B型肝炎病毒突變率之計算與公式推導

Derive the Formula of Mutation Rae of HBV

孫光天	魏至軒	張定宗	劉紋君	徐禕佑
Koun-Tem Sun	Jr-Shiuan Wei	Ting-Tsung Chang	Wen-Chun Liu	Yi-Yu Hsu
國立臺南大學 數位學習科技系	國立臺南大學 數位學習科技系	國立成功大學 醫學院基礎醫學所	國立成功大學 醫學院基礎醫學所	國立臺南大學 數位學習科技系
ktsun@mail.nutn.edu.tw	*rick70002@yahoo.com.tw*	*ttchang@mail.ncku.edu.tw*	*ttchang@mail.ncku.edu.tw*	*ktsun@mail.nutn.edu.tw*

摘要

本研究主旨在於建立B型肝炎病毒序列的長期突變率計算公式，我們利用短時間序列樣本的特性和行為，用來預測一段時間間隔後的演化結果。在本研究中，我們使用4個月的序列資料，並修改物種距離計算的方法(JC法)，為突變率預測公式，推估出1年後的突變率，我們將推估的結果與實際的突變率運用成對t檢定，檢定兩者有無差異，結果發現推估的結果與實際的突變率無顯著差異(P值達0.32)，證明本研究所建立的B型肝炎短期突變率計算公式的可靠性。最後，我們將其公式應用在B型肝炎病毒序列長期的演化，由短期(一年)的序列資料推估長期(四十年)的突變率，結果發現與其它學者研究文獻的真實四十年的演化數據有相似成果。經實驗結果證明，本方法有極高正確性與彈性，可準確預測B型肝炎病毒之突變率(短期與長期)，對B型肝炎病毒演化與治療將成為一有用工具。

關鍵字：演化率、演化公式、突變率、B型肝炎病毒

Abstract

The main of this research is to derive the formula of mutation rate of the hepatitis B virus. We use the characteristic and competent of short time sampling sequences to derive the formula for long term evolution. We make use of the formula to predict the evolution result after a period of time later. In this research, we modify the JC method to be a formula for estimating the evolutionary rate after 1 year by the sampling data within four months. The paired- t test is used to evaluate the hypothesis, and that is significant between the results of own prediction and the real evolutionary rate (the P-value is 0.32). The results verify the feasibility of the own method.Then, we apply own formula to calculate the mutation rate of the hepatitis B virus. The long-term (40 years) mutation rate is computed by our method, and the results are similar to related researches. Own formula would be a useful technology to HBV sequence phylogenetic analysis of bioinformatics.

Keywords: evolutionary rate、evolutionary equation、mutation rate、hepatitis B virus

1、前言

本研究方向是要導出一個B型肝炎序列長期(40年)與短期(1年)突變率計算公式，由於過去對於B型肝炎突變率研究的領域，最早是由Okamoto等學者[5]提出，Okamoto等學者[5]對一個54歲的B型肝炎女性患者體內的HBV進行40年的突變分析，於第1年與40年後取得的病毒序列進行研究，他假設病毒來自出生時感染的同一母源毒株，則在其體內突變率為平均每年每個位點會產生$1.4\times10^{-5} \sim 3.2\times10^{-5}$次突變(Mutations)，但由本研究團隊與成大醫學院張定宗醫師過去的研究，利用間隔一個月抽一次血，長達一年的方式計算月突變率和年突變率，結果顯示平均每年每個位點會產生$3\times10^{-3} \sim 7\times10^{-3}$次突變(Mutations)，與Okamoto等學者的研究結果相差甚遠，張定宗醫師認為短期所計算的

突變之所以會和長期所計算的突變相差這麼多，是因為基因序列中有些點會在短時間不斷的突變，但有些點卻因為高保留性(conserved)的關係，長達40年，還是不會改變，本研究將針對這個問題提出一個修正後的公式。

