

# 水稻抗白葉枯病基因表現蛋白質之多白胺酸重複 (LRR) 的分布及胺基酸組成<sup>1</sup>

林奕承<sup>2</sup> 呂椿棠<sup>3</sup> 魏夢麗<sup>3</sup> 呂秀英<sup>4,5</sup>

## 摘 要

林奕承、呂椿棠、魏夢麗、呂秀英。2009。水稻抗白葉枯病基因表現蛋白質之多白胺酸重複 (LRR) 的分布及胺基酸組成。台灣農業研究 58:93-105。

目前已知多白胺酸重複 (leucine-rich repeat, LRR) 為生物體抗病基因蛋白質序列上的特有序列,因此本研究針對水稻所有已定序之 *Xa* 基因,自 NCBI (National Center for Biotechnology Information) 及 KOME (Knowledge-based Oryza Molecular Biological Encyclopedia) 公共資料庫內擷取出它們已知的 CDS (coding domain sequence) 與全長 cDNA 序列,將其轉譯的蛋白質序列利用 Pfam 蛋白質家族資料庫尋找出各 LRR 片段,發現僅在 *Xa1*、*Xa21* 及 *Xa26* 基因之表現蛋白質結構中發現 LRR 存在。進而利用生物資訊與統計方法,進行不同 *Xa* 基因蛋白質之 LRR 序列的分布及胺基酸組成之比較分析。結果顯示,*Xa1* 之 LRR 在蛋白質序列中鬆散分布於 NB-ARC (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4) 區域序列之後,而 *Xa21* 及 *Xa26* 之 LRR 則密集分布在 NB-ARC 或 P-kinase (protein-kinase) 區域序列之前。*Xa1* 的蛋白質序列中靠近 NB-ARC 區域序列的前一、二個 LRR 似乎較遠離其他 LRR,但 *Xa21* 和 *Xa26* 的蛋白質之 LRR 區域結構並未有如此現象。三個 *Xa* 基因表現蛋白質之 LRR 序列的胺基酸組成皆以親水性胺基酸為最多且其中以白胺酸 (leucine) 出現頻率最高,其次為不帶電荷之絲胺酸 (serine)。*Xa1* 蛋白質之 LRR 在某些胺基酸的出現頻率相異於 *Xa21* 和 *Xa26*。這些結果有助於了解不同 *Xa* 基因間表現蛋白質中特有 LRR 序列的結構變異情形。

**關鍵詞：**水稻、*Xa* 基因、白葉枯病、多白胺酸重複、胺基酸組成及屬性。

## 前 言

水稻白葉枯病 (rice bacterial blight) 是由黃單孢桿菌 (*Xanthomonas oryzae* pv. *oryzae*)

引起的一種細菌性維管束病害,多發生於氣候溫暖、土壤肥沃之水田,尤其當植株受傷更易感染,故於颱風後發病較為嚴重,為世界性主要稻作病害之一。目前雖然已有藥劑針對此病

1. 行政院農業委員會農業試驗所研究報告第 2354 號。接受日期：98 年 6 月 26 日。
2. 私立朝陽科技大學生物技術研究所研究生。台灣 台中縣 霧峰鄉。
3. 本所作物組助理研究員。台灣 台中縣 霧峰鄉。
4. 本所技術服務組研究員兼組長。台灣 台中縣 霧峰鄉。
5. 通訊作者,電子郵件：iying@tari.gov.tw；傳真機：(04)23325176。

害具有防治效果，但育成抗病品種為解決此病害最有效之方法 (Chang 1995)。而抗病性是由基因控制，因此日本、中國及國際水稻研究所 (IRRI) 等研究單位，皆已積極對水稻抗白葉枯病基因 (*Xa*) 展開研究。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 *Xa* 基因。其中已知完整序列的有 *Xa1* (Yoshimura *et al.* 1996, 1998; The Rice Chromosome 10 Sequencing Consortium 2003)、*xa5* (Petpisit *et al.* 1977)、*Xa13* (Sanchez *et al.* 1999)、*xa13* (Ogawa *et al.* 1987)、*Xa21* (Song *et al.* 1995)、*Xa26* (Sun *et al.* 2004; Sun *et al.* 2006) 及 *Xa27* (Gu *et al.* 2005) 等七個基因。若能深入瞭解這些已定序之 *Xa* 基因的序列結構，將有利於水稻的分子育種及抗性機制之研究。

植物抗病機制中，有兩套基因系統先後作用，一為抗病基因 (resistant gene, R gene)，二為防衛基因。R 基因能特異性的識別病原菌對應之無毒基因 (*Avr*) 的產物，兩者可直接或間接地相互作用而刺激植物的訊號傳遞系統，最後誘導出植物防衛基因的表達，使植物表現出抗病性，同時 R 基因能不斷進化，可對病原菌不斷突變所產生的生理小種產生抗性 (Wang *et al.* 2002)。植物 R 基因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-酞胺酸激酶 (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-酞胺酸激酶 (intracellular serine-threonine kinase) 區域；及 (5) 其他。R 基因的發現，使水稻的抗病機制在分子層次研究上有了突破性的發展，且由於

