

水稻抗白葉枯病基因同義密碼子使用偏性之影響因素探討¹

魏夢麗² 鍾依涵² 呂椿棠² 呂秀英^{3,4}

摘 要

魏夢麗、鍾依涵、呂椿棠、呂秀英。2010。水稻抗白葉枯病基因同義密碼子使用偏性之影響因素探討。台灣農業研究 59:86-102。

同樣對白葉枯病菌產生抗性的不同 *Xa* 基因間，呈現出密碼子使用偏性的多樣性。因此，本研究進行 *Xa* 基因同義密碼子使用偏性之影響因素探討，以進而瞭解形成密碼子使用的演化機制。研究材料係取自 NCBI (National Center for Biotechnology Information) 之 GenBank 資料庫下載已完全定序的 *Xa* 基因 (*Xa1*, *xa5*, *xa13*, *Xa13*, *Xa21*, *Xa26* 及 *Xa27*) 及其家族之 17 個編碼區序列 (coding domain sequence, CDS)，透過計算這些 *Xa* 基因的 CDS 長度及各種同義密碼子使用偏性指標，如同義密碼子相對使用度 (relative synonymous codon usage, RSCU)、密碼子適應指標 (codon adaption index, CAI)、有效密碼子數 (effective number of codons, ENC)、同義密碼子第三位置出現 G 或 C 核苷酸之頻率 (GC3) 以及同義密碼子第一、二位置出現 G 或 C 核苷酸之平均頻率 (GC12)，並分析其間之關係，以闡明同義密碼子使用偏性之主要影響因素。結果發現 *Xa* 基因之 CDS 長度及基因表達程度與同義密碼子使用偏性有極顯著關係，即 *Xa* 基因長度較短者可有較高的基因表達程度、以及較高的同義密碼子使用偏性，如 *xa5*、*xa13*、*Xa13* 及 *Xa27* 基因。*Xa* 基因序列之 GC3 組成限制，也影響密碼子的使用。而 GC3 突變偏性又受到 GC12 之影響，造成 *Xa* 基因之同義密碼子使用偏性。

關鍵詞：同義密碼子使用偏性、基因表達程度、GC 含量、突變偏性、分子演化。

前 言

生物的蛋白質合成都是由 DNA 上連續排列的遺傳密碼 (genetic code) 所決定，DNA 的遺傳訊息轉錄到 mRNA 上，通過翻譯讀寫

mRNA 上的遺傳密碼，將 DNA 上遺傳訊息轉變為蛋白質的胺基酸排列順序。mRNA 分子中每 3 個相鄰鹼基代表一種胺基酸或終止信號，此 3 個相鄰鹼基稱為密碼子 (codon)。因此，密碼子是核苷酸攜帶訊息和蛋白質攜帶訊息間

1. 行政院農業委員會農業試驗所研究報告第 2400 號。接受日期：99 年 6 月 21 日。
2. 本所作物組助理研究員、計畫助理、助理研究員。台灣 台中縣 霧峰鄉。
3. 本所研究員兼主任秘書。台灣 台中縣 霧峰鄉。
4. 通訊作者，電子郵件：iying@tari.gov.tw；傳真機：(04)23302801。

對應的基本規則，是生物體內資訊傳遞的基本環節。編碼 20 種胺基酸的密碼子共 61 種，因此有 2 個或 2 個以上轉譯成同一胺基酸之情形，這些表現相同胺基酸的密碼子稱為同義密碼子 (synonymous codon)。大多數生物體或基因通常傾向使用部份特定的同義密碼子，此種生物中普遍存在的同義密碼子非均衡使用現象，稱為同義密碼子使用偏性 (synonymous codon usage bias)。例如密碼子 GUU、GUC、GUA、GUG 皆決定纈胺酸 (valine, Val)，若突變使密碼 GUU 改變成 GUC，則仍然可轉譯為 Val，不影響蛋白質的活性，此意味著 tRNA 的反密碼子 (anticodon) 的第一個鹼基 (從 5' 端算起) 與 mRNA 的密碼子之第三個鹼基並非一對一的關係，此即所謂的搖擺位點 (wobble position)。

同義密碼子使用偏性會因物種或基因不同而有所差異 (Gu *et al.* 2002; Fan *et al.* 2003)，此與諸多因素有關，如：鹼基組成突變偏性 (base compositional mutation bias) (Karlin & Mrazek 1996; Knight *et al.* 2001; Huo & Yang 2003)、自然選擇 (nature selection) (Sharp & Li 1986; Duret & Mouchiroud 1999; Peixoto *et al.* 2003; Romero *et al.* 2003)、基因長度 (Moriyama & Powell 1998)、tRNA 豐富度 (Percudani *et al.* 1997; Duret 2000)、mRNA 二級結構 (Gu *et al.* 2004)、密碼子與反密碼子之交互作用 (Shi *et al.* 2001)、蛋白質結構 (Gupta *et al.* 2000)、蛋白質親水程度及胺基酸保守性 (Romero *et al.* 2000)、轉譯起點效應 (Stenström *et al.* 2001) 等。然而，不同物種基因體發生密碼子使用偏性的原因也不盡相同 (Berezhnoy & Duplij 2005; Liu & Xue 2005)。大腸桿菌 (*Escherichia coli*) 及釀酒酵母 (*Saccharomyces cerevisiae*) 的密碼子用法取決於突變偏性 (mutational biases) 及對最優密碼子選擇 (selection for optimal codons) 的平衡，而與此相反，一些極

端富含 AT 或 GC 的原核生物及哺乳動物中基因的密碼子用法主要由突變偏性造成 (Bulmer 1988; Sharp *et al.* 1993)。但是，Duret & Mouchiroud (1999) 卻發現在線蟲 (*Caenorhabditis elegans*)、果蠅 (*Drosophila melanogaster*) 及阿拉伯芥 (*Arabidopsis thaliana*) 中該特性並非由突變引起，而是取決於對密碼子搖擺位點的選擇。Liu *et al.* (2003) 則發現水稻的 GC 含量決定了密碼子的使用，而主要與轉錄水平上的突變有關。

同義密碼子使用偏性究竟是由突變偏性或轉譯選擇 (translational selection) 產生，還是兩者共同作用產生，迄今結論不甚一致 (Salim & Cavalcanti 2008)。Sharp & Li (1987) 認為密碼子使用偏性在高效表達基因中是由轉譯水準上的選擇壓力主宰，在低表達基因或群體比較小的物種中則主要經由突變或漂移產生。但是 Chiapello *et al.* (1998) 發現轉譯時密碼子的選擇是造成阿拉伯芥密碼子使用偏性的重要角色，Karlin & Mrazek (1996) 卻認為人類基因體之密碼子使用偏性變異以突變偏性較轉譯選擇更為重要。在大腸桿菌、釀酒酵母、線蟲、果蠅、阿拉伯芥及水稻等生物基因體中，皆發現基因表達程度與密碼子使用偏性間存在顯著相關關係 (Ikemura 1985; Sharp & Li 1987; Duret & Mouchiroud 1999)。Liu *et al.* (2004a) 也發現基因表達程度是造成梗稻粒線體中密碼子使用偏性的主要因素，高度表達之粒線體基因的密碼子通常含有豐富的 A 或 T，且偏好以 A 或 U 結尾；但梗稻粒線體的基因長度與基因表達程度、密碼子使用偏性之間並無明顯關係。Liu *et al.* (2004b) 利用 NCBI (National Center for Biotechnology Information) 之 GenBank 資料庫中 1860 條大於 300 bp 的水稻 CDS 探討同義密碼子使用偏性之影響因素，認為水稻密碼子使用偏性之變異性是由於轉錄時核苷酸突變偏性所造成，水稻基因體中高度表達基因較偏好某

種 GC 含量豐富的同義密碼子。Liu *et al.* (2003) 以水稻品種 93-11 作為材料進行基因體同義密碼子使用偏性分析，則發現基因表達程度雖會影響密碼子使用偏性，但兩者間相關係數較小。顯然，影響水稻基因體之同義密碼子使用偏性的因素，可能因目標基因而有所不同。

