

水稻抗白葉枯病基因之密碼子使用偏性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。2009。水稻抗白葉枯病基因之密碼子使用偏性。台灣農業研究 58:155-168。

密碼子使用偏性分析對水稻白葉枯病抗病基因 (*Xa*) 之抗性分子機制之瞭解相當重要，因此本研究自 NCBI (National Center for Biotechnology Information) 公共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) 進行密碼子使用偏性分析。首先進行各 *Xa* 基因之同義密碼子相對使用度 (relative synonymous codon usage, RSCU) 的集群分析和對應分析，可將 *Xa* 基因之 CDS 明顯區分成 4 群。進而再分別計算各群基因之同義密碼子相對使用頻率 (relative frequency of synonymous codon, RFSC) 以篩選出高頻密碼子，結果發現蛋白質序列具有多白胺酸重複 (leucine-rich repeat, LRR) 區域結構且 CDS 長度大於 1000 bp 的 *Xa1*、*Xa21* 及 *Xa26*，其大多數胺基酸並無密碼子使用偏好性，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*，則幾乎所有胺基酸都偏好以 C 或 G 結尾的同義密碼子。*Xa1*、*Xa21*、*Xa26* 及 *Xa27* 的表現蛋白質皆含豐富之白胺酸 (leucine, Leu)，但 *Xa1*、*Xa21* 及 *Xa26* 的 Leu 並無偏好任何密碼子，亦即其表現蛋白質的 Leu 可選擇任意同義密碼子轉譯，而 *Xa27* 的 Leu 則高度偏好使用 CUC。由此顯見，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。

關鍵詞：*Xa* 基因、同義密碼子相對使用度、同義密碼子相對使用頻率、集群分析、對應分析。

前 言

遺傳訊息由 DNA 經 RNA 再到蛋白質的遺傳過程是生物學中心法則，其中由 A、T(U)、

C、G 四種核苷酸以 3 個為一組排列組合而成 1 個胺基酸，即所謂遺傳密碼子 (codon)，共有 64 種組合 ($4^3 = 64$)，但合成細胞中蛋白質的胺基酸只有 20 種，因此有 2 個或 2 個以上轉譯成

1. 行政院農業委員會農業試驗所研究報告第 2361 號。接受日期：98 年 7 月 17 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. 本所技術服務組研究員兼組長。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)23325176。

同一胺基酸之情形，此現象稱為密碼子的簡併性 (degeneracy)，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，這些同義密碼子稱為該生物體或基因的最優密碼子 (optimal codon)，此現象稱為密碼子使用偏性 (codon usage bias)。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)。

蛋白質高級結構由其一級結構 (胺基酸序列) 所決定，即蛋白質摺疊及天然結構所需要之訊息皆蘊含在一級胺基酸序列中，即 Anfinsen 原理 (Anfinsen 1973)。幾乎所有的蛋白質結構預測方法，無論其具體手段如何，事實上都以此為基礎，但多年來利用一級結構來預測高級結構，不斷改進預測方法但成效並不顯著，因此找尋影響蛋白質摺疊於胺基酸序列之外的訊息則更為重要；而基因的編碼區序列 (coding domain sequence, CDS) 可能攜有相關訊息，由於遺傳密碼的簡併性，基因序列較其編碼之胺基酸序列有更多的訊息 (Li & Song 2003)。密碼子是胺基酸的精細結構，蛋白質在核糖體上的合成過程中，由於密碼子和反密碼子 (anticodon) 之特性以及細胞環境之不同，造成同義密碼子轉譯成胺基酸的過程發生差異，這些差異會影響新勝肽鏈摺疊的動力學過程，進而可能影響到蛋白質的最終空間結構，因此研究基因之胺基酸組成與密碼子使用偏性，有利蛋白質高級結構之預測 (Gu *et al.* 2002; Hu & Lin 2005)。

造成同義密碼子使用性 (synonymous codon usage) 之因素多而複雜，其中包含有：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，主要造成密碼子使用偏性的因素為組成突變偏性及自然選擇，而不同的物種也會有不同的影響因素 (Bereznoy & Duplij 2005; Liu & Xue 2005)。Sharp *et al.* (1994) 把密碼子使用偏好性強和弱兩組基因之間同義密碼子相對使用度之差，達到統計學上顯著水準的密碼子定義為最優密碼子，提供了高度表達基因一個表示依據。Duret *et al.* (1999) 則認為，所謂最優密碼子只是隨著基因表達程度的升高而增加其出現頻率的密碼子。一般認為高表達基因的密碼子偏好程度遠大於低表達的基因，這可能是使用主要偏愛密碼子有優化轉譯的效應 (Sharp *et al.* 1993)。密碼子使用偏性上的差異，與蛋白質轉譯速度快慢有關 (Duret & Mouchiroud 1999)。而不同功能的基因，其密碼子使用偏性也存在較大的差異 (Liu *et al.* 2005)。因此，探討基因密碼子使用偏性與基因表達之間的關連，可了解基因在調控表達中對遺傳密碼的選擇性及細胞內蛋白質的轉譯速度，對基因演化規律、轉譯調控機制研究具有重要性。此外，密碼子使用偏性的分析，也能輔助編碼區鑑定、基因選殖體寡核苷酸探針之製備、基因晶片設計等 (Ma *et al.* 2002; Dos *et al.* 2003)。

同義密碼子相對使用度 (relative synonymous codon usage, RSCU) 是常用的同義密碼子使用性之統計評估指標，用以檢測轉譯成同一胺基酸之同義密碼子的出現頻率是否相等：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高，反之則表示出現次數相對較少 (Sharp *et al.* 1986)。利用 RSCU

之計算，可將資料集內的密碼子使用予以常態化，使不受胺基酸組成差異之影響 (Comeron & Aguade 1998)。而 Lin *et al.* (2002) 利用菸草具有高度表達基因的特性，提出了高頻密碼子 (high-frequency codon) 的判定方法，即利用某一密碼子在基因序列中之實際觀測數與該密碼子對應之胺基酸在基因序列中之實際觀測數的比值，計算同義密碼子相對使用頻率 (relative frequency of synonymous codon, RFSC)，作此計算可避免受到基因長短與胺基酸豐富度之影響：當密碼子之 RFSC 值大於 60% 或者超過其對應胺基酸之同義密碼子平均 RFSC 值 1.5 倍時，視為高頻密碼子。