水稻的 *Xa* 基因在演化上的保守性與差異性，揭示出 *Xa* 基因可能存在較複雜的分子機制 (Guo *et al.* 2005)。已知 *Xa1* 為上述第一類 R 基因，蛋白質結構特徵屬 CC-NBS-LRR (coiled-coil-NBS-LRR) 類蛋白，其中 LRR 結構域是由 6 個高度保守的 93 個胺基酸的重複單元所組成；而 *Xa21* 和 *Xa26* 為第四類 R 基因，是一種膜蛋白，其特徵是包含一個胞外 LRR 結構域及一個穿膜區和胞內激酶域；至於 *xa5*、*xa13*、*Xa13* 及 *Xa27* 之抗病防禦系統的生化途徑異於前四類 R 基因，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 (Jones & Jones 1997; Yoshimura *et al.* 1998; Khush & Angeles 1999; Dangl & Jones 2001; Anjali & McCouch 2004; Gu *et al.* 2005; Sun *et al.* 2006)。

許多 R 基因已被選殖出來並且定義其特性，發現在 R 基因的表現蛋白質結構上大多具有多白胺酸重複 (leucine-rich repeat, LRR) 序列的共同結構單元 (Dangl & Jones 2001)。已有很多研究顯示 LRR 與生物體的生長發育、抗蟲及抗病的表現有關 (Jones & Jones 1997; Zhang 1998)。LRR 具有多種的結構和功能表現，其結構區域主要傳導蛋白質與蛋白質之間的相互作用，而 LRR 的一級結構，由高度保守的 20–29 個胺基酸所組成，是由一個  $\alpha$ -helix (xaxx±a±±±±a±±x±±) 和一個  $\beta$ -sheet (LxxLxLxxN/CxL) 通過 loop 環接形成一個非球形的馬蹄形分子結構 (附錄 1)，其中 x 為任何一種胺基酸，a 代表脂肪族胺基酸 (aliphatic amino acid)，± 表示該位置可能出現插入 (+) 或缺失 (-)，L 表示白胺酸 (leucine)、N 為天門冬胺酸 (aspartic acid)、C 為半胱胺酸 (cysteine) (Kobe & Deisenhofer 1995, 1996)。在水稻 *Xa* 基因蛋白質結構也發現與 LRR 有密切關係存在 (Song *et al.* 1995; Yoshimura *et al.* 1996; Baker *et al.* 1997; Song *et al.* 1997; Sun *et al.* 2004, 2006; Gu *et al.* 2005; Wang 2006)。

為了解不同 Xa 基因表現蛋白質之 LRR 序列區域結構的變異形式，因此本研究首先自網站上各公共資料庫蒐集 Xa 基因相關之蛋白質序列，從中找出 LRR 片段後，探討各 Xa 基因蛋白質序列中 LRR 的位置分布差異性及胺基酸偏好性。

## 材料與方法

### Xa 基因蛋白質序列的蒐集

本研究首先根據 Oryzabase 資料庫內已定義的 7 個 Xa 基因 (*Xa1*、*Xa5*、*Xa13*、*xa13*、*Xa21*、*Xa26* 及 *Xa27*) 的基因符號，到 NCBI (National Center for Biotechnology Information, <http://www.ncbi.nlm.nih.gov/>, Yoshimura *et al.* 1998) 資料庫中搜尋及下載它們的編碼區序列 (coding domain sequence, CDS) 及蛋白質序列。如果在 NCBI 中尚未登錄有這些 Xa 基因的蛋白質序列，則再利用 BioEdit 生物資訊軟體 (Biological Sequence Alignment Editor for Win 95/98/NT/2K/XP, 7.0.5.2 版, <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>, Hall 2001)，將其 CDS 轉譯成蛋白質序列。

接著，另從 KOME (Knowledge-based Oryza Molecular Biological Encyclopedia, <http://cdna01.dna.affrc.go.jp/cDNA/>, Kikuchi *et al.* 2003) 資料庫中，以水稻白葉枯病學名“*Xanthomonas oryzae* pv. *oryzae*”及其俗名“bacterial blight”進行交叉搜尋，經逐一檢視含此字義之序列登錄號 (accession no.)，再以此登錄號連結至 NCBI 資料庫，以找出 Xa 基因的全長互補 DNA (full length-complementary DNA, FL-cDNA) 及其經由最長讀取框架 (longest open reading frame, longest ORF) 所轉譯的蛋白質序列。

最後將上述各 Xa 基因的蛋白質序列儲存成 FASTA 格式，以供進行後續各項分析。

### Xa 基因蛋白質之 LRR 序列片段的搜尋及其分布

Pfam (<http://pfam.janelia.org/>, Finn *et al.* 2006) 是利用 Swissprot 50.0 和 SP-TrEMBL 33.0 蛋白質序列資料庫內容，依據多序列排比 (multiple sequence alignment) 與隱匿馬可夫模型 (hidden Markov model, HMM) 所建構而成的蛋白質家族資料庫，儲存有關的蛋白質結構、多序列排比、蛋白質結構域 (protein structure domain) 等資訊，且能進行蛋白質結構預測。至 2008 年 7 月止，第 23 版的 Pfam 資料庫總共蒐集了 10,340 個蛋白質家族。於是本研究將所蒐集之 Xa 基因蛋白質序列，個別輸入 Pfam 網站查詢畫面，使用其內設的 Global & local (merged) 及 E-value = 1.0 二種條件進行查詢，以搜尋 Xa 基因蛋白質序列中的蛋白質結構域，藉此找出符合 LRR 結構的序列片段。然後，根據 Pfam 所顯示的 LRR 在蛋白質序列上的位置，利用 AutoCAD 繪圖軟體畫出其分布狀況，並將不同 Xa 基因結果並列，以比較不同 Xa 基因之蛋白質序列中各 LRR 片段的位置分布。