在The phylogenetic handbook[6]書中提及突變分為三類，為高保留性的固定突變(fixed mutation)、多態性突變(polymorphism)、看不見的突變(lost mutation)，如圖1所示，固定突變是一種長期(long-term)的演化結果，在計算物種演化距離時，只計算這一種突變，其突變的比率，我們種之為突變率(mutation rate)或演化率(evolutionary rate)，多態性突變是指說其位點有二種或二種以上的核甘酸循環變化，換句話說，就是該位點會在短時間內一直重複，當長期觀察時，則看不出其突變的速度，看不見的突變指其位點極其重要，當突變發生時會導致物種的死亡，本研究將針對固定突變、多態性突變的差異，提出一個修正後的公式，以解決資料上計算之誤差。

由圖 1中的P是指突變率(與第一個時間點比)會隨著時間的延伸，漸漸的接近某個上限機率在Jukes與Cantor[2]兩位學者所提出的JC演化距離計算公式中就把這個上限機率定為0.75，因為在他們所提出的方法，認為核苷酸要轉變成任一個自身以外的核苷酸，都只有0.25的機率，$0.25\times 3=0.75$，所以JC法的P值上限為0.75，而K值則是指推估出來的平均每個位置的突變次數，且K值沒有上限，在計算演化樹時，會選用K值做為物種間的距離。

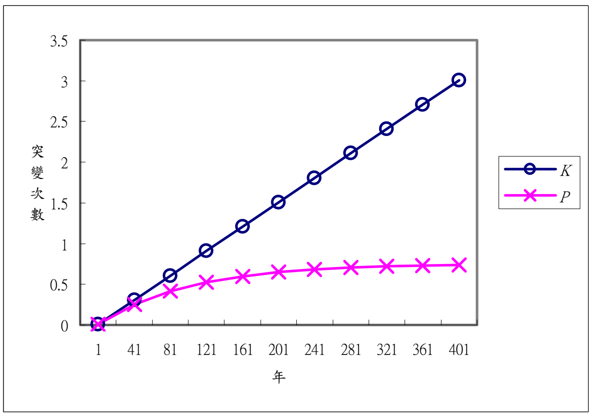

圖 1 不同時間距離, JC法(P值)與突變次數(K值)之變化

在本研究中，我們將利用短期一年內的變化量，將其當作K線的單位時間變化量，去推估40年後的突變率，並將其40年間的平均單位時間突變率求出來。

2、方法

2.1、文獻探討

在這個部份是要找一個B型肝炎序列長期(40年)與短期(1年)突變率差異的修正公式，Okamoto等學者[5]於1987 年對一個54歲的B型肝炎女性患者體內的HBV進行40年的突變分析，於第1年與40年後取得的病毒序列進行研究，他假設病毒來自出生時感染的同一母源毒株，則在其體內突變率為平均每年每個位點會產生$1.4\times10^{-5} \sim 3.2\times10^{-5}$個突變(Mutations)。其突變率公式是採用JC法[2]和其平均年突變率(Mutation rate)計算如下：

$$Mutation\_rate = -\frac{3}{4}\ln(1-\frac{4}{3}D) \quad (1)$$

D ：兩物種序列之差異個數(突變個數)÷序列全長。

$$Mutation\_rate\_per\_year = \frac{Mutation\_rate}{2t} \quad (2)$$

其中t為年，除以2倍的年是由於突變率(Mutation rate)[1]最早的定義是指兩個物種和共同祖先的距離(突變次數)，如圖 2所示，所以除以2倍的時間，才能計算出真正單年的突變率真實的情況。

圖 2兩個物種共同演化

圖 3單一物種獨自演化

但真實情形並非如此，如圖 3所示，由於Okamoto等學者[5]是拿第1年和40年後的病毒序列去作分析，且做了病毒來自出生時感染同一母源毒株的假設，所以兩者是祖先(ancestor)與後代(posterity)的關係，所以在考慮時間的部份，應該只除以單倍的時間。

由於B型肝炎序的transitional與transversional[3]的差異不大，所以我們選用JC法，作為公式修改的依據，由JC法的公式我們可以看出來，原本公式是要用P去求K，推估真實的突變次數，但在這個部份，我們是要反過來，用K去求P，因為我們要推估的不再是突變次數，而是和第1年比，40年後的突變率，所以我們要用現在成大張定宗醫師所提供的5位病患的1年內12次，以月為間隔的抽血毒病序列資料作為資料來源。