水稻白葉枯病 (rice bacterial leaf blight) 是由黃單孢桿菌 (*Xanthomonas oryzae* pv. *oryzae*, *Xoo*) 引起的一種細菌性維管束病害，為世界性主要稻作病害之一。為瞭解水稻白葉枯病抗病基因 (*Xa*) 之抗性分子機制，前作 (Wei *et al.* 2009) 自 NCBI 之 GenBank 資料庫 (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>, Guisez *et al.* 1993) 下載已完全定序的 *Xa* 基因 (*Xa1*, *xa5*, *xa13*, *Xa13*, *Xa21*, *Xa26* 及 *Xa27*) 及其家族之 17 個編碼區序列 (coding domain sequence, CDS) 進行密碼子使用偏性分析，發現同樣對 *Xoo* 產生抗性的不同 *Xa* 基因間呈現出密碼子使用偏性的多樣性，*Xa1*、*Xa21* 及 *Xa26* 的 CDS 僅少數胺基酸具有高頻密碼子 (high-frequency codon) 且大多偏好以 A 或 U 結尾的密碼子，而 *xa5*、*xa13*、*Xa13* 及 *Xa27* 之 CDS 的所有胺基酸都偏好以 C 或 G 結尾的密碼子，由此揭示出選擇在抗病基因的演化中有著重要的作用。因此，值得進一步探討 *Xa* 基因密碼子使用偏性之影響因素，可有利 *Xa* 基因分子機制演化上之研究，並促進蛋白質結構預測與基因改良。

材料與方法

Xa 基因序列之蒐集

本研究自 NCBI 之 GenBank 資料庫下載已定序之 *Xa* 基因的 CDS，這些已知 *Xa* 基因的相關資訊整理如表 1，*Xa1* 分別於 1999 及 2004 年登錄有兩條序列 (本文簡稱 *Xa1-1999* 及 *Xa1-2004*)，各位於第 4 及第 10 條染色體；*xa5*

為一隱性基因，位於第 5 條染色體；*Xa13* 及 *xa13* 分別為顯、隱性基因，皆位於第 8 條染色體；*Xa21* 基因家族位於第 11 條染色體，包含 7 個成員分別為 *Xa21-A1*、*Xa21-A2*、*Xa21-B*、*Xa21-C*、*Xa21-D*、*Xa21-E* 及 *Xa21-F*；*Xa26* 基因家族位於第 11 條染色體，包含 4 個成員分別為 *Xa26-MRKA*、*Xa26-MRKB*、*Xa26-MRKC* 及 *Xa26-MRKD* (本文簡稱 *Xa26a*、*Xa26b*、*Xa26c* 及 *Xa26d*)；*Xa27* 位於第 6 條染色體。

同義密碼子使用偏性指標值之計算

將上述各 *Xa* 基因之 CDS，利用 CodonW 密碼子分析軟體 (1.4 版，<http://bioweb.pasteur.fr/seqana/interfaces/codonw.html>, Peden 1999)，先求出各基因序列上各種密碼子的個數，然後進行以下各種同義密碼子使用偏性指標之計算。編碼 20 種胺基酸的密碼子共有 61 種 (附錄 1)，其中由於甲硫胺酸 (methionine, Met)、色胺酸 (tryptophan, Trp) 各只對應一個密碼子，因此在以下計算過程中皆未將其納入，而只針對各基因之 59 種同義密碼子進行同義密碼子使用偏性指標值之估算。

同義密碼子相對使用度 (relative synonymous codon usage, RSCU) 及其對應分析 (correspondence analysis, COA)：RSCU 通常用來研究整個基因的密碼子使用變異。編譯成同一種胺基酸的同義密碼子從理論上期望是有相同的使用頻率 (具有隨機性)，RSCU 是實際觀測值與理論期望值間之比率，其大小反映出胺基酸內同義密碼子的偏性變異。為探討不同 *Xa* 基因間密碼子使用性的相似及相異程度，在不考慮 Met、Trp 下，計算出各基因之 59 種同義密碼子之 RSCU 值，其計算式 (Sharp *et al.* 1993) 為

$$RSCU_{ij} = \frac{X_{ij}}{\frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij}} \quad [1]$$

表 1. 已知序列之水稻 Xa 基因之解碼區序列 (CDS) 長度及各種同義密碼子使用偏性指標值

Table 1. Length of coding domain sequence (CDS) and indices of synonymous codon usage bias of bacterial leaf blight resistant genes (Xa) of rice with known sequences^z

Label	Gene symbol	Accession no. in NCBI	Submitted year	Gene length (bp)	Chromosome location	CDS Length (bp)	CAI ^y	ENC ^y	GC3 ^y	GC12 ^y
Xa1-1999	Xa1	AB002266	1999	5910	4	5409	0.183	54.25	0.440	0.459
Xa1-2004	Xa1	NM_194685	2004	4206	10	4206	0.209	56.91	0.543	0.462
xa5	xa5	AY643716	2004	318	5	318	0.273	39.93	0.673	0.429
xa13	xa13	DQ421394	2006	924	8	924	0.266	31.06	0.956	0.508
Xa13	Xa13	DQ421395	2006	924	8	924	0.261	31.08	0.953	0.510
Xa21-A1	Xa21	U72725	1997	8416	11	2991	0.196	55.94	0.446	0.465
Xa21-A2	Xa21	U72727	1998	5940	11	2991	0.192	55.44	0.443	0.463
Xa21-B	Xa21	U37133	1995	3921	11	3078	0.168	55.56	0.464	0.458
Xa21-C	Xa21	U72723	1998	19639	11	2978	0.195	57.12	0.445	0.472
Xa21-D	Xa21	U72726	1997	13341	11	1839	0.166	58.44	0.513	0.445
Xa21-E	Xa21	U72724	1997	9424	11	2442	0.202	57.11	0.459	0.454
Xa21-F	Xa21	U72728	2005	7204	11	3071	0.205	55.70	0.504	0.454
Xa26a	Xa26	DQ355952	2006	3401	11	3297	0.180	57.34	0.439	0.469
Xa26b	Xa26	DQ355952	2006	3631	11	3312	0.199	57.99	0.480	0.464
Xa26c	Xa26	DQ355952	2006	3456	11	3342	0.176	57.02	0.475	0.459
Xa26d	Xa26	DQ355952	2006	11396	11	3327	0.200	57.24	0.482	0.470
Xa27	Xa27	AY986491	2005	2393	6	342	0.213	33.09	0.798	0.663

^z Source: Oryzabase and NCBI databases, searched on Feb. 29, 2008.

^y CAI, codon adaption index; ENC, effective number of codons; GC3, frequency of G or C nucleotides present at the third position of synonymous codons; GC12, mean frequencies of G or C nucleotides present at the first and second positions of synonymous codons.

式中 X_{ij} 是第 i 個胺基酸中第 j 個密碼子出現的次數， n_i 為第 i 個胺基酸同義密碼子的數目。RSCU 用以觀測同義密碼子出現頻率是否相等，當 RSCU 值大於 1 時表示相對應的密碼子出現的次數比期望次數（即該胺基酸所有同義密碼子出現平均頻率，意味著其出現頻率都相等）為高，反之則表示出現次數相對較少 (Sharp *et al.* 1993)。本研究同義密碼子使用性是由不同 Xa 基因 (樣本) 與 59 種同義密碼子出現頻率 (變數) 所構成的資料矩陣來決定，此係一種多維 (multiple dimensions) 的資料形式，最適合利用多變數分析 (multivariate

analysis) 之統計技術來解析其所蘊含之訊息。對應分析 (correspondence analysis) 為 Benzécri 於 1973 年所提出，是一種廣泛運用到生態學 (尤其是植物生態學) 的多變數統計分析技術，現已大量應用到其他各方面的研究領域 (Beh 2004)，包括基因組分析上之各種應用 (Gupta & Ghosh 2001; Tekaiia *et al.* 2002; Tan *et al.* 2004)。對應分析對於數據的要求比較簡單，不需要有嚴格的線性關係，也無須限定所處理的資料是連續或離散變數，只需要是一個列聯表 (contingency table) 的資料形式，即可將多維之變數或樣本於一個低維空間上以一個