水稻白葉枯病 (rice bacterial 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。而水稻品種之抗病性是由基因控制，目前在 Oryzabase 資料庫 (<http://www.shigen.nig.ac.jp/rice/oryzabase/top.jsp>, Kurata & Yamazaki 2006) 中現已登記命名 32 個水稻抗白葉枯病基因 (*Xa*)，其中已完全定序的有 *Xa1*、*xa5*、*xa13*、*Xa13*、*Xa21*、*Xa26* 及 *Xa27* 七個基因 (Song *et al.* 1995; Yoshimura *et al.* 1995, 1998; Anjali & McCouch 2004; Gu *et al.* 2005; Sun *et al.* 2006)。植物之抗病基因 (resistant gene, R gene) 的表現蛋白質結構分為五類 (Baker *et al.* 1997)：(1) 含有核苷酸結合位置 (nucleotide binding sites, NBS) 的細胞質受體類蛋白 (cytoplasmic receptor-like proteins) 和多白胺酸重複 (leucine-rich repeat, LRR) 序列區域，簡稱為 NBS-LRR；(2) 絲胺酸-酰胺酸激酶 (serine-threonine kinase)；(3) 含有大量細胞質外 LRR 區域的穿膜受體 (transmembrane receptor)；(4) 含細胞外 LRR 區域之穿膜受體和細胞內絲胺酸-酰胺酸激酶 (intracellular serine-threonine kinase) 區域；及 (5) 其他。已知 *Xa1* 為上述第一類 R 基因，而 *Xa21* 和 *Xa26*

為第四類 R 基因，至於 *xa5*、*xa13*、*Xa13* 及 *Xa27* 之抗病防禦系統的生化途徑異於 *Xa1*、*Xa21* 和 *Xa26*，其表現蛋白質結構尚未清楚，目前被歸於其他之第五類 R 基因 (Yoshimura *et al.* 1998; Khush & Angeles 1999; Anjali & McCouch 2004; Gu *et al.* 2005; Sun *et al.* 2006)。為促進 *Xa* 蛋白質序列的結構研究，前作從 NCBI (National Center for Biotechnology Information) 之 GenBank 資料庫 (<http://www.ncbi.nlm.nih.gov/Genbank/index.html>, Guisez *et al.* 1993) 下載已知序列的 7 個 *Xa* 基因共 17 條蛋白質序列為供試材料，探討不同 *Xa* 基因之蛋白質胺基酸組成的變異形式，結果發現不同 *Xa* 基因各有偏好的胺基酸，其中 *Xa1*、*Xa21*、*Xa26* 及 *Xa27* 基因之蛋白質序列中皆以白胺酸 (leucine, Leu) 出現頻率最高，*xa13*、*Xa13* 及 *Xa27* 的丙胺酸 (alanine, Ala) 較其他 *Xa* 基因出現甚多，*xa5* 出現麩胺酸 (glutamic acid, Glu) 和酥胺酸 (threonine, Thr) 的頻率則遠高於其他基因 (Wei *et al.* 2008)。然而，對於 *Xa* 基因轉譯成胺基酸的密碼子使用特徵迄今尚無任何報導。因此，本研究進而利用 NCBI 資料庫內已知序列的 *Xa* 基因之 CDS 為試驗材料，首先估算其 RSCU 並結合多變數分析 (multivariate analysis)，以了解不同 *Xa* 基因之同義密碼子使用性的變異形式，再藉由 RFSC 篩選出高頻密碼子，期能深入剖析 *Xa* 基因之密碼子使用偏性，以有利於水稻的分子育種及抗性機制之研究。

材料與方法

Xa 基因序列之蒐集與下載

本研究自 NCBI 之 GenBank 資料庫下載已知定序之 *Xa* 基因的 CDS，這些已知 *Xa* 基因的相關資訊整理如表 1，其 CDS 長度介於 318 bp–5409 bp 之間，其中 *xa5*、*xa13*、*Xa13* 及 *Xa27* 之 CDS 長度皆小於 1000 bp。*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。

同義密碼子使用性的多變數分析

將上述各 *Xa* 基因之 CDS，利用 CodonW 密碼子分析軟體 (1.4 版，<http://bioweb.pasteurur.fr/seqana/interfaces/codonw.html>, Peden 1999)，先求出密碼子個數，再計算 RSCU，其計算式為 (Sharp *et al.* 1986)：

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

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頻率都相等) 為高，反之則表示出現次數相對較少 (Sharp *et al.* 1986)。

同義密碼子使用性是由多個物種或基因 (觀測值) 與各同義密碼子出現頻率 (變數) 所構成的資料矩陣來決定，這種多維資料形式所含之訊息，最適合利用多變數分析之統計技術來解析。多變數分析是一種綜合分析方法，能夠在多個研究對象和多個指標相互關聯的情況下分析出它們的統計規律。集群分析 (cluster analysis) 是多變數分析方法的一種，係利用一般邏輯程序，能根據多個變數間或研究對象的

相似性與相異性，客觀地將相似者歸在同一群內，集群分析之結果最後以樹狀圖 (dendrogram) 的分支狀況來判斷群數 (Leps & Smilauer 1999)。對應分析 (correspondence analysis) 為 Benzecri 於 1973 年所提出，用於分析簡單二維及多維表格的行與列之間的對應關係，可將多維之變數或樣本於一個低維空間上以一個序列分佈圖 (ordination diagram) 呈現，從各變數或樣本落在序列分佈圖上之距離遠近，可用來輔助集群分析之樹狀圖的分群結果判斷，倘發現點間有群團現象表示可能是同一群。