### Xa 基因蛋白質之 LRR 序列的胺基酸組成分析

**胺基酸組成比例：**由於所搜集的各 Xa 基因之蛋白質序列不只一條 (分別來自於 NCBI 和 KOME 資料庫)，因此首先計算不同 Xa 基因蛋白質各來源序列中 LRR 的胺基酸組成比例 (%)，進而求得各基因蛋白質之 LRR 的胺基酸組成之平均比例和變異，再利用 Excel 軟體繪製雷達圖來呈現各 Xa 基因間表現蛋白質之 LRR 的胺基酸組成差異性。

**胺基酸屬性含量：**根據胺基酸的支鏈性質 (Mathews *et al.* 2000)，將 20 種胺基酸區分為極性 (polar) 與非極性 (non-polar)，即分別為親水性 (hydrophile) 與疏水性 (hydrophobic) 胺基酸。非極性胺基酸有丙胺酸 (alanine, Ala)、苯丙胺酸 (phenylalanine, Phe)、異白胺酸 (isoleucine, Ile)、白胺酸 (leucine, Leu)、甲硫胺

酸 (methionine, Met)、脯氨酸 (proline, Pro)、纈氨酸 (valine, Val) 及色氨酸 (tryptophan, Trp)；極性胺基酸依其帶電與否，又分為帶正電荷 (positively charged) (即鹼性)、帶負電荷 (negatively charged) (即酸性) 與不帶電荷 (uncharged) (即中性)；不帶電荷胺基酸有胱氨酸 (cysteine, Cys)、甘氨酸 (glycine, Gly)、天冬醯胺酸 (asparagine, Asn)、麩醯胺酸 (glutamine, Gln)、絲氨酸 (serine, Ser)、蘇氨酸 (threonine, Thr) 及酪氨酸 (tyrosine, Tyr)，帶正電荷胺基酸有組氨酸 (histidine, His)、離氨酸 (lysine, Lys)、精氨酸 (arginine, Arg)，帶負電荷胺基酸有天冬氨酸 (aspartic acid, Asp) 及麩氨酸 (glutamic acid, Glu)。首先計算不同 *Xa* 基因蛋白質各來源序列之 LRR 的不同屬性胺基酸所佔比例和變異，然後進行其在基因間的差異顯著性測驗。

## 結 果

### *Xa* 基因蛋白質序列的蒐集結果

本研究透過 NCBI 搜尋到七個已被定序的 *Xa* 基因之相關資訊，整理如表 1，這些基因的來源包含有印度型水稻 (*Oryza sativa indica* cultivar-group)、日本型水稻 (*japonica*)、長花藥野生稻 (*Oryza longistaminata*)。目前在 NCBI 的 *Xa1* 有分別於 1999 及 2004 年登錄之位於第 4 及第 10 條染色體上的兩條序列，本研究各以 *Xa1\_1999* 及 *Xa1\_2004* 代號表示；*xa5*、*xa13*、*Xa13* 及 *Xa27* 各一條序列，分別位於第 5、8、8 及 6 條染色體上，其中 *xa5* 和 *xa13* 為隱性基因；*Xa21* 位於第 11 條染色體上，含 *Xa21-A1*、*Xa21-A2*、*Xa21-B*、*Xa21-C*、*Xa21-D*、*Xa21-E* 及 *Xa21-F* 共七個基因家族；*Xa26* 也在第 11 條染色體上，含 *Xa26-MRKA*、*Xa26-MRKB*、*Xa26-MRKC* 及 *Xa26-MRKD* 四個基因家族，其 DNA 序列儲存在 NCBI 資料庫的登錄號皆為 DQ355952，但長度不同，本研究各以 *Xa26-A*、

*Xa26-B*、*Xa26-C*、*Xa26-D* 簡稱之。由於 *Xa21-A2*、*Xa21-C*、*Xa21-F* 及 *Xa26-D* 在 NCBI 中尚未登錄有完整的蛋白質序列，故本研究另外利用 BioEdit 軟體將此四條 CDS 轉譯成蛋白質序列。除了 *xa5*、*xa13*、*Xa13* 及 *Xa27* 基因之蛋白質序列長度小於 500 外，其餘 *Xa* 基因之蛋白質序列介於 612–1820 間 (表 1)。

接著，透過 KOME 連結至 NCBI 資料庫共搜尋到 11 條 FL-cDNA 序列，這些序列中皆含有 *Xa* 基因的蛋白質序列，相關資訊整理如表 2，*Xa1*、*Xa21* 及 *Xa26* 基因各找到 9、1 及 1 條序列，但沒有找到其他 *Xa* 基因之 FL-cDNA 序列。由於這些序列與表 1 序列所處的染色體位置不同，並分屬於不同染色體上 (表 2)，因此本研究根據它們所在的染色體代號來命名這些來源序列，若同一染色體上有二個以上不同序列，則再根據 NCBI 登錄號大小於染色體代號後加上不同小寫字母，例如 *Xa1-2a*、*Xa1-2b* 各代表 *Xa1* 基因在第 2 條染色體上的第一、二條 FL-cDNA 序列。其中 *Xa1* 基因的 *Xa1-4a* 和 *Xa1-4b* 與 NCBI 之 *Xa1\_1999* 都位於第 4 條染色體上 (表 1、2)，利用 NCBI 的兩序列排比工具 blast2 (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>) 兩兩排比結果，發現 *Xa1-4b* 與 *Xa1\_1999* 並不相似 (no significant similarity)，但 *Xa1-4a* 之 4–1101 bp 序列區域與 *Xa1\_1999* 之 4312–5409 bp 範圍內的序列一致性比值 (identities) 高達 1035/1098 (94%)，此有待後續 Pfam 資料庫搜尋 LRR 結果，倘 LRR 位置分布一致時，才能確認它們是相同序列。這些 *Xa* 基因 FL-cDNA 序列的來源包含印度型及日本型水稻，其經由最長讀取框架轉譯的蛋白質序列之長度在 366–1575 之間 (表 2)。