序列分布圖 (ordination diagram) 呈現, 藉此不但可用來檢測變數間或樣本間的關係, 也能從各變數或樣本落在序列分布圖上之距離遠近來判斷分群結果 (Benzécri 1973)。密碼子使用性之對應分析, 能區分高表達基因和低表達基因 (Gouy & Gautier 1982; Holm 1986; Peden 1999), 並可有效檢測出不同基因間之密碼子使用性的變異形式 (Gupta & Ghosh 2001; Wei *et al.* 2009)。因此, 本研究根據 RSCU 值利用 STATISTICA 統計軟體 (7.0 版, Statsoft Inc. 2002) 進行對應分析, 由各 X_a 基因落在所得序列分布圖之因子軸 (COA axis) 上的相對位置, 來探討 X_a 基因之同義密碼子使用偏性。

密碼子適應指標 (codon adaption index, CAI) : CAI 是透過計算某基因中每個密碼子的相對貢獻及所有密碼子的頻率, 來做為測量該基因之同義密碼子使用偏性大小的一個指標, 用以評估哪一種選擇在形成密碼子使用模式時是有效的, 故可有效運用於預測基因的表達程度、估計病毒基因對宿主的適應性及比較不同生物之間密碼子之使用模式 (Sharp & Li 1987)。Nakamura & Tabata (1997) 以 7 個基因組的表達模式, 證實了 CAI 能有效預測基因表達程度。自此, CAI 已成為最常用的基因表達程度之估計值 (Salim & Cavalcanti 2008)。CAI 的計算步驟如下 (Sharp & Li 1987): 首先, 利用 RSCU 值計算出密碼子相對適應值 (relative adaptiveness of a codon, W), W 為同一胺基酸中各密碼子使用頻率與最優密碼子 (即出現頻率最高之密碼子) 使用頻率的比值, 即

$$W_{ij} = \frac{RSCU_{ij}}{RSCU_{i_{max}}} = \frac{X_{ij}}{X_{i_{max}}} \quad [2]$$

式中 $RSCU_{i_{max}}$ 為第 i 個胺基酸中最大的 RSCU 值; $X_{i_{max}}$ 為第 i 個胺基酸中出現頻率最高的密碼子個數。

$$CAI = CAI_{obs}/CAI_{max} \quad [3]$$

$$\text{其中 } CAI_{obs} = \sqrt[L]{\prod_{k=1}^L RSCU_k} \quad [4]$$

$$CAI_{max} = \sqrt[L]{\prod_{k=1}^L RSCU_{k_{max}}} \quad [5]$$

式中 $RSCU_k$ 為基因中第 k 個密碼子的 RSCU 值; $RSCU_{k_{max}}$ 為基因中第 k 個密碼子所編碼之胺基酸中最大 RSCU 值; L 為基因中密碼子的數目。因此, CAI 的精確計算式可寫成

$$CAI = \sqrt[L]{\prod_{k=1}^L W_k} \quad [6]$$

由該公式可知, CAI 是密碼子相對適應值 (W) 的幾何平均值 (geometric mean), 顯示出其與基因表達程度之間具有正向相關性。CAI 值介於 0-1 之間, 高度表達之基因, 通常 CAI 值也越高。但 Sharp & Li (1987) 強調 CAI 只能大致估計出基因內密碼子使用性的適合度 (suitability), 而不考慮密碼子在基因上的分布。

有效密碼子數 (effective number of codons, ENC) : ENC 為觀察在基因序列中密碼子被真正使用的密碼子使用偏性之測量方法, 其數值範圍在 20-61 之間 (Wright 1990)。當每個胺基酸都使用到一種密碼子時, ENC 為 20; 若各胺基酸皆無偏好任何同義密碼子時, 則 ENC 當近於 61。ENC 的計算步驟如下 (Wright 1990):

先計算各胺基酸的密碼子均質性 (homozygosity, F), 計算式為

$$\hat{F} = \frac{n \sum_{i=1}^k p_i^2 - 1}{n - 1} \quad [7]$$

式中 k 為同義密碼子數, 例如白胺酸 (leucine, Leu) 有 6 個密碼子 (附錄 1), 則 $k = 6$; n 為某一胺基酸之同義密碼子的總和, 例如 Leu 中 $n = n_1 + n_2 + \dots + n_6$; p_i 為某一胺基酸之第 i 個同義密碼子頻率, 例如 Leu 中的 $p_i = \frac{n_i}{n}$ 。

再求 ENC，其在 Wright (1990) 使用的符號為 \hat{N}_c ，計算公式為

$$\hat{N}_c = 2 + \frac{9}{\hat{F}_2} + \frac{1}{\hat{F}_3} + \frac{5}{\hat{F}_4} + \frac{3}{\hat{F}_6} \quad [8]$$

式中 \hat{F}_i 為每一種同義密碼子家族，如 \hat{F}_6 表示所有包含 6 個同義密碼子之同義密碼子家族的均質性，如 Leu、精胺酸 (arginine, Arg)、絲胺酸 (serine, Ser) 之 3 種胺基酸 (附錄 1)。由於在計算時，可能有些基因會缺少某些密碼子，針對此類情況，可循以下幾種方法處理：(1) 若某一同義密碼子家族的 \hat{F} 計算結果為 0，則對於之前 \hat{N}_c 的計算應加以調整，即在計算某些同義密碼子家族的平均均質性時，應按照實際出現的同義密碼子家族之數目來計算；(2) 包含 3 個同義密碼子之同義密碼子家族的異白胺酸 (isoleucine, Ile) (附錄 1) 若缺失或很少被使用，則 \hat{F}_3 應計算為 \hat{F}_2 與 \hat{F}_4 之平均值；(3) 由於某些基因之密碼子之使用可能具有極度偏性，而造成 \hat{N}_c 值可能大於 61，此時該基因之 \hat{N}_c 值則以 61 計算。

同義密碼子第三位置出現 G 或 C 核苷酸之頻率 (GC3)：GC3 含量可能是影響同義密碼子使用偏性的因素之一。一般而言在計算各基因之同義密碼子的 GC3 時，並不納入只對應一個密碼子的 Met 和 Trp。

同義密碼子第一、二位置出現 G 或 C 核苷酸之平均頻率 (GC12)：GC12 為在不考慮 Met、Trp 下同義密碼子第一位置出現 G 或 C 之頻率 (GC1) 及其第二位置出現 G 或 C 之頻率 (GC2) 的平均值，其計算式為

$$GC12 = \frac{(GC1 + GC2)}{2} \quad [9]$$

密碼子使用偏性影響因素之分析

CDS 長度對基因表達程度之影響：將各基因的 CDS 長度與 CAI 進行簡單相關，並繪製散佈圖，以探討 CDS 長度是否影響基因表達程度。

CDS 長度對同義密碼子使用偏性之影響：將各基因的 CDS 長度與在 RSCU 之對應分析所得序列分布圖中解釋變異具有最大貢獻率的第一軸 (COA axis 1) 上位置、以及與 ENC 分別進行簡單相關，並繪製散佈圖，以探討 CDS 長度是否為影響同義密碼子使用偏性之因素。

基因表達程度對同義密碼子使用偏性之影響：將各基因的 CAI 與在 RSCU 之 COA axis 1 上位置及 ENC 分別進行簡單相關，並繪製散佈圖，以探討基因表達程度是否為影響同義密碼子使用偏性之因素。

GC3 含量對同義密碼子使用偏性之影響：為探討基因中 GC3 含量是否會影響同義密碼子使用偏性，因此本研究以 GC3 為 X 軸而 ENC 為 Y 軸，繪製一個 Nc-plot 圖 (Wright 1990)，可用來檢視同義密碼子使用偏性是否因 GC3 組成的壓力限制 (compositional constraints) 所造成。若密碼子為隨機使用，ENC 與 GC3 的期望關係可根據 Wright (1990) 提出之方法求得，其使用的符號及公式如下：

$$Nc = 2 + s + \left[\frac{29}{s^2 + (1-s)^2} \right] \quad [10]$$

式中符號 Nc 即 ENC，s 為 GC3 值 (註：Wright 之原著印刷有誤，Nc 公式的分母被誤寫成 $s^2 + (1-s^2)$ ，本文在此予以修正)。