為探討不同 *Xa* 基因間密碼子使用性的相似及相異程度，本研究在不考慮 Met、Trp 及終止密碼子 (UGA、UAG、UAA) 共 5 個無意義密碼子 (non-sense codon) 下，將所計算出各基因之 59 種同義密碼子之 RSCU 值，利用 STATISTICA 統計軟體 (7.0 版，Statsoft Inc. 2002) 以非加權配對算術平均法 (unweighted pair group method using arithmetic average, UPGMA) 進行集群分析，以求得歸群樹狀圖。UPGMA 是集群分析的演算方法之一，以群間個體所有成對距離的平均值為判斷，較不受離群值影響 (Mohna *et al.* 1992; Mumm *et al.* 1994)。接著，同樣再根據 RSCU 值利用 STATISTICA 統計軟體進行對應分析，由其行列之間的對應關係，求得序列分佈圖，以呈現基因間密碼子使用性之關係。最後綜合集群分析之樹狀圖和對應分析之序列分佈圖的結果，來共同決定群數。

高頻密碼子之判定

根據 RFSC 可用來篩選基因的高頻密碼子，其計算式為 (Lin *et al.* 2002)：

$$RFSC_{ij} = \frac{X_{ij}}{\sum_{j=1}^{n_i} X_{ij}} \times 100\%$$

X_{ij} 是第 i 個胺基酸中第 j 個密碼子出現的次

表 1. 已知序列之水稻 *Xa* 基因相關資訊

Table 1. Information of rice *Xa* genes with known sequences ^z

Label	Gene symbol	Submitted year	Accession no. in NCBI	Definition	Gene length	Chr. no.	CDS ^y	
							Location	Length
<i>Xa1-1999</i>	<i>Xa1</i>	1999	AB002266	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) mRNA for <i>Xa1</i>	5910	4	113..5521	5409
<i>Xa1-2004</i>	<i>Xa1</i>	2004	NM_194685	<i>Oryza sativa</i> (<i>japonica</i> cultivar-group) putative bacterial blight resistance protein <i>Xa1</i> -like protein, mRNA	4206	10	1..4206	4206
<i>xa5</i>	<i>xa5</i>	2004	AY643716	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) transcription factor IIA gamma subunit (TFIIA _γ) mRNA, partial cds.	318	5	1..318	318
<i>xa13</i>	<i>xa13</i>	2006	DQ421394	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) cultivar IRBB13 disease resistant allele <i>xa13</i> (<i>xa13</i>) gene, complete cds.	924	8	1..924	924
<i>Xa13</i>	<i>Xa13</i>	2006	DQ421395	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) cultivar IR64 disease resistant allele <i>Xa13</i> (<i>Xa13</i>) gene, complete cds.	924	8	1..924	924
<i>Xa21-A1</i>	<i>Xa21</i>	1997	U72725	<i>Oryza longistaminata</i> receptor kinase-like protein gene	8416	11	4771..7384, 7676..8052	2991
<i>Xa21-A2</i>	<i>Xa21</i>	1998	U72727	<i>Oryza longistaminata</i> receptor kinase-like protein, pseudogene sequence	5940	11	2151..4764, 4872..5248	2991
<i>Xa21-B</i>	<i>Xa21</i>	1995	U37133	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) receptor kinase-like protein gene	3921	11	1..2677, 3521..3921	3078
<i>Xa21-C</i>	<i>Xa21</i>	1998	U72723	<i>Oryza longistaminata</i> receptor kinase-like protein gene, pseudogene	19639	11	15118..17720, 17827..18201	2978
<i>Xa21-D</i>	<i>Xa21</i>	1997	U72726	<i>Oryza longistaminata</i> receptor kinase-like protein	13341	11	2367..4205	1839
<i>Xa21-E</i>	<i>Xa21</i>	1997	U72724	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) receptor kinase-like protein gene	9424	11	2819..5260	2442
<i>Xa21-F</i>	<i>Xa21</i>	2005	U72728	<i>Oryza longistaminata</i> receptor-like kinase protein, pseudogene sequence	7204	11	1683..4352, 5147..5547	3071
<i>Xa26a</i>	<i>Xa26</i>	2006	DQ355952	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) receptor kinase MRK _a	3401	11	14225..17135, 17240..17625	3297
<i>Xa26b</i>	<i>Xa26</i>	2006	DQ355952	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) bacterial blight resistance protein <i>Xa26</i>	3631	11	21725..24659, 24765..25141	3312
<i>Xa26c</i>	<i>Xa26</i>	2006	DQ355952	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) receptor kinase MRK _c	3456	11	34496..34911, 35026..37951	3342
<i>Xa26d</i>	<i>Xa26</i>	2006	DQ355952	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) cultivar Minghui 63 receptor kinase pseudogene	11396	11	2021..3423, 9305..10839, 13028..13416	3327
<i>Xa27</i>	<i>Xa27</i>	2005	AY986491	<i>Oryza sativa</i> (<i>indica</i> cultivar-group) <i>Xa27</i> gene, <i>Xa27</i> -IR24 allele	2393	6	1590..1931	342

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

^y CDS: coding domain sequence.

數， n_i 為第 i 個胺基酸同義密碼子的數目。若某一密碼子之 RFSC 值大於 60% 或者超過其對應胺基酸之同義密碼子平均 RFSC 值 1.5 倍，即視為高頻密碼子。由於 CodonW 分析軟體並未提供 RFSC 之計算，本研究是利用 Excel 軟體編寫公式而求得各密碼子之 RFSC 值及判斷高頻密碼子。