### *Xa* 基因蛋白質之 LRR 序列片段的搜尋結果及其分布情形

將上述找到的 28 條不同 *Xa* 基因來源之蛋白質序列，透過 Pfam 資料庫搜尋結果，僅

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<sup>z</sup>

Label	Gene symbol	Chr. no.	Submitted year	Accession no. in NCBI	Definition	Gene length	Location	CDS <sup>y</sup>	Length	Accession no.	Protein sequence Length
Xa1_1999	Xa1	4	1999	AB002266	<i>Oryza sativa (indica cultivar-group)</i> mRNA for Xa1	5910	113..5521		5409	BAA25068	1802
Xa1_2004	Xa1	10	2004	NM_194685	<i>Oryza sativa (japonica cultivar-group)</i> putative bacterial blight resistance protein Xa1-like protein, mRNA	4206	1..4206		4206	NP_919667	1401
xa5	xa5	5	2004	AY643716	<i>Oryza sativa (indica cultivar-group)</i> I transcription factor IIA gamma subunit (TFIIA) mRNA, partial cds.	318	1..318		318	AAV53715	106
xa13	xa13	8	2006	DQ421394	<i>Oryza sativa (indica cultivar-group)</i> cultivar IRBB13 disease resistant allele xa13 (xa13) gene, complete cds.	924	1..924		924	ABD78942	307
Xa13	Xa13	8	2006	DQ421395	<i>Oryza sativa (indica cultivar-group)</i> cultivar IR64 disease resistant allele Xa13 (Xa13) gene, complete cds.	924	1..924		924	ABD78943	307
Xa21-A1	Xa21	11	1997	U72725	<i>Oryza longistaminata</i> receptor kinase-like protein gene	8416	4771..7384,7676..8052		2991	AAB82755	996
Xa21-A2	Xa21	11	1998	U72727	<i>Oryza longistaminata</i> receptor kinase-like protein, pseudogene sequence	5940	2151..4764,4872..5248		2991	-	996 <sup>x</sup>
Xa21-B	Xa21	11	1995	U37133	<i>Oryza sativa (indica cultivar-group)</i> /receptor kinase-like protein gene	3921	1..2677,3521..3921		3078	AAC49123	1025
Xa21-C	Xa21	11	1998	U72723	<i>Oryza longistaminata</i> receptor kinase-like protein gene, pseudogene	19639	15118..17720,17827..18201		2978	-	991 <sup>x</sup>
Xa21-D	Xa21	11	1997	U72726	<i>Oryza longistaminata</i> receptor kinase-like protein	13341	2367..4205		1839	AAB82753	612
Xa21-E	Xa21	11	1997	U72724	<i>Oryza sativa (indica cultivar-group)</i> receptor kinase-like protein gene	9424	2819..5260		2442	AAB82756	813
Xa21-F	Xa21	11	2005	U72728	<i>Oryza longistaminata</i> receptor-like kinase protein, pseudogene sequence	7204	1683..4352,5147..5547		3071	-	1023 <sup>x</sup>
Xa26-A	Xa26	11	2006	DQ355952	<i>Oryza sativa (indica cultivar-group)</i> receptor kinase MRKa	3401	14225..17135,17240..17625		3297	ABD36511	1098
Xa26-B	Xa26	11	2006	DQ355952	<i>Oryza sativa (indica cultivar-group)</i> bacterial blight resistance protein Xa26	3631	21725..24659,24765..25141		3312	ABD36512	1103
Xa26-C	Xa26	11	2006	DQ355952	<i>Oryza sativa (indica cultivar-group)</i> receptor kinase MRKc	3456	34496..34911,35026..37951		3342	ABD36513	1113
Xa26-D	Xa26	11	2006	DQ355952	<i>Oryza sativa (indica cultivar-group)</i> cultivar Minghui 63 receptor kinase pseudogene	11396	2021..3423,9305..10839,13028..13416		3327	-	1108 <sup>x</sup>
Xa27	Xa27	6	2005	AY986491	<i>Oryza sativa (indica cultivar-group)</i> Xa27 gene, Xa27-IR24 allele	2393	1590..1931		342	AAV54163	113

<sup>z</sup> Source: Oryzabase and NCBI databases, searched from Feb. 29, 2008.

<sup>y</sup> CDS: coding domain sequence.

<sup>x</sup> Protein sequence translated from CDS using BioEdit software.