透過 Nc-plot 可看出，若基因受到 GC3 組成之壓力限制影響，則其 Nc 會落在期望值平滑曲線正上面或曲線之下方，表示基因可能受到 GC 突變偏性之影響或者受到密碼子第三位置對 G 或 C 之選擇，而造成密碼子使用偏性 (Wright 1990)。

同時，為評估 ENC 期望值與估計值間之差異，計算 ENC 的估計值與期望值之比率 (ENC ratio)：

$$ENC \text{ ratio} = \frac{(ENC_{\text{exp}} - ENC_{\text{obs}})}{ENC_{\text{exp}}} \quad [11]$$

ENC_{obs} 為基因經由算式 [8] 所計算出之 ENC 值； ENC_{exp} 為某基因所含之 GC3 含量時，利用 ENC 與 GC3 之期望關係的公式 [10] 所算出之期望值。然後，以基因為 X 軸，ENC ratio 為 Y 軸，繪製曲線分布圖，其結果可用以佐證 Nc-plot 之結果 (Kawabe & Miyashita 2003)。

GC12 含量對 GC3 突變偏性之影響：為瞭解突變偏性與密碼子使用偏性之間的關係，以 GC3 為 X 軸，GC12 為 Y 軸，繪製中性圖 (neutrality plot)，用以觀察 GC12 含量是否為 GC3 含量突變偏性之影響因素 (Sueoka 1988)。若 GC3 和 GC12 之間存在顯著相關關係，則表示基因會受突變偏性影響，若他們之間無顯著關係，則表示該等基因有較低的突變偏性或較高的保守性。

結 果

已知序列之 *Xa* 基因之 CDS 的長度及各種同義密碼子使用偏性指標值 (除 RSCU 以外)，整理如表 1。由表 1 可知，就密碼子適應指標 CAI 越高，表示基因之表達程度越高來看，所有 *Xa* 基因的表達程度並不高；而較短 CDS (318–924 bp) 之 *xa5*、*xa13*、*Xa13* 及 *Xa27* 基因的 CAI 值介於 0.213–0.273 之間，至於 CDS 較長 (1839–5409 bp) 的 *Xa1*、*Xa21* 及 *Xa26* 基因家族之 CAI 值則在 0.168–0.209 之間，似乎 *Xa* 基因之 CDS 長度較短者可有較高的基因表達程度。*xa5*、*xa13*、*Xa13* 及 *Xa27* 基因之 CDS 的 ENC 值在 31.06–39.93 之間，該有效密碼子數較接近於 20，表示胺基酸的同義密碼子使用偏性較高；然而，*Xa1*、*Xa21* 及 *Xa26* 基因家族之 CDS 的 ENC 值在 54.25–58.44 之間，較接近於 61，表示其序列胺基酸較不偏好任何密碼子。*xa13* 及 *Xa13* 基因之 CDS 的 GC3 值甚近於 1，表示其同義密碼子第三位置的 G 或 C 含量很高，其次為 *Xa27* 基因；*Xa1*、*Xa21* 及 *Xa26* 基因家族之 CDS 的同義密碼子第三位置的 G

或 C 含量則相對較低，GC3 值在 0.5 左右。所有 *Xa* 基因之 CDS 的同義密碼子第一、二位置之 G 或 C 含量則並無偏高現象，其 GC12 值在 0.5 左右，未接近於 1。

將各 *Xa* 基因之 CDS 的 RSCU 值進行對應分析結果 (圖 1)，兩個主要因子軸能解釋變異的貢獻分別為 56.94% 和 19.61%，總貢獻量合計佔 76.55%，此意謂著利用對應分析所得之二維序列分布圖的兩個因子軸，當足以充分說明 *Xa* 基因間密碼子使用偏性的差異趨勢，且以第一軸 (COA axis 1) 能夠解釋過半數的密碼子使用偏性。因此，由各基因落在 RSCU 之 COA axis 1 上的位置，可顯示各基因密碼子使用性的變異形式；基因在 RSCU 之 COA axis 1 上的落點越偏離 0，表示其同義密碼子使用偏性越高；而各基因落點點之間有群團現象，表示這些基因的同義密碼子使用性相近。由圖 1 發現，*xa5*、*xa13*、*Xa13* 及 *Xa27* 都甚遠離於其他 *Xa* 基因，且 *xa5* 及 *Xa27* 亦各偏離於 *xa13* 及 *Xa13*，此表示 *xa13* 及 *Xa13* 之 CDS 的胺基酸之同義密碼子使用偏性與 *xa5* 及 *Xa27* 所得者有差異，而此四個 *Xa* 基因之同義密碼子使用偏性又甚異於 *Xa1*、*Xa21* 及 *Xa26* 基因家族。

將各 *Xa* 基因的 CDS 長度與 CAI 進行相關分析，結果發現 CDS 長度會影響基因表達程度，其間存在極顯著的負相關關係 ($P < 0.001$)，即 CDS 長度越短則基因表達程度越高 (圖 2)。透過 *Xa* 基因的 CDS 長度與在 RSCU 之 COA axis 上位置間存在之極顯著負相關關係，以及 CDS 長度與 ENC 間之極顯著正相關關係，可知 CDS 長度越短則同義密碼子使用偏性也越高 (圖 3、4)。再將 CAI 分別與 RSCU 之 COA axis 1 上位置及 ENC 進行簡單相關，發現 CAI 與在 RSCU 之 COA axis 1 上位置之間呈極顯著正相關 (圖 5)，而 CAI 與 ENC 間則呈極顯著負相關 (圖 6)，顯示 *Xa* 基因之表達程度確實會影響其同義密碼子使用偏性。為探討

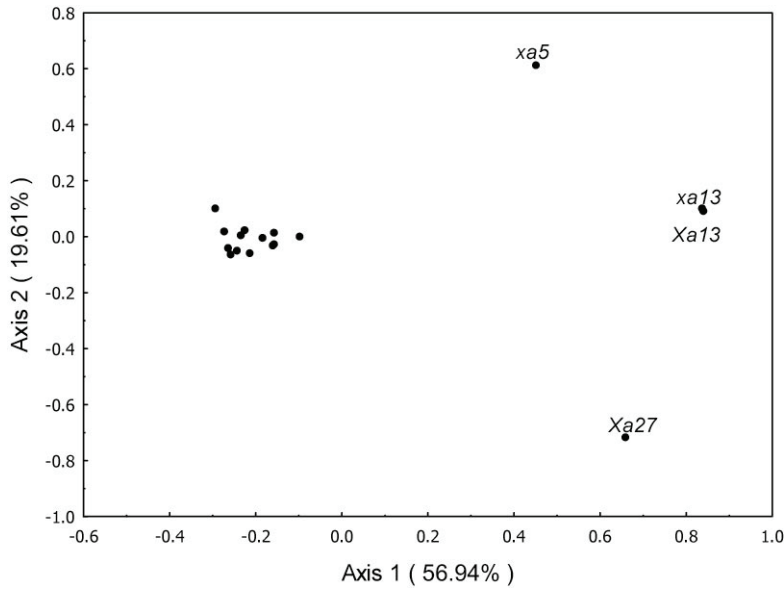


圖 1. 水稻 *Xa* 基因根據同義密碼子相對使用度 (RSCU) 的對應分析之序列分布圖。

Fig. 1. Ordination diagram of correspondence analysis based on the relative synonymous codon usage (RSCU) for bacterial leaf blight resistant genes (*Xa*) of rice.