結 果

Xa 基因同義密碼子使用性之多變數分析

根據 17 個 *Xa* 基因及其家族之 CDS 的各 59 種同義密碼子之 RSCU 值進行集群分析及對應分析之結果，列如圖 1。由集群分析之樹狀圖顯示 (圖 1A)，在歐氏距離 (Euclidean distance) 7.23 時，*Xa* 基因及其家族根據 RSCU 值可被分成 4 群：*Xa1-1999*、*Xa1-2004*、*Xa21-C*、*Xa26a*、*Xa26b*、*Xa26c*、*Xa21-F*、*Xa26d*、*Xa21-A1*、*Xa21-A2*、*Xa21-E*、*Xa21-B* 及 *Xa21-D* 為同一群，*xa13* 和 *Xa13* 自成一類，*xa5* 和 *Xa27* 則各自成群。而對 17 個 *Xa* 基因及其家族的 RSCU 值進行對應分析，其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%，說明存在影響其密碼子使用頻率的主要趨勢，於是由圖 1B 之二維序列分佈圖可知，*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*，即對應分析與集群分析所獲分群結果相當一致，17 個 *Xa* 基因及其家族根據同義密碼子使用性之 RSCU 值大致可被區分成 4 群：第 I 群為 *Xa1-1999*、*Xa1-2004*、*Xa21-C*、*Xa26a*、*Xa26b*、*Xa26c*、*Xa21-F*、*Xa26d*、*Xa21-A1*、*Xa21-A2*、*Xa21-E*、*Xa21-D* 及 *Xa21-B*，第 II 群為 *xa5*，第 III 群為 *xa13* 及 *Xa13*，以及第 IV 群為 *Xa27*。

Xa 基因高頻密碼子之判定

根據上述 RSCU 值分群結果，接著分別計算各群基因之 RFSC，並依照其密碼子之 RFSC

值大於 60% 或者超過其對應胺基酸之同義密碼子平均 RFSC 值 1.5 倍的原則，找出各群基因之高頻密碼子，結果如表 2 所示：在不考慮 Met、Trp 不具同義密碼子外，第 I 群基因 (*Xa1-1999*、*Xa1-2004*、*Xa21-C*、*Xa26a*、*Xa26b*、*Xa26c*、*Xa21-F*、*Xa26d*、*Xa21-A1*、*Xa21-A2*、*Xa21-E*、*Xa21-D* 及 *Xa21-B*) 僅 6 種胺基酸 (Pro、Thr、Ala、Asp、Cys、Arg) 具有高頻密碼子，第 II 群基因 (*xa5*) 除 Thr 和 Asp 外其餘 16 種胺基酸皆有高頻密碼子，第 III 群基因 (*xa13*、*Xa13*) 之所有胺基酸存在高頻密碼子，而第 IV 群基因 (*Xa27*) 除 Asp 不具特定密碼子偏性外其餘 16 種胺基酸都具有高頻密碼子 (*Xa27* 之 CDS 轉譯之蛋白質序列中不含 Lys)；由此顯示 *xa5*、*xa13*、*Xa13* 及 *Xa27* 基因表現蛋白之胺基酸大都具有特定偏好的密碼子之特徵，而且其高頻密碼子幾乎都以 C 或 G 結尾。針對不同 *Xa* 基因各自偏好的幾個主要胺基酸 (Wei *et al.* 2008) 而言，由表 2 可看出，*Xa1*、*Xa21* 及 *Xa26* 之表現蛋白質中的 Leu 並無偏好任何密碼子，而 *xa5*、*xa13*、*Xa13* 及 *Xa27* 的 Leu 皆高度偏好使用 CUC；所有 *Xa* 基因在 Ala 都具有高頻密碼子，第 I 群基因之 *Xa1*、*Xa21* 及 *Xa26* 的 Ala 較偏好 GCA，而 *xa5*、*xa13*、*Xa13* 及 *Xa27* 的 Ala 都偏好使用 GCC；此外，*xa5*、*xa13*、*Xa13* 及 *Xa27* 在 Glu 也都偏好 GAG，但僅有 *xa13*、*Xa13* 及 *Xa27* 的 Thr 偏好 ACC，而 *xa13* 及 *Xa13* 的 Thr 除偏好 ACC 外也偏好 ACG。

由於 *Xa1*、*Xa21*、*Xa26* 分屬於不同類別的 R 基因，因此在第 I 群基因內再進而分別檢視這 3 個基因各胺基酸之同義密碼子的 RFSC 值 (表 3)，結果顯示這 3 個基因在 Ala 都偏好 GCA，除此第一類 R 基因的 *Xa1* 在 Thr 及 His 也各自偏好 ACA 及 CAU；同屬第四類 R 基因的 *Xa21* 和 *Xa26* 在 Pro、Ala、Asp 及 Arg 都一致具有高頻密碼子，其 Pro、Ala 及 Asp 都同樣各偏好

