 光分析表明,螺旋参数将随序列的不同而变化。对于 B-DNA,螺旋扭角可在  $26^\circ-42^\circ$  之间变化。Dickerson<sup>[1]</sup> 总结出一套规则,可由碱基序列预测螺旋扭角的变化。结果与实验符合甚好。

我们认为,两个 DNA 序列,其差别首先表现它们的一级结构,即碱基顺序的差异上。为了给这种差异以定量的描述,我们将通过螺旋扭角序列来定义两个 DNA 片断之间的“距离”。因为根据 Dickerson 规则<sup>[1]</sup>,碱基序列唯一地确定了扭角序列。设两 DNA 序列之长度各为  $N_1$  及  $N_2$ (bp),则通过 Dickerson 规则可唯一地求出它们各自的螺旋扭角序列:  $x_1, x_2, \dots, x_{N_1-1}$  及  $y_1, y_2, \dots, y_{N_2-1}$ 。则我们定义这两个 DNA 片断之距离  $D_{xy}$  为

$$D_{xy} = \left[ \sum_{n=1}^{M-1} (x_n - y_n)^2 \right]^{1/2}, \quad (1)$$

其中  $M$  为  $N_1$  与  $N_2$  的最小公倍数。为了便于比较,将(1)式除以某些常数,并经过简单变换,得到下列约化距离  $d_{xy}$

$$d_{xy} = 100 \times \sqrt{1 - \rho_{xy}}, \quad (2)$$

其中  $\rho_{xy}$  为序列  $\{x_n\}$  及  $\{y_n\}$  之间的相关系数,定义为

$$\rho_{xy} = \frac{\sum_{n=1}^{M-1} x_n y_n}{\left[ \sum_{n=1}^{M-1} x_n^2 \cdot \sum_{n=1}^{M-1} y_n^2 \right]^{1/2}}. \quad (3)$$

公式(2)定义的约化距离具有下列性质: 1.  $0 \leq d_{xy} \leq 100$ ; 2.  $d_{xy} = 0$ , 仅当  $x_n = y_n$ ; 3.  $d_{xy} = 100$ , 仅当  $\rho_{xy} = 0$ ; 4.  $d_{xy} = d_{yx}$ ; 5.  $d_{xz} + d_{yz} \geq d_{xy}$

五种革兰氏阴性菌的起始复制区的 DNA 序列由文[2]给出。它们的长度均为 245bp。这

五种细菌是: 1. 大肠杆菌; 2. 鼠伤寒沙门氏菌; 3. 产气肠杆菌; 4. 肺炎杆菌; 5. 胡萝卜软腐欧文氏菌。这五种 DNA 序列之间的距离排成矩阵  $A$ , 其中  $A_{12}$  表示 1、2 两种细菌对应 DNA 序列之间的距离,其它依此类推。五种人猿科动物:

1. 人类; 2. 黑猩猩; 3. 大猩猩; 4. 猩猩; 5. 长臂猿, 它们体内都有一种蛋白质——烟酰胺腺嘌呤二核苷酸脱氢酶 V。编码该蛋白质的主要 79 个密码子由文[3]给出。它们之间的距离见  $B$  矩阵。

$$A = \begin{bmatrix} 0 & 18.8 & 20.2 & 22.5 & 27.4 \\ 18.8 & 0 & 14.2 & 19.5 & 23.2 \\ 20.2 & 14.2 & 0 & 8.2 & 15.9 \\ 22.5 & 19.5 & 8.2 & 0 & 13.7 \\ 27.4 & 23.2 & 15.9 & 13.7 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 9.7 & 13.3 & 19.1 & 19.3 \\ 9.7 & 0 & 4.3 & 13.0 & 15.3 \\ 13.3 & 4.3 & 0 & 12.0 & 15.3 \\ 19.1 & 13.0 & 12.0 & 0 & 9.9 \\ 19.3 & 15.3 & 15.3 & 9.9 & 0 \end{bmatrix}$$

矩阵  $A$  一般是符合细菌分类学的。关于人类与黑猩猩的关系更近,还是与大猩猩的关系更近这一问题,在生物学界历来是有争议的。矩阵  $B$  中的数据支持前一种说法。当然,用一种生物的局部 DNA 片断来对整个生物作出论断,一般是不合理的。但是对相近生物的相对应 DNA 序列作比较,还是有意义的。由于本工作已全部计算机化,因此对于迅速增长的已知 DNA 序列,可望进行快速分析。

## 参 考 文 献

[1] Dickerson, R. E.: *J. Mol. Biol.*, 1983, **166**, 419.  
 [2] 王德宝等:《核酸》,上册,科学出版社,北京,1986, 111—112页。  
 [3] Wilson, A. C.: *Scientific American*, 1985, **253**, 148.

[本文于 1988 年 10 月 31 日收到]