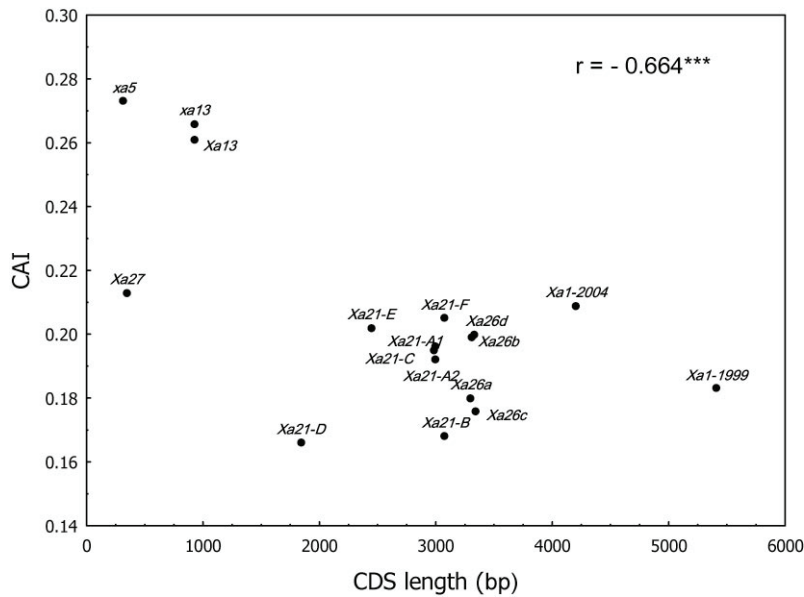


圖 2. 水稻 *Xa* 基因的編碼區序列 (CDS) 長度與密碼子適應指標 (CAI) 間之相關關係。

Fig. 2. Correlation between coding domain sequence (CDS) length and codon adaption index (CAI) for bacterial leaf blight resistant genes (*Xa*) of rice. *** Significant at 0.1% level.

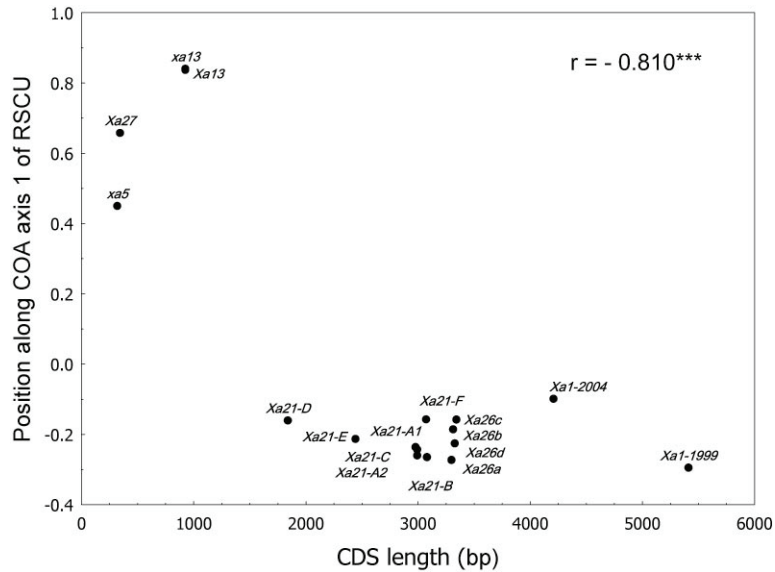


圖 3. 水稻 *Xa* 基因的編碼區序列 (CDS) 長度與在同義密碼子相對使用度 (RSCU) 之對應分析第一因子軸 (COA axis 1) 上位置間之相關關係。

Fig. 3. Correlation between coding domain sequence (CDS) length and the position along the first axis of correspondence analysis (COA axis 1) of relative synonymous codon usage (RSCU) for bacterial leaf blight resistant genes (*Xa*) of rice. *** Significant at 0.1% level.

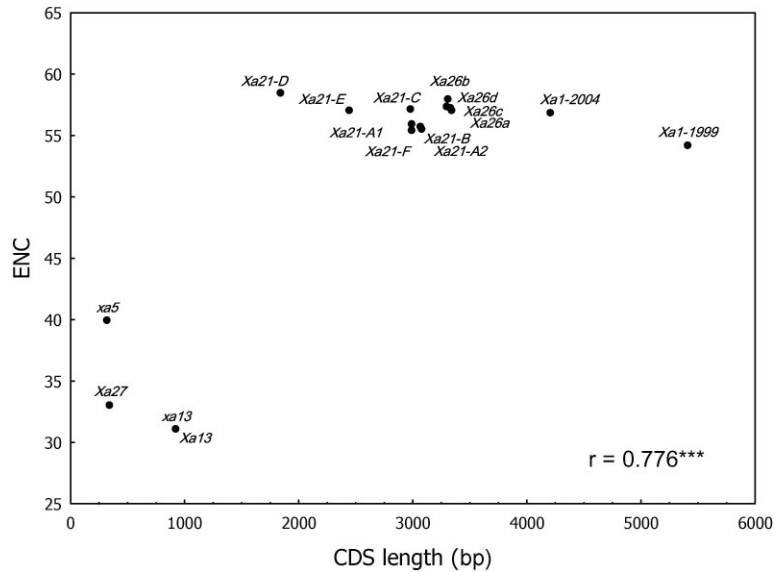


圖 4. 水稻 *Xa* 基因的編碼區序列 (CDS) 長度與有效密碼子數 (ENC) 間之相關關係。

Fig. 4. Correlation between coding domain sequence (CDS) length and effective number of codon (ENC) for bacterial leaf blight resistant genes (*Xa*) of rice. *** Significant at 0.1% level.

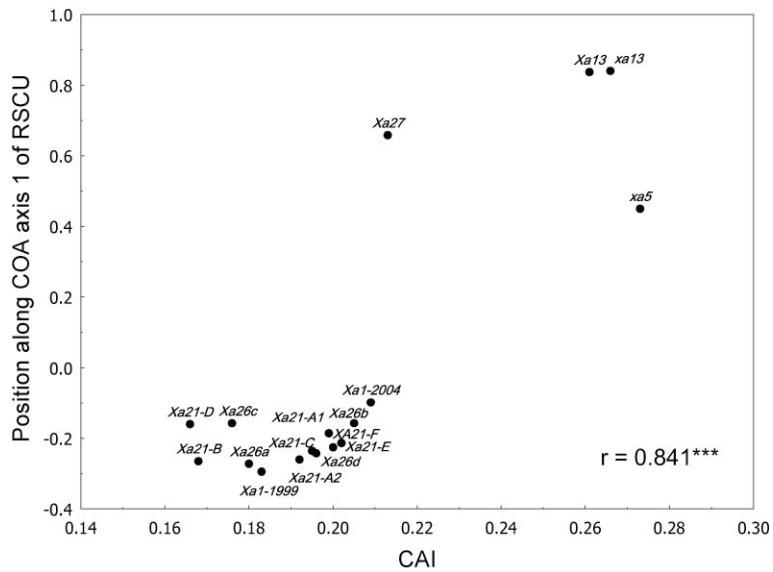


圖 5. 水稻 *Xa* 基因之密碼子適應指標 (CAI) 與在同義密碼子相對使用度 (RSCU) 之對應分析第一因子軸 (COA axis 1) 上位置間之相關關係。

Fig. 5. Correlation between codon adaption index (CAI) and the position along the first axis of correspondence analysis (COA axis 1) of relative synonymous codon usage (RSCU) for bacterial leaf blight resistant genes (*Xa*) of rice. *** Significant at 0.1% level.

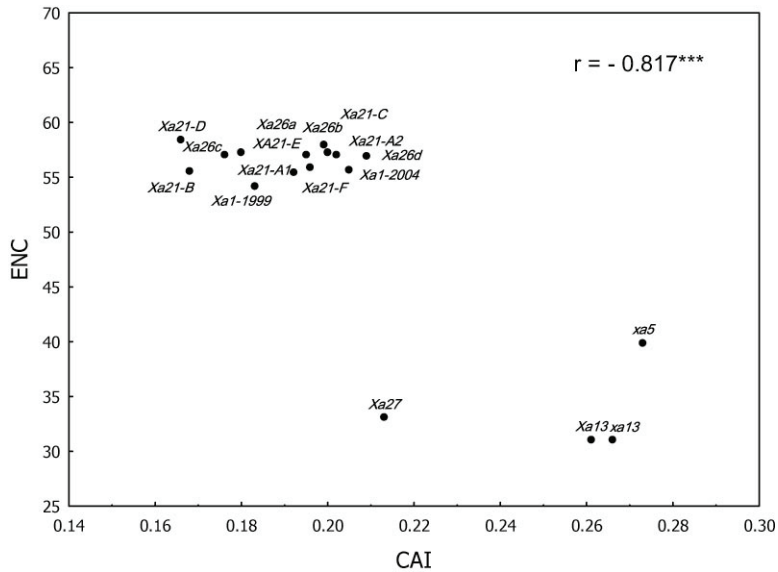


圖 6. 水稻 *Xa* 基因的密碼子適應指標 (CAI) 與效密碼子數 (ENC) 間之相關關係。

Fig. 6. Correlation between codon adaption index (CAI) and effective number of codon (ENC) for bacterial leaf blight resistant genes (*Xa*) of rice. *** Significant at 0.1% level.