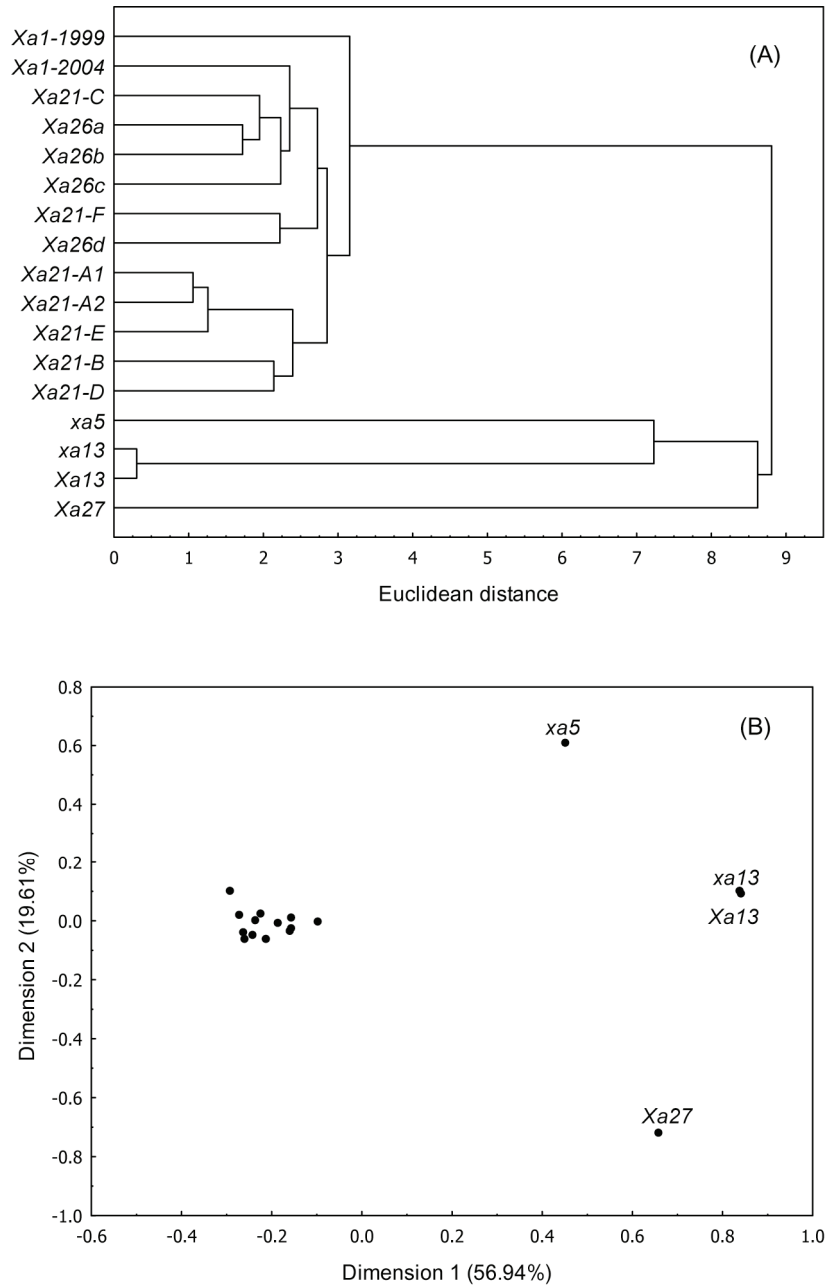


圖 1. 17 個水稻 *Xa* 基因及其家族之 CDS 序列根據同義密碼子相對使用度 (RSCU) 的分群結果。(A) 集群分析採 UPGMA 法之樹狀圖，(B) 對應分析之序列分布圖。

Fig. 1. Grouping results of 17 CDSs of rice *Xa* genes and their families according to relative synonymous codon usage (RSCU). (A) Dendrogram of cluster analysis using UPGMA, (B) Ordination diagram of correspondence analysis.

表 2. 水稻 *Xa* 基因根據同義密碼子偏好性相對使用頻度 (RFSC) 所判定之高頻密碼子
 Table 2. High-frequency codon of rice *Xa* genes identified from relative frequency of synonymous codon (RFSC)

Amino acid	Codon	I ^z	II	III	IV	Amino acid	Codon	I	II	III	IV
Phe	UUU	42.8	20.0	0.0	0.0	Ala	GCU	28.1	0.0	2.4	11.1
	UUC	57.2	80.0 ^y	100.0	100.0		GCC	22.6	80.0	72.3	61.1
	UUA	10.3	0.0	0.0	0.0		GCA	38.0	20.0	7.2	5.6
Leu	UUG	17.7	18.2	5.9	0.0	Tyr	GCG	11.4	0.0	18.1	22.2
	CUU	19.3	9.1	2.9	4.0		UAU	49.8	0.0	3.8	0.0
	CUC	19.2	54.5	64.7	68.0		UAC	50.2	100.0	96.2	100.0
Ile	CUA	11.1	9.1	0.0	8.0	His	CAU	55.8	0.0	0.0	0.0
	CUG	22.4	9.1	26.5	20.0		CAC	44.2	100.0	100.0	100.0
	AUU	34.9	60.0	0.0	50.0		CAA	53.1	80.0	0.0	33.3
Val	AUC	36.5	40.0	100.0	50.0	Gln	CAG	46.9	20.0	100.0	66.7
	AUA	28.6	0.0	0.0	0.0		AAU	50.8	33.3	0.0	0.0
	GUU	29.9	37.5	2.6	14.0		AAC	49.2	66.7	100.0	100.0
Ser	GUC	19.0	25.0	47.4	86.0	Lys	AAA	51.0	12.5	0.0	0.0
	GUA	18.1	12.5	0.0	0.0		AAG	49.0	87.5	100.0	0.0
	GUG	33.0	25.0	50.0	0.0		GAU	62.1	50.0	0.0	50.0
Pro	UCU	21.2	20.0	0.0	20.0	Asp	GAC	37.9	50.0	100.0	50.0
	UCC	17.1	30.0	50.0	40.0		GAA	54.1	27.3	0.0	33.3
	UCA	19.3	10.0	0.0	0.0		GAG	45.9	72.7	100.0	66.7
Thr	UCG	11.4	0.0	18.2	40.0	Cys	UGU	39.2	66.7	0.0	0.0
	AGU	12.3	0.0	0.0	0.0		UGC	60.8	33.3	100.0	100.0
	AGC	18.7	40.0	31.8	0.0		CGU	11.5	0.0	0.0	0.0
Phe	CCU	27.9	0.0	5.6	50.0	Arg	CGC	12.6	0.0	37.5	75.0
	CCC	14.4	0.0	27.8	50.0		CGA	11.2	0.0	0.0	25.0
	CCA	43.7	0.0	5.5	0.0		CGG	11.8	33.3	12.5	0.0
Thr	CCG	14.0	100.0	61.1	0.0	Gly	AGA	26.0	0.0	0.0	0.0
	ACU	42.4	36.3	0.0	25.0		AGG	26.8	66.7	50.0	0.0
	ACC	20.5	27.3	56.8	75.0		GGU	30.0	0.0	11.8	25.0
Pro	ACA	26.9	18.2	0.0	0.0	Gly	GGC	27.8	100.0	52.9	66.7
	ACG	10.2	18.2	43.2	0.0		GGA	25.2	0.0	17.7	8.3
							GGG	17.0	0.0	17.6	0.0

^z Gene group: I, *Xa1-1999*, *Xa21-A1*, *Xa21-A2*, *Xa21-B*, *Xa21-D*, *Xa21-E*; *Xa1-2004*, *Xa21-C*, *Xa21-F*, *Xa26a*, *Xa26b*, *Xa26c*, *Xa26d*; II, *xa5*; III, *xa13* and *xa13*; IV, *Xa27*.
^y High-frequency codon identified from RFSC (bold and underline).