表 2. 水稻 *Xa* 基因全長 cDNA 序列相關資訊Table 2. Information of rice *Xa* genes with full-length cDNA sequence <sup>z</sup>

Label	Chr. no.	Accession no.	Nucleotide analysis			Amino acid analysis		
			GenBank hit (by BLASTN)	Genome annotation (from GenBank hit)	cDNA length	Longest ORF <sup>y</sup>	PIR (by BLASTX)	Length
<i>Xa1-2a</i>	2	AK073576	<i>Oryza sativa</i> (japonica cultivar-group) cDNA clone:J033046M12, full insert sequence	<i>Oryza sativa</i> (japonica cultivar-group) genomic DNA, chromosome 2, BAC clone: OSJNBa0011103.	4511	1190..4294	bacterial blight-resistance protein <i>Xa1</i> -rice	1034
<i>Xa1-2b</i>	2	AK242786	<i>Oryza sativa</i> (japonica cultivar-group) genomic DNA, chromosome 2, complete sequence	AP004084 <i>Oryza sativa</i> (japonica cultivar-group) genomic DNA, chromosome 2, BAC clone: OJ1217_F02.	3642	701..2644	bacterial blight-resistance protein <i>Xa1</i> -rice	647
<i>Xa1-4a</i>	4	AK105096	<i>Oryza sativa</i> (japonica cultivar-group) cDNA clone:001-100-B05, full insert sequence	—	1873	413..1513	bacterial blight-resistance protein <i>Xa1</i> -rice	366
<i>Xa1-4b</i>	4	AK241237	<i>Oryza sativa</i> (japonica cultivar-group) genomic DNA, chromosome 4, complete sequence	OSJN00106 <i>Oryza sativa</i> genomic DNA, chromosome 4, BAC clone: OSJNBb0060E08, complete sequence.	2943	254..2866	bacterial blight-resistance protein <i>Xa1</i> -rice	870
<i>Xa1-6</i>	6	AK100253	<i>Oryza sativa</i> (japonica cultivar-group) cDNA clone:J023061C20, full insert sequence	<i>Oryza sativa</i> (japonica cultivar-group) genomic DNA, chromosome 6, PAC clone:P0621D05.	2552	202..1392	bacterial blight-resistance protein <i>Xa1</i> -rice	396
<i>Xa1-8a</i>	8	AK066438	<i>Oryza sativa</i> (japonica cultivar-group) cDNA clone:J013067B14, full insert sequence	<i>Oryza sativa</i> (japonica cultivar-group) genomic DNA, chromosome 8, BAC clone: OJ1212_C09.	3385	218..2998	bacterial blight-resistance protein <i>Xa1</i> -rice	926
<i>Xa1-8b</i>	8	AK121298	<i>Oryza sativa</i> (japonica cultivar-group) cDNA clone:J023110L09, full insert sequence	<i>Oryza sativa</i> (japonica cultivar-group) genomic DNA, chromosome 8, PAC clone:P0024C06.	5161	158..4885	bacterial blight-resistance protein <i>Xa1</i> -rice	1575
<i>Xa1-11</i>	11	AK106715	<i>Oryza sativa</i> genomic DNA, chromosome 4, BAC clone: OSJNBb0085C12, complete sequence.	<i>Oryza sativa</i> genomic DNA, chromosome 4, BAC clone: OSJNBb0085C12, complete sequence.	2202	46..1200	bacterial blight-resistance protein <i>Xa1</i> -rice	384
<i>Xa1-12</i>	12	AK065947	<i>Oryza sativa</i> (japonica cultivar-group) cDNA clone:J013047A22, full insert sequence	<i>Oryza sativa</i> chromosome 12, BAC OSJNBa0035B12 of library OSJNBa from chromosome 12 of cultivar Nipponbare of ssp. japonica of <i>Oryza sativa</i> (rice), complete sequence.	4851	473..4288	bacterial blight-resistance protein <i>Xa1</i> -rice	1271
<i>Xa2-1</i>	1	AK072480	<i>Oryza sativa</i> (japonica cultivar-group) cDNA clone:J023126F05, full insert sequence	<i>Oryza sativa</i> (japonica cultivar-group) genomic DNA, chromosome 1, PAC clone:P0431H09.	3828	1250..3343	protein kinase <i>Xa21</i> (EC 2.7.1.-) AI, receptor type-long-staminate rice	697
<i>Xa26-11</i>	11	AK241618	<i>Oryza sativa</i> (japonica cultivar-group) chromosome 11 clone OSJNBa0004O15 map near R3342S, complete sequence	AY364476 <i>Oryza sativa</i> (indica cultivar-group) bacterial blight resistance protein ( <i>Xa26</i> ) gene, complete cds.	3617	45..1760	protein kinase <i>Xa21</i> (EC 2.7.1.-) AI, receptor type - long-staminate rice	571

<sup>z</sup> Source: KOME databases, searched from Feb. 29, 2008.<sup>y</sup> ORF: open reading frame.

*Xa1*、*Xa21*、*Xa26* 三個基因找到 LRR 序列區域，而 *xa5*、*Xa13*、*xa13*、*Xa27* 則完全沒有。根據 Pfam 資料庫搜尋結果，整理出 *Xa1*、*Xa21*、*Xa26* 基因各來源序列之 LRR 的數量及其在蛋白質序列上的位置，如圖 1 所示。由於 *Xa1-4a* 與 *Xa1\_1999* 經 blast2 排比結果

相似度很高 (如前節所述)，而在圖 1 也發現兩者的 LRR 分布完全相同，因此確認 *Xa1-4a* 應是 *Xa1\_1999* 的部份片段，即屬相同序列。於是 *Xa1-4a* 序列在後續統計分析時將不予納入，亦即本研究用以分析的 *Xa1*、*Xa21* 及 *Xa26* 三個基因的來源序列分別有 10、8 及 4 條。