Xa 基因 CDS 的 GC3 含量是否影響其同義密碼子使用性，由圖 7 之 Nc-plot 可看出，所有 *Xa* 基因皆落於期望曲線之下方，表示所有 *Xa* 基因皆因 GC3 組成而造成其同義密碼子使用偏性，尤以 CDS 長度較短 (< 1000 bp) 的 *xa5* 及 *Xa27* 基因受到 GC3 組成之影響更鉅 (更明顯偏離期望曲線)；此結論亦可由圖 8 之基因對 ENC ratio (基因間 ENC 期望值與估計值間之差異) 的曲線分布圖獲得佐證，各基因之 ENC ratio 皆為正值，而 *xa5* 及 *Xa27* 之 ENC ratio 遠大於其他 *Xa* 基因，表示 *xa5* 及 *Xa27* 確實距離期望曲線較遠，其同義密碼子使用偏性受到 GC3 組成限制之影響更甚於其他 *Xa* 基因。繼由圖 9 之中性圖可知，GC12 與 GC3 之間呈顯著正相關 ($P < 0.05$)，顯示 GC 含量在 *Xa* 基因之同義密碼子使用偏性的產生過程中扮演著決定性的作用，且 *Xa* 基因密碼子之 GC3 組成也

受到其 GC12 之影響，導致可能發生較高的突變偏性而造成同義密碼子使用上之偏性。

討 論

本研究發現不同 *Xa* 基因存在密碼子使用偏性變異，*Xa1*、*Xa21* 及 *Xa26* 基因家族之序列胺基酸較不偏好任何密碼子，而 *xa5*、*xa13*、*Xa13* 及 *Xa27* 基因卻有較高的同義密碼子使用偏性 (表 1、圖 1)。進而根據本研究同義密碼子使用偏性之影響因素探討，顯示 *Xa* 基因長度較短者可有較高的基因表達程度、以及較高的同義密碼子使用偏性，如 *xa5*、*xa13*、*Xa13* 及 *Xa27* 基因 (圖 2-圖 6)。因此，CDS 長度及基因表達程度是影響 *Xa* 基因同義密碼子使用偏性的部份因素。而在其他水稻基因，如稈稻粒線體基因之表達程度是造成密碼子使用偏性的因素之一，但其 CDS 長度則並不影響基因表

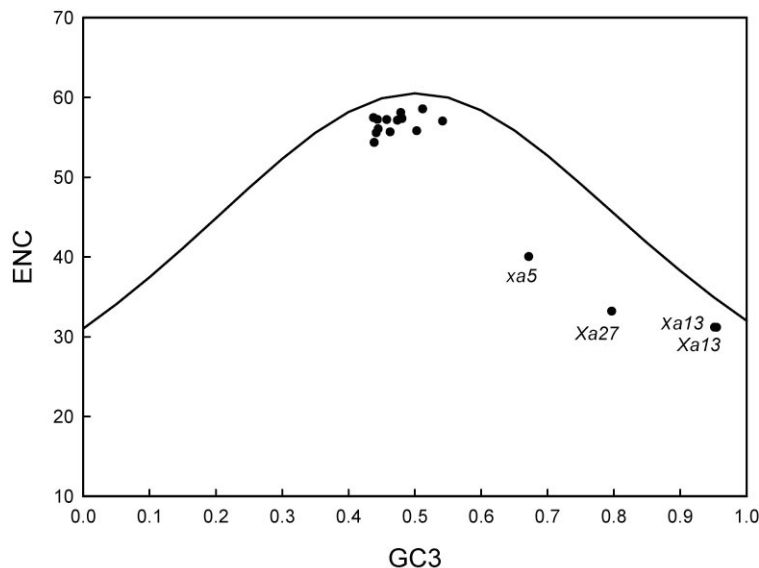


圖 7. 水稻 *Xa* 基因之 Nc-Plot：圖中實線為在密碼子隨機使用下之密碼子第三位置出現 G 或 C 核苷酸之頻率 (GC3) 與有效密碼子數 (ENC) 之期望值。

Fig. 7. Nc-plot for bacterial leaf blight resistant genes (*Xa*) of rice. The solid curve represents the expected curve between frequency of G or C nucleotides present at the third position of synonymous codons (GC3) and effective number of codon (ENC) under random codon usage.

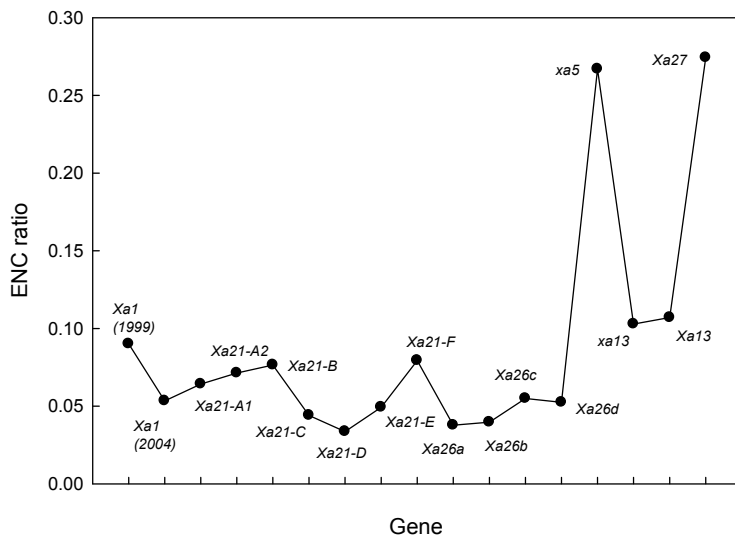


圖 8. 水稻 *Xa* 基因的有效密碼子數估計值與期望值之比率 (ENC ratio) 的曲線分布圖。

Fig. 8. Distribution curve of the ratio between observed and expected effective number of codon (ENC ratio) for bacterial leaf blight resistant genes (*Xa*) of rice.

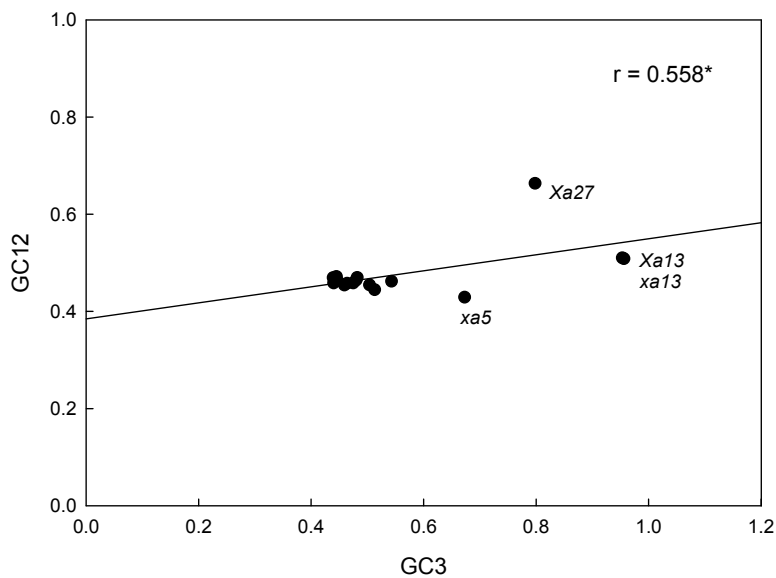


圖 9. 水稻 *Xa* 基因之中性圖。

Fig. 9. Neutrality plot for bacterial leaf blight resistant genes (*Xa*) of rice. * Significant at 5% level.