CCA、GCA 及 GAU，不過 *Xa21* 及 *Xa26* 在 Arg 的高頻密碼子並不相同；此外，*Xa21* 的 Thr、Glu 及 Cys 也都具有高頻密碼子，但 *Xa26* 則無。從表 3 可發現，顯然 *Xa1*、*Xa21* 及 *Xa26* 僅少數胺基酸具有高頻密碼子，而這些胺基酸幾乎偏好以 A 或 U 結尾的密碼子。

討 論

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)，本研究利用 RSCU 之多變數分析進行 *Xa* 基因同義密碼子使用性分析，可大致將 *Xa* 基因分為 4 群 (圖 1)，據此進而利用 RFSC 找出各群 *Xa* 基因表現蛋白之胺基酸的高頻密碼子 (表 2)。綜合各項結果顯示，CDS 長度大於 1000 bp 的 *Xa* 基因 (第 I 群之 *Xa1-1999*、*Xa21-A1*、*Xa21-A2*、*Xa21-B*、*Xa21-D*、*Xa21-E*、*Xa1-2004*、*Xa21-C*、*Xa21-F*、*Xa26a*、*Xa26b*、*Xa26c*、*Xa26d*) 只有少數胺基酸具有高頻密碼子，即大多數胺基酸並無明顯偏好任何密碼子。而 CDS 長度較短 (318–924 bp) 的 *Xa* 基因 (第 II、III 及 IV 群基因的 *xa5*、*xa13*、*Xa13* 及 *Xa27*)，絕大多數的胺基酸都有高頻密碼子，顯示這些基因表現蛋白之胺基酸具有高度的密碼子偏好性。此發現與 Liu *et al.* (2003) 利用水稻品種 93-11 為材料探討基因長度對密碼子使用偏性之影響研究所得結果一致，即水稻中序列較短之基因的密碼子偏好性較大。此外，Liu (2006) 進行輻射菌 (*Deinococcus radiodurans*) 基因密碼子偏性分析，也發現基因之 CDS 長度為影響基因密碼子使用性之主要因素之一。

密碼子使用偏性會直接影響到基因轉錄的效率，進而決定 mRNA 的濃度 (Xia 1996)。Coghlan *et al.* (2000) 研究酵母菌 (*Saccharomyces cerevisiae*) 的密碼子使用偏性與 mRNA 的濃度及蛋白質的相關性時，也發現密碼子使用偏性可能與增加轉錄及轉譯的效率

(如提高準確度) 及減少轉譯過程中能量消耗有關。*Xa1*、*Xa21* 及 *Xa26* 基因的表現蛋白質皆含有多白胺酸重複 (leucine-rich repeat, LRR) 序列區域結構 (Song *et al.* 1995; Yoshimura *et al.* 1998; Sun *et al.* 2006)，Wei *et al.* (2008) 指出 *Xa* 基因之表現蛋白質中的 Leu 為水稻具有白葉枯病抗性之重要胺基酸，*Xa1*、*Xa21* 及 *Xa26* 的表現蛋白質序列皆含有豐富的 Leu。而本研究發現 *Xa1*、*Xa21* 及 *Xa26* 在 Leu 並無偏好任何密碼子 (表 2、3)，顯示這 3 個 *Xa* 基因表現蛋白質的 Leu 可任意選擇其同義密碼子轉譯，此結果有助於穩定 *Xa* 基因之蛋白質結構，可不受突變與外力而影響其抗性。相較於其他 *Xa* 基因，*Xa1*、*Xa21* 及 *Xa26* 僅有少數胺基酸具有高頻密碼子，且大多偏向選擇 A 或 U 結尾的密碼子。第一類 R 基因的 *Xa1* 與第四類 R 基因的 *Xa21* 及 *Xa26* 僅在 Ala 都偏好 GCA，在其他胺基酸則密碼子偏好性不盡相同；而同屬於第四類 R 基因的 *Xa21* 及 *Xa26* 在某些胺基酸的密碼子偏好特性仍多少有異。

至於 CDS 長度較短的 *xa5*、*xa13*、*Xa13* 及 *Xa27* 基因都不具 LRR 蛋白質結構 (Khush & Angeles 1999; Anjali & McCouch 2004; Gu *et al.* 2005; Sun *et al.* 2006)，且其胺基酸組成偏好性明顯不同 (Wei *et al.* 2008)，由本研究結果發現這些基因幾乎所有胺基酸都特別偏好以 C 或 G 結尾的同義密碼子 (表 2)，表示 *xa5*、*xa13*、*Xa13* 及 *Xa27* 基因編碼區的鹼基組成 (G + C 含量) 強烈影響其同義密碼子。*Xa27* 蛋白質序列中有非常多的 Leu 且含量遠高於其他 *Xa* 基因 (Wei *et al.* 2008)，按可能與 *Xa27* 之產物內含 3 個核定位信號 (nuclear localized signal, NLS) 模組之 C 端保守區域和一個轉錄活化區的細胞核局部性第 III 型作用器有關 (Gu *et al.* 2005)，由本研究發現 *Xa27* 的 Leu 高度偏好使用 CUC (表 2)，此與 *Xa1*、*Xa21* 及 *Xa26* 可由任何密碼子轉譯成 Leu 的結果完全不同 (表 2、3)。

表 3. 水稻 *Xa1*, *Xa21*, *Xa26* 基因根據同義密碼子偏好性相對使用頻度 (RFSC) 所判定之高頻密碼子
Table 3. High-frequency codon of rice *Xa1*, *Xa21*, *Xa26* genes identified from relative frequency of synonymous codon (RFSC)