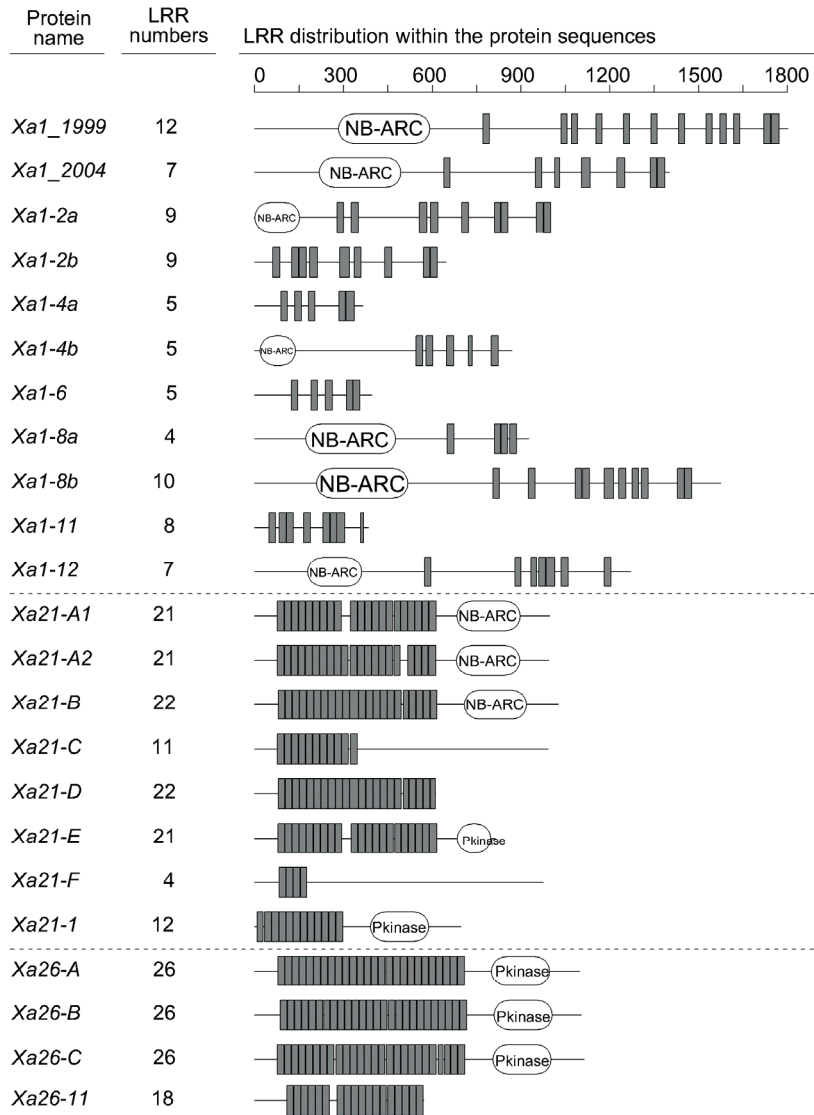


圖 1. *Xa* 蛋白質序列之 LRR 位置分布圖。■：LRR。

Fig. 1. Distribution of leucine-rich repeat (LRR) in the protein sequences of rice *Xa* genes. ■：LRR.

在這些來源序列中，可發現 *Xa1* 在蛋白質序列上的 LRR 數量都較少 (4–12 條)，而 *Xa21* 與 *Xa26* 的 LRR 數量則相對較多 (除 *Xa21-F* 僅有 4 條外)，在 11–26 條之間 (圖 1)。三個 *Xa* 基因的 LRR 序列之平均長度相近，約 23–24 個胺基酸，其中雖然 *Xa1* 之 LRR 的長度變異性較大，範圍在 11–33 個胺基酸之間，但出現頻率最多的 LRR 長度也是 24 個胺基酸，與其他兩個 *Xa* 基因表現蛋白質之 LRR 長度相當 (表 3)。

三個 *Xa* 基因蛋白質之 LRR 的分布有很大的差異：*Xa1* 之 LRR 在蛋白質序列上的分布較疏鬆分散 (圖 1)，其間距的變異程度甚大，最大距離可達 289 個胺基酸，但最短距離僅 1 個胺基酸，平均距離為  $55.6 \pm 63.7$  個胺基酸 (表 3)；而 *Xa21* 與 *Xa26* 之 LRR 在蛋白質序列中的分布則緊密集中 (圖 1)，平均距離分別為  $3.0 \pm 4.4$  及  $2.6 \pm 2.8$  個胺基酸，該兩基因之 LRR 序列的間距差異不大 (表 3)。此外，*Xa1* 之 LRR 片段大多位於 NB-ARC (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4) 區域序列之後，而 *Xa21* 與 *Xa26* 之 LRR 則在 NB-ARC 或 P-kinase (protein-kinase) 區域序列之前 (圖 1)。*Xa1* 的蛋白質序列中靠近 NB-ARC 區域序列的前一個或兩個 LRR 似乎遠離於其他 LRR，但 *Xa21* 和 *Xa26* 的蛋白質之 LRR 區域結構並未有如此現象 (圖 1)。

表 3. *Xa* 基因之蛋白質序列之 LRR 的長度與間距

Table 3. Length and distance of leucine-rich repeat (LRR) in protein sequences of rice *Xa* genes

Gene	Length				Distance			
	Mean $\pm$ sd <sup>z</sup>	Minimum	Maximum	Mode <sup>y</sup>	Mean $\pm$ sd	Minimum	Maximum	Mode
<i>Xa1</i>	23.6 $\pm$ 3.0	11	33	24	55.6 $\pm$ 63.7	1	289	2
<i>Xa21</i>	23.0 $\pm$ 1.2	17	29	23	3.0 $\pm$ 4.4	2	32	2
<i>Xa26</i>	23.0 $\pm$ 0.8	18	24	23	2.6 $\pm$ 2.8	2	27	2

<sup>z</sup> sd: standard deviation.