達程度及密碼子使用偏性 (Liu *et al.* 2004a) ; Liu *et al.* (2003) 在水稻品種 93-11 之基因體上發現 CDS 長度較短之基因的密碼子使用偏性較高，而基因表達程度多少影響密碼子使用偏性，但兩者間相關係數較小；Liu *et al.* (2004b) 認為水稻密碼子使用偏性之變異性是由於轉錄時核苷酸突變偏性所造成，水稻基因體中高度表達基因較偏好某種 GC 含量豐富的同義密碼子，且 CDS 長度越短，水稻轉錄活性表達程度越高。一般而言，CDS 長度為影響水稻密碼子使用偏性的因素之一，當 CDS 長度越短，其密碼子使用偏性越高；但當目標基因不同，影響程度也會不同 (Liu *et al.* 2004b; Liu *et al.* 2005)。在多細胞真核生物中，基因長度與密碼子使用偏性之關係的內在機制和原因還不是很清楚 (Salim & Cavalcanti 2008)；有可能是因為遺傳上連鎖的選擇位點的進化動力學所造成 (Comeron *et al.* 1999; McVean & Charlesworth 2000)，即所有的都是相等，但較長基因內的位點將與更大數目的分離非中性突變位點緊密相連，如此形成有更低的密碼子使用偏移 (Comeron *et al.* 1999)。Liu *et al.* (2005) 則指出高表達基因比低表達基因對密碼子偏好性有顯著的變異，是因為高表達基因有較強的天擇限制，會選擇密碼子以進行有效的轉譯，而影響高表達基因最主要之因素為 tRNA 的豐富度，即高表達基因會要求密碼子的使用頻率與 tRNA 數量及轉錄能力相符。

影響 *Xa* 基因之同義密碼子使用偏性，還包括 GC3 之組成限制 (圖 7·8)。Wei *et al.* (2009) 發現具有多白胺酸重複 (leucine-rich repeat, LRR) 區域結構且 CDS 長度大於 1000 bp 的 *Xa1*、*Xa21* 及 *Xa26* 基因的大多數胺基酸並無密碼子使用偏性，僅少數胺基酸偏好以 A 或 U 結尾的同義密碼子；而不具 LRR 區域蛋白質結構且 CDS 長度較短的 *xa5*、*xa13*、*Xa13* 及 *Xa27*

基因，則幾乎所有胺基酸都偏好以 C 或 G 結尾的同義密碼子。由此顯見，同樣對白葉枯病菌產生抗性的不同 *Xa* 基因間所呈現的密碼子使用偏性的多樣性，揭示出選擇在抗病基因的演化中有著重要的作用。按照中性進化理論的觀點，核苷酸序列中密碼子第 3 位鹼基的變化受到突變和自然選擇的影響是中性或近中性的 (Sharp *et al.* 1993)，由此似乎意味著可能由於鹼基突變偏好 G + C 或 A + U 而造成這些 *Xa* 基因之同義密碼子使用偏性。Liu *et al.* (2003) 也指出水稻基因序列 GC 含量強烈影響了密碼子的使用，其密碼子使用偏性之產生主要與轉錄水平上的突變有關。此外，本研究也發現 GC12 之突變也會影響 GC3 的突變偏性，進而造成 *Xa* 基因之同義密碼子使用偏性 (圖 9)。同樣這一特點也存在於其他生物，但其影響程度不同。Kawabe & Miyashita (2003) 利用 4 種單子葉植物 (小麥、大麥、水稻及玉米) 以及 3 種雙子葉植物 (阿拉伯芥、菸草、豌豆) 的 CDS 進行密碼子使用偏性分析，結果發現單子葉與雙子葉植物的密碼子使用偏性皆受到 GC3 組成限制之因素影響，又單子葉植物之 GC3 組成則受到 GC12 之影響，但雙子葉植物則否。Liu & Xue (2005) 利用水稻、玉米、小麥及阿拉伯芥等四種植物的細胞核基因與葉綠體基因進行密碼子使用偏性之比較，發現密碼子使用偏性在葉綠體及細胞核中皆受到 GC3 組成限制影響，而細胞核基因體 GC3 突變偏性又受到 GC12 所影響，葉綠體基因體中 GC3 則否，表示細胞核基因體之 GC 含量會影響突變偏性，而葉綠體基因體則有較低的突變偏性或較高的保守性。

誌 謝

本研究承蒙國家科學委員會 (計畫編號：NSC95-2317-B-055-004) 補助經費，特致謝忱。

引用文獻 (Literature cited)

- Beh, E. J. 2004. A Bibliography of the Theory and Application of Correspondence Analysis. Vol. II-By Publication. School of Quantitative Methods and Mathematical Sciences, Univ. Western Sydney, Australia. 101 pp.
- Benzécri, J. P. 1973. Lanalyse des Donnees. II. Lanalyse des Correspondances. Dunod. Paris, France. 619 pp.
- Berezhnoy, A.Y. and S. A. Duplij. 2005. Dependence of nucleotide physical properties on their placement in codons and determinative degree. J. Zhejiang Univ. Sci. 6:948–960.
- Bulmer, M. 1988. Are codon usage patterns in unicellular organisms determined by selection-mutation balance? J. Mol. Bio. 1:15–26.
- Chiappello, H., F. Lisacek, M. Caboche, and A. Henaut. 1998. Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. Gene 209:GC1–GC38.
- Cameron, J. M., M. Kreiman, and M. Agude. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. Genetics 151:239–249.
- Duret, L. 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. Trends Genet. 16:287–289.
- Duret, L. and D. Mouchiroud. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc. Natl. Acad. Sci. USA 96:4482–4487.
- Fan, S. H., A. G. Guo, L. W. Shan, and X. P. Hu. 2003. Analysis of genetic code preference in *Arabidopsis thaliana*. Prog. Biochem. Biophys. 30:221–225. (in Chinese with English abstract)
- Gouy, M. and C. Gautier. 1982. Codon usage in bacteria correlation with gene expressivity. Nucleic Acids Res. 10:7055–7074.
- Gu, W. J., J. M. Ma, T. Zhou, X. Sun, and Z. H. Lu. 2002. Codon usage in genes coding for proteins with different tertiary structures. Acta Biophys. Sin. 18:81–86. (in Chinese with English abstract)
- Gu, W. J., J. M. Ma, T. Zhou, X. Sun, and Z. H. Lu. 2004. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. BioSystems 73:89–97.
- Guisez, Y., J. Robbins, E. Remaut, and W. Fiers. 1993. Folding of the MS2 coat protein in *Escherichia coli* is modulated by translational pauses resulting from mRNA secondary structure and codon usage: a hypothesis. J. Theor. Biol. 162:243–252.
- Gupta, S. K., S. K. Majumdar, T. Bhattacharya, and T. C. Ghosh. 2000. Studies on the relationships between the synonymous codon usage and protein secondary structural units. Biochem. Biophys. Res. Commun. 269:692–696.
- Gupta, S. K. and T. C. Ghosh. 2001. Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*. Gene 273:63–70.
- Holm, L. 1986. Codon usage and gene expression. Nucleic Acids Res. 14:3075–3087.
- Hou, Z. C. and N. Yang. 2003. Factors affecting codon usage in *Yersinia pestis*. Acta Biochem. Biophys. Sin. 35:580–586.
- Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2:13–34.
- Karlin, S. and J. Mrazek. 1996. What drives codon choices in human genes? J. Mol. Biol. 262:459–472.
- Kawabe, A. and N. T. Miyashita. 2003. Patterns of codon usage bias in three dicot and four monocot plant species. Genes Genet. Syst. 78:343–352.
- Knight, R. D., J. F. Stephen, and L. F. Landweber. 2001. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol. 2:1–13.
- Liu, Q. P., J. Tan, and Q. Z. Xue. 2003. Synonymous codon usage bias in the rice cultivar 93-11 (*Oryza sativa* L. ssp. *Indica*) Acta Genet. Sin. 30:335–340. (in Chinese with English abstract)
- Liu, Q. P., Y. Feng, and Q. Z. Xue. 2004a. Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa*. Mitochondrion 4:313–320.
- Liu, Q. P., Y. Feng, X. Zhao, H. Dong, and Q. Z. Xue. 2004b. Synonymous codon usage bias in *Oryza sativa*. Plant Sci. 167:101–105.
- Liu, Q. P. and Q. Z. Xue. 2005. Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. J. Genet. 84: 55–62.