2.2、修正公式推導三個步驟

2.2.1、將平均年突變率(mutation rate)計算公式中的2t改成*t*(時間)

在此原本在計算 P_t 轉 P_{t+1} 的改變量時，要乘上2，但因為這裡是單一物種在做演化，所以要將2拿掉：

$$
\begin{aligned}
&A \leftrightarrow A \to T : (1-3P_t)\times\alpha \\
&A \leftrightarrow T \to T : P_t \times(-3\alpha) \\
&A \leftrightarrow C \to T : P_t \quad \times\alpha \\
&A \leftrightarrow G \to T : P_t \quad \times\alpha
\end{aligned}
$$

α ：每個位置平均突變成任一核苷酸的平均每年的突變次數。

所以突變率公式改為：

$$K = 3\alpha \times t = -\frac{3}{4}\ln(1-\frac{4}{3}D) \quad (3)$$

，但因為我們是要用K去求P推做40年後的突變率所以改成：

$$D_1 = \frac{3}{4} - \frac{3}{4}e^{-\frac{4}{3}\mu\times t} \quad (4)$$

在此我們令 $\mu = 3\alpha$ ，以利於計算。

$\mu = 3\alpha$ ：每個位置平均每年的突變次數，又稱平均時間突變率。

2.2.2、將多態性突變 (polymorphism mutation)從突變個數中去除，並另外計算

在物種的演化，序列中的某些位點或片段，由於其有重要的功能或意義，以致於若發生突變，該物種就會死亡，所以該物種的後代的這些位點或片段不容易發現突變，這個特性，我們稱它叫高度保留性(conserved)，於B型肝炎序列中，有些位點的保留性比其他位置低，所以我們希望透過保留性低的位點，來計算突變率，避免保留性高的位點或片段影響到突變率的計算。

多態性突變(polymorphism)[6]，是指說其位點有二種或二種以上的核甘酸循環變化，換句話說，就是該位點會在短時間內一直重複，當長期觀察時，則看不出其突變的速度，所以我們希望能將這個保留性低的部份，從JC法中去除，並額外計算，然而亦呼應了之前所提及的「不考慮反覆突變的情形」下計算。

我們運用GeneDoc來呈現序列的突變情形，如圖 4，在此我們將同一位點出現最多次的相同核苷酸去除，並以"．"表示，透過圖 4的觀察，發現有些位點在一年內的抽血資料中突變2次以上，如第66個位點就突變了兩次，我們將這些點視為多態性突變 (polymorphism mutation)，我們在此稱之為離異值(Outlier)，將其從突變個數中去除，突變率公式修正為：

$$D_2 = \frac{Outlier}{N} + \frac{3}{4} - \frac{3}{4}e^{-\frac{4}{3}\mu_{without\_outlier}\times t} \quad (5)$$

$\mu_{without\_outlier}$ ：指去除*Outlier*的每個位置平均突變突變個數。

N：序列全長(3215)

Outlier：指多態性突變 (polymorphism mutation)的個數。

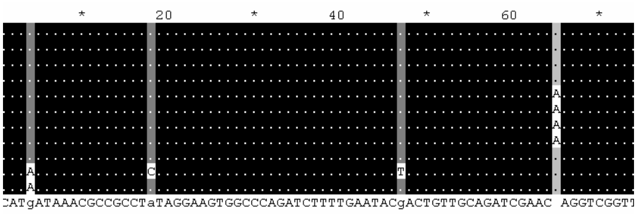

圖 4 GeneDoc去除相同核苷酸圖

2.2.3、修正突變率計算範圍

於上一步驟中有提到，各物種的序列中都存在著不同的高、低保留性的位點與片段[4]，為了避免高保留性位點與片段響影到突變率的計算，所以我們將修改突變率公式的上限。

JC法的公式之中：

$$K = 3\alpha \times t = -\frac{3}{4}\ln(1-\frac{4}{3}D) \quad (6)$$

，Nature-log內的 $\frac{4}{3}$ 是JC法的突變率上限，而P最大不能超過 $\frac{3}{4}$ ，但由於B型肝炎病毒序列中，大部份的位點都是高保留性(conserved)，所以我們將突變率上限作向下修正的動作，將上限乘上 $\frac{N_{parsimony}}{N}$ （$N_{parsimony}$ 指會產生改變的位點數)作為修正後的突變率上限，突變率公式修正為：