<sup>y</sup> mode: the value that occurs the most frequently in a data set.

### *Xa* 基因蛋白質之 LRR 序列的胺基酸組成分析結果

各 *Xa* 基因蛋白質之 LRR 序列的胺基酸組成，如圖 2 所示，可明顯看出皆以 Leu 含量最多且 Ser 次之，但 Cys、Glu、Lys 和 Pro 所佔比例以 *Xa1* 較 *Xa21* 及 *Xa26* 稍多，而 Gly、Ile 和 Asn 所佔比例以 *Xa1* 少於 *Xa21* 及 *Xa26*。

顯著性測驗結果，LRR 序列之疏水性胺基酸以比較各 *Xa* 基因蛋白質之 LRR 序列的胺基酸屬性含量，顯示皆以親水性之不帶電胺基酸最多，其次是疏水性胺基酸 (圖 3)。由於各屬性胺基酸的含量均符合常態分布，因此經統計 *Xa21* 和 *Xa26* 顯著多於 *Xa1*，而不帶電胺基酸以 *Xa1* 和 *Xa26* 顯著較 *Xa21* 多，帶正電與帶負電胺基酸則以 *Xa1* 顯著多於 *Xa21* 和 *Xa26*。

### 討 論

檢視表 1 與表 2 之序列的定義內容，表 1 皆為已被定序完成的 *Xa* 基因，而表 2 則屬於 FL-cDNA 序列且當中皆含有與 *Xa* 基因有關的蛋白質序列，唯目前尚查無文獻以確認這些有相同基因符號但位於不同染色體之序列，是否屬於同一個基因。由於已知 *Xa* 基因中存在 R 基因結構，本研究也於 11 條 FL-cDNA 序列中發現含有此種結構，因此將這些序列合併以擴大分析之序列內容，希望能從中更瞭解與 *Xa* 基因有關的抗病性序列結構。

已知 *Xa1* 屬於第一類 R 基因 (Dangl & Jones 2001)，其基因產物的結構為 CC-NBS-LRR 類蛋白，不同於其他 NB-LRR 抗病蛋白質的結構特徵 (Jones & Jones 1997)。*Xa21* 與 *Xa26* 同屬於第四類 R 基因 (Dangl & Jones 2001)，它的基因產物是一種膜蛋白，其特徵是包含一個胞外 LRR 結構域及一個穿膜區和胞內激酶域 (Jones & Jones 1997)。從本研究結果發現，在 *Xa1*、*Xa21* 和 *Xa26* 的蛋白質之 LRR 結構域中每一個 LRR 的長度未必相等，最多由 23–24 個胺基酸所組成，而 *Xa1* 的 LRR 長度之變異性遠大於 *Xa21* 和 *Xa26*。此外，第一類 R 基因的 *Xa1* 之 LRR 在蛋白質序列中的分布呈現疏鬆分散，而 *Xa21* 和 *Xa26* 的蛋白質之 LRR 則分布緊密集中；同時也發現，*Xa1* 的蛋白質之 LRR 片段大多位於 NB-ARC 區域序列之後，而 *Xa21* 與 *Xa26* 的蛋白質之 LRR 則在 NB-ARC 或 P-kinase 區域序列之前；*Xa1* 在其蛋白質序列中靠近 NB-ARC 區域序列的 LRR 明顯遠離其他 LRR，但在 *Xa21* 和 *Xa26* 的蛋白質序列上並無發現此現象。由此顯示，就 LRR 序列的長度及位置分布而言，不同 R 基因類別的 *Xa* 基因存在相當的差異性，而同一 R 基因類別的 *Xa21* 和 *Xa26* 則無明顯差異，此等發現有助於初步了解這些已定序 *Xa* 基因間特有 LRR 序列之結構的變異情形。根據 Belkadir *et al.* (2004) 的研究推測，胞內 LRR 結構域的不同位置 LRR 可能具有不同的功能；靠近胺基端 (N-terminal) 的 LRR 可能具有與胞內因子互相作用而刺激調節的功能，而位於羧基端 (C-terminal) 的 LRR 可能參與與病原蛋白的識別。因此未來值得針對特定位置的 LRR 之結構特性再深入研究。

蛋白質的一級構造是蛋白質最終構成的根本，各級構造的訊息都決定於胺基酸的序列，即蛋白質摺疊所含之訊息揭示在其胺基酸殘基的線性序列中 (Shi *et al.* 2005)，而胺基酸的分