- Liu, Q., S. Dou, Z. Ji, and Q. Xue. 2005. Synonymous codon usage and gene function are strongly related in *Oryza sativa*. *Biosystems* 80:123–131.
- McVean, G. A. T. and B. Charlesworth. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecularly evolution and variation. *Genetics* 155:929–944.
- Moriyama, E. N. and J. R. Powell. 1998. Gnen length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res.* 26:3188–3193.
- Nakamura, Y. and S. Tabata. 1997. Codon-anticodon assignment and detection of codon usage trends in seven microbial genomes. *Microb. Comp. Genomics* 2:299–312.
- Peden, J. F. 1999. Analysis of Codon Usage. PhD thesis. Dept. Genetics, Univ. Nottingham. 226 pp.
- Peixoto, L., A. Zavala, H. Romero, and H. Musto. 2003. The strength of translational selection for codon usage varies in the three replicons of *Sinorhizobium melioli*. *Gene* 320:109–116.
- Percudani, R., A. Pavesi, and S. Ottonello. 1997. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 268: 322–330.
- Romero, H., A. Zavala, and H. Musto. 2000. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic Acids Res.* 28:2084–2090.
- Romero, H., A. Zavala, H. Musto, and G. Bernardi. 2003. The influence of translational selection on codon usage in fishes from the family *Cyprinidae*. *Gene* 317:141–147.
- Salim, H. M. W. and A. R. O. Cavalcanti. 2008. Factors influencing codon usage bias in genomes. *J. Braz. Chem. Soc.* 19:257–262.
- Sharp, P. M. and W. H. Li. 1986. An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24:28–38.
- Sharp, P. M. and W. H. Li. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15:1281–1295.
- Sharp, P. M., M. Stenico, J. F. Peden, and A. T. Lloyd. 1993. Codon usage: mutational bias, translational selection, or both? *Biochem. J.* 21:835–841.
- Shi, X. F., J. F. Huang, C. R. Liang, S. Q. Liu, J. Xie, and C. Q. Liu. 2001. Is there a close relationship between synonymous codon bias and codon-anticodon binding strength in human genes? *Chinese Sci. Bull.* 12:1015–1019.
- Statsoft Inc. 2002. STATISTICA: The Small Book Chinese Version. USA. 144 pp.
- Stenström, C. M., H. Jin, L. L. Major, W. P. Tate, and L. A. Isaksson. 2001. Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli*. *Gene* 263:273–284.
- Sueoka, N. 1988. Directional mutation pressure and neutral molecular evolution. *Proc. Nat. Acad. Sci. USA* 85:2653–2657.
- Tan, Q., K. Brusgaard, T. A. Kruse, E. Oakeley, B. Hemmings, H. Beck-Nielsen, L. Hansen, and M. Gaster. 2004. Correspondence analysis of microarray time-course data in case-control design. *J. Biomed. Inform.* 37:358–365.
- Tekaia, F., E. Yeramouan, and B. Dujon. 2002. Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. *Gene* 297:51–60.
- Wei, M. L., Y. H. Chung, C. T. Lu, and H. Y. Lu. 2009. Synonymous codon usage bias in rice genes with resistance to bacterial blight (*Xanthomonas oryza*). *J. Taiwan Agric. Res.* 58:155–168. (in Chinese with English abstract)
- Wright, F. 1990. The 'effective number of codons' used in a gene. *Gene.* 87:23–29.

附錄 1. 各種胺基酸之同義密碼子

Appendix 1. Synonymous codons of each amino acid

Amino acid			Amino acid			Amino acid		
Name	Code	Codon	Name	Code	Codon	Name	Code	Codon
Phenylalanine	Phe	UUU	Proline	Pro	CCU	Aspartic acid	Asp	GAU
		UUC			CCC			GAC
Leucine	Leu	UUA	Threonine	Thr	CCA	Glutamic acid	Glu	GAA
		UUG			CCG			GAG
		CUU			ACU			UGU
		CUC			ACC			UGC
Isoleucine	Ile	CUA	Alanine	Ala	ACA	Tryptophan	Trp	UGG
		CUG			ACG			CGU
		AUU			GCU			CGC
		AUC			GCC			CGA
Methionine	Met	AUA	Tryptophan	Tyr	GCA	Glycine	Gly	CGG
		AUG			GCG			AGA
		GUU			UAU			AGG
Valine	Val	GUC	Histidine	His	UAC	Glutamine	Gln	GGU
		GUA			CAU			GGC
		GUG			CAC			GGA
		UCU			CAA			GGG
Serine	Ser	UCC	Asparagine	Asn	CAG	Lysine	Lys	AAA
		UCA			AAU			AAG
		UCG			AAC			
		AGU			AAA			
		AGC			AAG			

Factors Influencing the Synonymous Codon Usage Bias in the Genes for Resistance to Bacterial Leaf Blight (*Xanthomonas oryzae* pv. *oryza*) of Rice (*Oryza sativa*)¹

Meng-Li Wei², Yi-Han Chung², Chun-Tang Lu², and Hsiu-Ying Lu^{3,4}

Abstract

Wei, M. L., Y. H. Chung, C. Y. Lu, and H. Y. Lu. 2010. Factors influencing the synonymous codon usage bias in the genes for resistance to bacterial leaf blight (*Xanthomonas oryzae* pv. *oryza*) of rice (*Oryza sativa*). J. Taiwan Agric. Res. 59:86–102.

Codon usage bias varied among genes (*Xa*) conferring resistance of rice (*Oryza sativa*) to bacterial leaf blight caused by *Xanthomonas oryzae* pv. *oryzae*. This study was conducted to elucidate the evolution of synonymous codon usage by analysis of factors influencing the synonymous codon usage pattern of *Xa* genes. A total of 17 coding domain sequences (CDS) of the completely sequenced genes, *Xa1*, *xa5*, *xa13*, *Xa13*, *Xa21*, *Xa26*, and *Xa27*, were obtained through the access of the GenBank database of National Center for Biotechnology Information (NCBI) and they were used to unravel the main factors influencing codon usage bias. The length of CDS and indices of synonymous codon usage bias such as relative synonymous codon usage (RSCU), codon adaption index (CAI), effective number of codons (ENC), frequency of G or C nucleotides present at the third position of synonymous codons (GC3), and mean frequencies of G or C nucleotides present at the first and second positions of synonymous codons (GC12) were estimated to examine relationships among them. The length of CDS and level of gene expression were highly correlated to codon usage bias in *Xa* genes; the *Xa* genes with shorter CDS (*xa5*, *xa13*, *Xa13*, and *Xa27*) had higher gene expression and higher codon usage bias. The GC3 compositional constraints also affected the codon usage in *Xa* genes. Moreover, the GC12 had a direct effect on GC3 mutation bias, leading to further high codon usage bias in *Xa* genes.

Key words: Synonymous codon usage bias, Gene expression level, GC3 content, Mutation bias, Molecular evolution, Bacterial leaf blight, *Xanthomonas oryzae*.

-
1. Contribution No.2400 from Taiwan Agricultural Research Institute (TARI), Council of Agriculture. Accepted: June 21, 2010.
 2. Respectively, Assistant Researcher, Project Assistant, and Assistant Researcher, Crop Science Division, TARI, Wufeng, Taichung, Taiwan, ROC.
 3. Secretary General, TARI, Wufeng, Taichung, Taiwan, ROC.
 4. Corresponding author, e-mail: iying@tari.gov.tw; Fax: (04)23302801.