$$D_3 = \frac{Outlier}{N} + \frac{N_{parsimony}}{N} \times \left(\frac{3}{4} - \frac{3}{4} e^{-\frac{4}{3}\mu_{without\_outlier} \times t} \right) \quad (7)$$

3、臨床實驗

研究樣本對象由成功大學分子醫學所張定宗教授實驗室所提供，其樣本特徵分述如下。

1、病患數目：5位

2、觀察期間為連續12個月之血液參數及病毒序列，

3、其中4位病患有第18個月抽血參數及基因序列。

4、序列長度為3215。

實驗分成兩個部份：

1、用4個月的序列資料，推估1年後的突變率

2、用1年內的序列資料，推估40年後的突變率

第一部份：

我們用4個月的序列資料透過修正後的公式(D_3)計算1年後的突變。

表 1用4個月的序列資料，推估1年後的突變率

	病患ID				
方法	221	224	241	2996	4035
我們方法	5.06E-03	1.69E-02	3.01E-03	1.58E-02	9.70E-03
真實	7.62E-03	5.83E-03	3.73E-03	1.04E-02	9.02E-03
JC法	2.44E-02	2.60E-02	1.62E-02	4.28E-02	3.74E-02

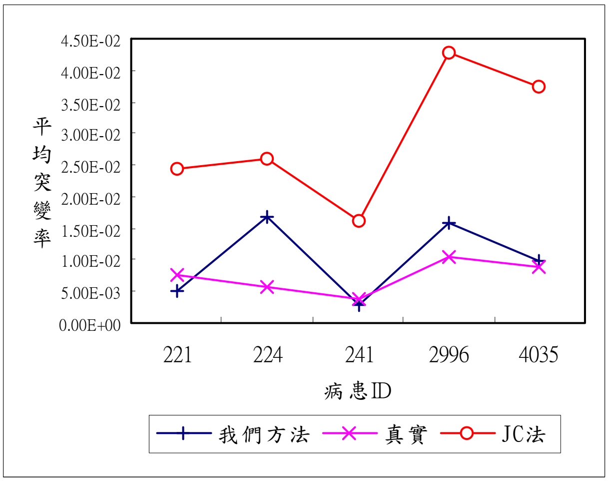

圖 5 以4個月推估1年後的突變率比較圖

從上述的表1 與圖 5中，發現我們方法與真實之突變率相去不遠，我們進一步將5位病患之資料作成對母體平均數差異檢定，發現P值在雙尾時高達0.322，無顯著差異；接著我們用JC法計算之結果與真實之突變率 ，做成對母體平均數差異檢定，發現P值在雙尾時為0.003，達到非常顯著之差異，證明了我們方法比JC法，更適用於推估B型肝炎病毒株之突變，接下來的部份，我們用不同的三個公式去計算推估後40年的突變率。

第二部份：

在這個部份，我們用我們方法與JC法公式去計算推估後40年的突變率(我們方法：D_3與JC法：D_1)。

$$D_1 = \frac{3}{4} - \frac{3}{4} e^{-\frac{4}{3}\mu \times t}$$

1、JC法公式：

這個公式是運用JC法符合指數分配的方式去計算。

2、我們方法：

$$D_3 = \frac{Outlier}{N} + \frac{N_{parsimony}}{N} \times \left(\frac{3}{4} - \frac{3}{4} e^{-\frac{4}{3}\mu_{without\_outlier} \times t} \right)$$