Amino acid	Codon	<i>Xa1</i>	<i>Xa21</i>	<i>Xa26</i>	Amino acid	Codon	<i>Xa1</i>	<i>Xa21</i>	<i>Xa26</i>
Phe	UUU	45.5	41.1	44.0	Ala	GCU	24.4	27.5	31.1
	UUC	54.5	58.9	56.0		GCC	23.2	24.1	20.5
	UUA	10.5	10.0	10.5		GCA	37.8	38.6	37.4
Leu	UUG	17.1	17.6	18.3	Tyr	GCG	14.6	9.8	11.0
	CUU	20.8	19.5	18.2		UAU	60.5	42.1	51.0
	CUC	21.4	19.4	17.4		UAC	39.5	57.9	49.0
Ile	CUA	8.2	9.8	14.9	His	CAU	60.9	53.5	55.0
	CUG	22.0	23.7	20.6		CAC	39.1	46.5	45.0
	AUU	33.1	36.4	33.9		CAA	52.2	54.4	51.9
Val	AUC	40.6	36.1	34.7	Asn	CAG	47.8	45.6	48.1
	AUA	26.3	27.5	31.4		AAU	57.0	54.7	42.5
	GUU	22.3	33.3	30.9		AAC	43.0	45.3	57.5
Ser	GUC	19.9	18.4	19.1	Lys	AAA	47.1	54.5	50.0
	GUA	24.7	17.3	14.4		AAG	52.9	45.5	50.0
	GUG	33.1	31.0	35.6		GAU	57.3	66.2	61.5
Pro	UCU	22.7	19.9	22.3	Asp	GAC	42.7	33.8	38.5
	UCC	16.4	19.0	14.7		GAA	46.9	64.6	46.5
	UCA	20.1	18.3	20.6		GAG	53.1	35.4	53.5
Thr	UCG	5.4	15.1	9.2	Glu	UGU	32.6	38.4	48.1
	AGU	18.1	10.5	11.4		UGC	67.4	61.6	51.9
	AGC	17.4	17.2	21.9		CGU	12.6	11.3	10.8
Cys	CCU	27.3	30.9	24.4	Arg	CGC	9.3	14.7	12.6
	CCC	20.2	13.0	13.8		CGA	8.2	10.9	15.0
	CCA	35.4	42.1 ^z	49.3		CGG	12.6	9.6	15.0
Gly	CCG	17.2	14.0	12.4	Gly	AGA	22.0	32.1	19.8
	ACU	31.5	26.2	32.4		AGG	35.2	21.5	26.9
	ACC	24.8	21.5	30.4		GGU	23.4	33.0	29.0
Gly	ACA	33.9	39.1	24.0	Gly	GGC	26.3	29.0	26.7
	ACG	9.7	13.2	13.2		GGA	26.9	22.4	28.7
						GGG	23.4	15.6	15.6

^z High-frequency codon indentified from RFSC (bold and underline).

由此顯見，同樣對白葉枯病菌產生抗性的不同 *Xa* 基因間所呈現的胺基酸組成及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。蛋白質序列具有 LRR 區域結構之 *Xa1*、*Xa21* 及 *Xa26* 的 CDS 長度較長，僅少數胺基酸具有高頻密碼子，且大多偏好以 A 或 U 結尾的密碼子；而 *xa5*、*xa13*、*Xa13* 及 *Xa27* 之蛋白質序列不具 LRR 區域結構且 CDS 較短，其幾乎所有胺基酸都偏好以 C 或 G 結尾的密碼子。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。近來也已有研究發現水稻的 G + C 含量決定了密碼子的使用，其密碼子使用偏性之產生主要與轉錄水平上的突變有關 (Liu *et al.* 2003)。同義密碼子使用偏性究竟是由突變偏性或轉譯選擇產生，還是兩者共同作用產生，迄今結論不甚一致 (Salim & Cavalcanti 2008)，未來將值得進一步探討 *Xa* 基因密碼子使用偏性之影響因素，以有利 *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。

誌 謝

本研究承蒙行政院農業委員會農業生物技術國家型科技計畫 [計畫編號：94 農科-5.2.1-農-C1(8)] 及國家科學委員會 (計畫編號：NSC95-2317-B-055-004) 補助經費，特致謝忱。

引用文獻 (Literature cited)

- Anfinsen, C. B. 1973. Principal that govern the folding of protein chains. *Science* 181:223–230.
- Anjali, S. I. and S. R. McCouch. 2004. The rice bacterial blight resistance gene *xa5* encodes a novel form of disease resistance. *Mol. Plant-Microbe Interact.* 17:1348–1354.
- Baker, B., P. Zambryski, B. Staskawicz, and S. P. Dinesh-Kumar. 1997. Signaling in plant-microbe interactions. *Science* 276:726–733.
- Benzecri, 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.
- Coghlan, A. and K. H. Wolfe. 2000. Relationship of codon bias to mRNA concentration and protein length in *Saccharomyces cerevisiae*. *Yeast* 16:1131–1145.
- Comeron, J. and M. Aguade. 1998. An evaluation of measure of synonymous codon usage bias. *J. Mol. Evol.* 47:268–274.
- Dos, R. M., L. Wernisch, and R. Savva. 2003. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.* 31:6976–6985.
- 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)
- 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. Sinica.* 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.
- Gu, K., B. Yang, D. Tian, L. Wu, D. Wang, C. Sreekala, F. Yang, Z. Chu, G. L. Wang, F. F. White, and Z. Yin. 2005. R gene expression induced by a type-III effector triggers disease resistance in rice. *Nature* 435:1122–1125.

