

## 壹、研究動機與目的

1971 年，美國國家癌症研究所啟動了「向癌症開戰」計畫；在計畫中，科學家們認識到包括癌症在內的許多疾病與基因之間存在有緊密的聯繫性。因此，解開基因的秘密是征服這些疾病的先決條件；作為基因研究的基礎，基因中的鹼基序列（即 DNA 序列）測定也開始受到廣泛的重視。在國際人類基因組計劃啟動後，1999 ~ 2006 年間便完成了全部 23 條染色體的定序工作。

在生物的 DNA 序列中存在有簡單重複序列（Simple Sequence Repeats, SSRs），通常指的是長度約二到六個鹼基對的單元，以重複相鄰的形式存在著。目前已經知道簡單重複序列與部分癌症有所關聯（Eshleman & Markowitz, 1996; Gurin, Federici, Kang, & Boyd, 1999; Han, Yanagisawa, Kato, Park, & Nakamura, 1993; Lazzereschi et al., 1999; Orlow, Lianes, Lacombe, Dalbagni, Reuter, & Cordon-Cardo, 1994; Rhyu, Park, & Meltzer, 1994），同時也是法醫學領域個體識別和親權鑑定的重要分析標的（Deka, Shriver, Yu, Ferrell, & Chakraborty, 1995; Dib et al., 1996）。因此，研究者所將要探討的是的四種以二聚體為單元的簡單重複序列—— $(AC)_n$ 、 $(AG)_n$ 、 $(AT)_n$  及  $(CG)_n$ 。在此，將以各種 SSRs 的重複長度  $n$  與出現次數  $f(n)$  的分布關係作為分析資料，以便往後能再進行更深入的生物物理探討與其他醫學研究。而本研究中，將採用 ORIGIN 來對人體第 22 條染色體中 SSRs 的分布曲線進行擬合。

## 貳、對 SSRs 分布曲線的擬合

目前在 SSRs 分布曲線的研究上所提出的分布函數形式有二：

一、指數函數（Durrett & Kruglyak, 1999; Kruglyak, Durrett, Schug, & Aquadro, 1998; Ohta & Kimura, 1973）：

$$f = Ae^{-n} \quad (\text{公式 1})$$

二、冪函數（Buldyrev et al., 1998; Dokholyan, Buldyrev, Havlin, & Stanley, 1997）：

$$f = An^{-B} \quad (\text{公式 2})$$

雖然以冪函數來擬合 SSRs 分布曲線已經有很高的相似度，但是對於  $15 < n < 25$  範圍的分布曲線就不是那麼地吻合，所以將進一步以冪函數疊加上高斯分

布函數：

$$f = An^{-B} + C \exp[-D(n - n_0)^2] \quad (\text{公式 3})$$

來擬合 SSRs 的分布曲線。

首先，將人體第 22 條染色體中 SSRs 的分布資料輸入 ORIGIN 程式中（見圖 1），再點取選項 Analysis 中的選項 Nonlinear Curve Fit（見圖 2），調整參數

	ACn(X1)	AC(Y1)	AGn(X2)	AG(Y2)	ATn(X3)	AT(Y3)	CGn(X4)	CG(Y4)
Long								
1	3	10185	3	14288	3	11229	3	862
2	4	1864	4	1777	4	2334	4	92
3	5	575	5	341	5	541	5	18
4	6	230	6	160	6	238	6	5
5	7	138	7	87	7	123	7	2
6	8	70	8	47	8	60	8	1
7	9	60	9	27	9	47	9	1
8	10	34	10	13	10	34	10	1
9	11	43	11	16	11	22		
10	12	32	12	9	12	16		
11	13	24	13	1	13	13		
12	14	39	14	5	14	7		
13	15	26	15	5	15	10		
14	16	26	16	1	16	9		
15	17	24	17	4	17	8		
16	18	24	19	1	18	5		
17	19	27	20	1	19	8		
18	20	15	21	2	20	7		
19	21	24	24	1	21	6		
20	22	18	25	1	22	6		
21	23	13	26	1	23	5		
22	24	8	28	1	24	5		
23	25	13	33	1	25	4		
24	26	1			26	2		
25	27	2			27	3		
26	28	5			28	1		
27	30	1						
28	32	1						
29	37	1						

圖 1 輸入人類染色體第 22 條染色體中 SSRs 的分布資料

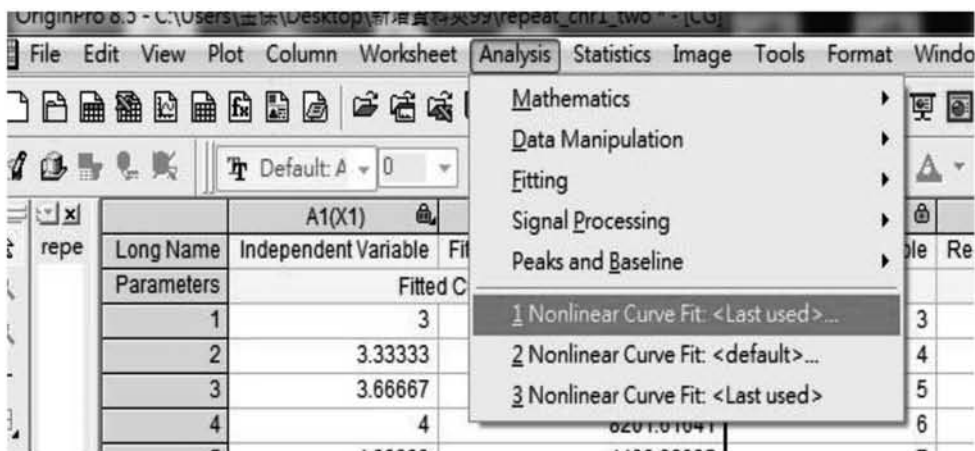


圖 2 點取選項 Analysis 中的選項 Nonlinear Curve Fit