我們方法將一年內突變超過2次的，改為離異值(Outlier)，並修正突變率上限。

對(1)與(2)所計算出來的40年突變率求平均年突變率，為求公平起見，所以同文獻的作法，除以兩倍的時間。

$$K = \frac{-\frac{3}{4}\ln(1 - \frac{4}{3}D)}{2t} \qquad (8)$$

表 2用1年內的序列資料，推估40年後的突變率

	病患ID				
方法	221	224	241	2996	4035
JC法	2.644E-03	4.821E-03	1.089E-03	3.733E-03	2.644E-03
我們方法	6.633E-05	1.830E-04	2.726E-05	1.248E-04	9.725E-05

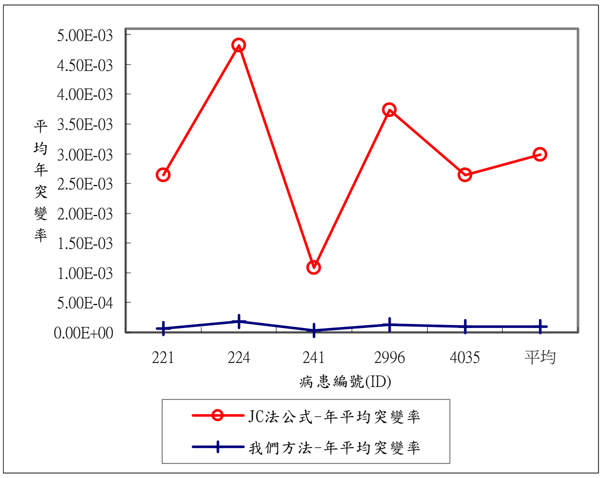

圖 6 JC法與修正後公式之年平均突變率比較圖

從上述的表2 與圖 6中，我們得知經過了修正後的公式，雖未完全推得文獻[5]所提及的$1.4\times10^{-5} \sim 3.2\times10^{-5}$的平均年突變率，但平均年突變率為$2.7\times10^{-5} \sim 1.8\times10^{-4}$了，比直接用JC法求出的平均年突變率$1\times10^{-3} \sim 4.8\times10^{-3}$，更接近其文獻學者早期研究之結果[5]。

4、結果與討論

本研究中，我們修改JC法並應用在B型肝炎病毒序列的突變率計算上，讓其計算可由短期(一年)的演化可推估長期(四十年)的演化，結果發現修正後的JC法所計算出來的結果與其它學者研究文獻的真實四十年的演化數據有相似成果，我們方法計算出來的平均年突變率為9.97×10^{-5}，比JC法所計算出來的平均年突變率2.99×10^{-3}更為接近學者Okamoto[5]研究中的平均年突變率$1.4\times10^{-5} \sim 3.2\times10^{-5}$，顯示出我們的公式不僅能夠提供短期演化之計算，也可以進行長期演化數據預測，進而評估臨床數據的差異，並判斷不同病情(如：慢性肝炎與猛爆性肝炎)病患之間的突變率有否不同，因此，本研究提供此病毒突變率公式可做為臨床醫學在病情判斷上一有用之技術。

參考文獻

[1] Y.-X. Fu, "Estimating Mutation Rate and Generation Time from Longitudinal Samples of DNA Seqeunces," *Molecular Biology Evolution*, vol. 18, pp. 620-626, 2001.

[2] T. H. Jukes and C. R. Cantor, "Evolution of protein molecules," *Mammalian Protein Metabolism*, vol. 3, pp. 21–132, 1969.

[3] M. Kimura, "Estimation of evolutionary distances between homologous nucleotide sequences," *Proc. Natl. Acad. Sci.*, vol. 78, pp. 454-458, 1981.

[4] W. Li, C. Wu, and C. Luo, "A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes," *Mol. Biol. Evol*, vol. 2, pp. 150-174,

1985.

[5] H. Okamoto, M. Imai, M. Kametani, T. Nakamura, and M. Mayumi, "Genomic heterogeneity of hepatitis B virus in a 54-year-old woman who contracted the infection through materno-fetal transmission," *Jpn J Exp Med*, vol. 57, pp. 231-6, 1987.

[6] M. Salemi and V. A., *The phylogenetic handbook.*, 200