- 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.
- Hou, Z. C. and N. Yang. 2003. Factors affecting codon usage in *Yersinia pestis*. *Acta Biochem. Biophys. Sinica* 35:580–586.
- Hu, X. Z. and Q. Z. Li. 2005. The further discussion for context correlation of codons in secondary structures of proteins. *Acta Scientiarum Naturalium Univ. NeMongol.* 36:294–300. (in Chinese with English abstract)
- Karlin, S. and J. Mrazek. 1996. What drives codon choices in human genes? *J. Mol. Biol.* 262:459–472.
- Khush, G. S. and E. R. Angeles. 1999. A new gene for resistance to race 6 of bacterial blight in rice, *Oryza sativa* L. *Rice Genet. Newsl.* 16:92–93.
- 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.
- Kurata, N. and Y. Yamazaki. 2006. Oryzabase. An integrated biological and genome information database for rice. *Plant Physiol.* 140:12–17.
- Leps, J. and P. Smilauer. 1999. *Multivariate Analysis of Ecological Data*. Faculty of Biological Science. Univ. South Bohemia, Ceske Budejovice. 110 pp.
- Li, W. J. and J. N. Song. 2003. How much protein secondary structural information do synonymous codons carry? *Acta Biophys. Sinica* 19:41–46. (in Chinese with English abstract)
- Lin, T., Z. H. Ni, M. S. Shen, and L. Chen. 2002. High-frequency codon analysis and its application in codon analysis of tobacco. *J. Xiamen Univ. (Nat. Sci.)* 5:551–554. (in Chinese with English abstract)
- 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. 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.
- Liu, Q. P. 2006. Analysis of codon usage pattern in the radioresistant bacterium *Deinococcus radiodurans*. *BioSystems* 85:99–106.
- Ma, J. M., W. J. Gu, T. Zhou, X. Sun, Z. H. Lu. 2002. Method for the design of mRNA detecting genechip based on codon usage bias. *J. Southeast Univ. (Nat. Sci. Ed.)* 32:581–585. (in Chinese with English abstract)
- Mohna, F. I., P. Shen, S. C. Jong, and K. Orikono. 1992. Molecular evidence supports the separation of *Lentinula edodes* from *Lentinus* and related genera. *Can. J. Bot.* 70:2446–2452.
- 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.
- Mumm, R. H., J. Hubert, and J. W. Dudley. 1994. A classification of 148 U.S. maize inbreds: II. Validation of cluster analysis based on RFLPs. *Crop Sci.* 34:852–865.
- Peden, J. F. 1999. *Analysis of Codon Usage*. PhD thesis. Depart. 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., M. Stenico, J. F. Peden, A. T. Lloyd. 1993. Codon usage: mutational bias, translational selection, or both? *Biochem. J.* 21:835–841.
- Sharp, P. M., M. Stenico, and A. T. Lloyd. 1994. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res.* 22:2437–2446.
- 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.
- Song, W. Y., G. L. Wang, L. L. Chen, H. S. Kim, L. Y. Pi, T. Holsten, J. Gardner, B. Wang, W. X. Zhai, L. H. Zhu, C. Frauquet, and P. Ronald. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* 270:1804–1806.
- 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.
- Sun, X. L., Y. L. Cao, and S. P. Wang. 2006. Point mutations with positive selection were a major force during the evolution of a receptor-kinase resistance gene family of rice. *Plant Physiol.* 140:998–1008.
- Wei, M. L., Y. H. Chung, C. T. Lu, and H. Y. Lu. 2008. Multivariate analysis of amino acid composition in proteins from rice bacterial blight resistance genes. *J. Taiwan Agric. Res.* 57:161–174. (in Chinese with English abstract)
- Xia, X. 1996. Maximizing transcription efficiency causes codon usage bias. *Genetics* 144:1309–1320.
- Yoshimura, S., A. Yoshimura, N. Iwata, S. R. McCouch, M. L. Abenes, M. Baraoidan, T. W. Mew, and R. J. Nelson. 1995. Tagging and combining bacterial blight resistance genes in rice using RAPD and RFLP markers. *Mol. Breed.* 1:375–387.
- Yoshimura, S., U. Yamanouchi, Y. Katayose, S. Toki, Z. X. Wang, I. Kono, N. Kurta, M. Yano, N. Iwata, and T. Sasaki. 1998. Expression of *Xa1*, a bacterial blight-resistance gene in rice, is induced by bacterial inoculation. *Proc. Nat. Acad. Sci.* 95: 1663–1668.

Synonymous Condon Usage Bias in Rice Genes with Resistance to Bacterial Blight (*Xanthomonas oryza*)¹

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

Abstract

Wei, M. L., Y. H. Chung, C. T. Lu, and H. Y. Lu. 2009. Synonymous condon usage bias in rice genes with resistance to bacterial blight (*Xanthomonas oryza*). J. Taiwan Agric. Res. 58:155–168.

Analysis of codon usage is important in understanding the molecular mechanism of rice genes with resistance to bacterial blight caused by *Xanthomonas oryzae* pv. *oryzae*. A total of 17 coding domain sequences (CDSs) of the completely sequenced genes, i.e., *Xa1*, *xa5*, *xa13*, *Xa13*, *Xa21*, *Xa26*, and *Xa27* were obtained through the access from public database of NCBI (National Center for Biotechnology Information) and they were used to analyze the condon usage bias. The use of correspondence analysis and cluster analysis of relative synonymous condon usage (RSCU) clearly separated the CDSs of *Xa* genes into four distinct groups. The relative frequency of synonymous condon (RFSC) was further used to indentify the high-frequency codons for each gene group. The results showed that *Xa1*, *Xa21*, and *Xa26* with leucine-rich repeat (LRR) domain protein structure and CDS length more than 1000 bp were found no condon usages in most amino acids, and preferred A- or U-ending codons in a few amino acids; whereas *xa5*, *xa13*, *Xa13*, and *Xa27* with non-LRR domain protein structure and shorter CDSs preferred C- or G-ending codons in almost all amino acids. The protein sequences of *Xa1*, *Xa21*, *Xa26*, and *Xa27* has been known to have abundant leucine (Leu). The *Xa1*, *Xa21*, and *Xa26* showed no significant condon bias in their Leu, that is they could choose any synonymous condon for Leu when translating to protein; whereas, *Xa27* exhibited CUC as the most preferred condon in Leu. It indicated that condon usage bias varied among rice genes with resistance to bacterial blight.

Key words: *Xa* genes, Relative synonymous condon usage, Relative frequency of synonymous condon, Cluster analysis, Correspondence analysis.

-
1. Contribution No.2361 from Taiwan Agricultural Research Institute (TARI), Council of Agriculture. Accepted: July 17, 2009.
 2. Respectively, Assistant Researcher, Project Assistant, and Assistant Researcher, Crop Science Division, TARI, Wufeng, Taichung, Taiwan, ROC.
 3. Senior Researcher and Director, Technical Service Division, TARI, Wufeng, Taichung, Taiwan, ROC.
 4. Corresponding author, e-mail: iying@tari.gov.tw; Fax: (04)23325176.