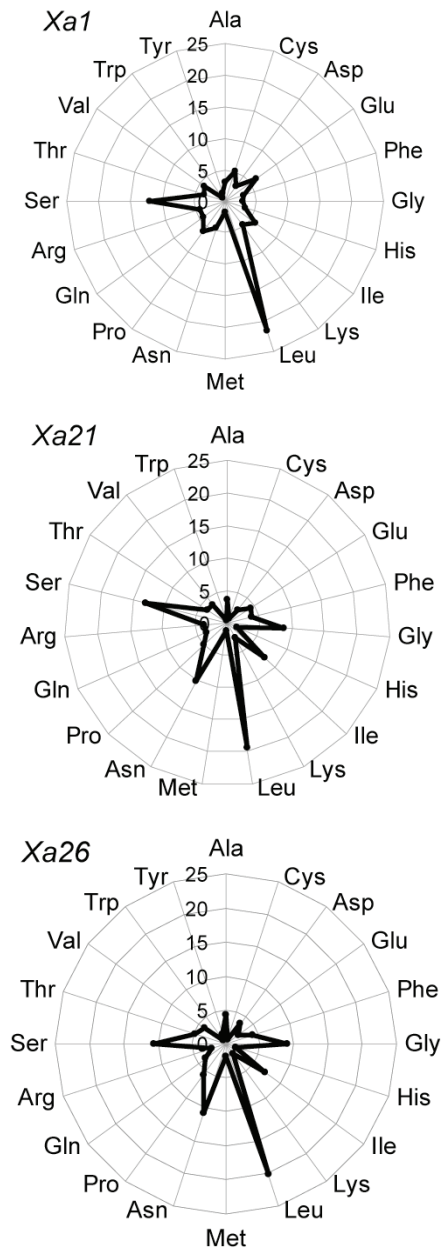


圖 2. *Xa* 基因蛋白質序列之 LRR 的胺基酸組成差異的雷達圖。

Fig. 2. Polar plots showing the compositional differences among amino acids of leucine-rich repeat (LRR) in protein sequences of rice *Xa1* (top), *Xa21* (middle) and *Xa26* (bottom) genes.

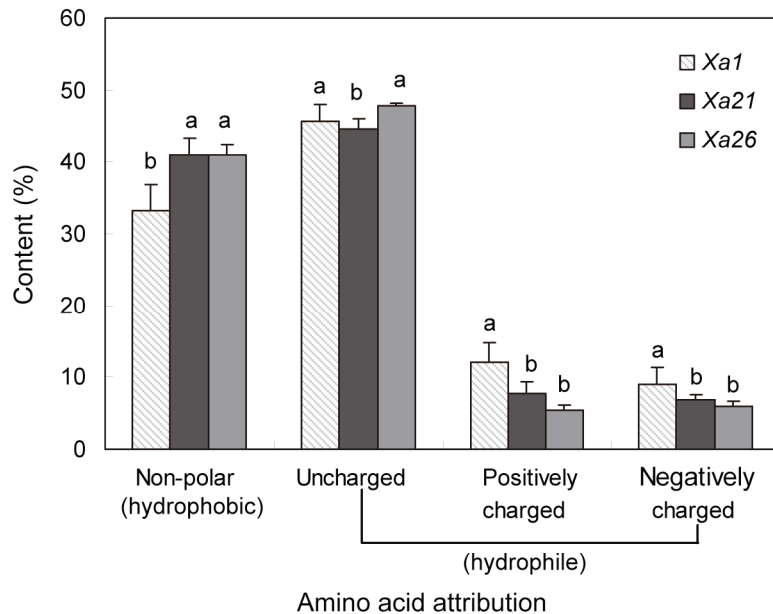


圖 3. *Xa* 基因蛋白質序列之 LRR 的胺基酸屬性含量。數值以平均值±標準差表達。各平均值上示以相同字母者為 5% 水準下經 LSD 測驗未達顯著差異。

Fig. 3. Contents of amino acid attributes of leucine-rich repeat (LRR) in protein sequences of rice *Xa1*, *Xa21* and *Xa26* genes. Data are expressed as mean  $\pm$  standard deviation. Means within each attribute followed by the same letters among genes are not significantly different at 5% level by LSD test.

類和序列中應含有蛋白質二級結構的訊息 (Wang & Wang 1999)。三個 *Xa* 基因之蛋白質序列 LRR 皆以親水性胺基酸含量較多且其中以 Leu 出現頻率最高，這是因為它們皆屬於 LRR 序列區域結構 (Song *et al.* 1995; Yoshimura *et al.* 1998; Sun *et al.* 2006)。由於 *Xa21* 和 *Xa26* 的抗病機制中含有細胞外 LRR 區域之穿膜受體和細胞內絲胺酸-酰胺酸激酶區域 (Song *et al.* 1995; Sun *et al.* 2006)，因此發現這兩個 *Xa* 基因之蛋白質 LRR 序列中不帶電之疏水性胺基酸的 Ser 也較多，而 *Xa1* 也含有多量的 Ser。20 種胺基酸出現在 LRR 序列中之頻率，僅 *Xa1* 的 Cys、Glu、Lys 和 Pro 所佔比例較 *Xa21* 和 *Xa26* 稍多，而 *Xa1* 的 Gly、Ile 和 Asn 所佔比例較 *Xa21* 和 *Xa26* 為少。由此得知，分屬於不同類別 R 基因的 *Xa* 之表現蛋白質

LRR 序列的胺基酸組成有其特定的偏好模式。蛋白質中的胺基酸組成隱藏著許多訊息，其胺基酸之同義密碼子特性會因不同物種或物種內不同基因而異，因此蛋白質的胺基酸密碼子之偏好性分析亦成為生物資訊研究的重要探討領域。未來將繼續針對這些 *Xa* 基因之 LRR 序列的密碼子偏好性、高頻密碼子確定及一致序列尋找等相關課題，深入探討，以釐清不同 *Xa* 基因在其表現蛋白質之 LRR 序列區域結構的差異性。

## 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。

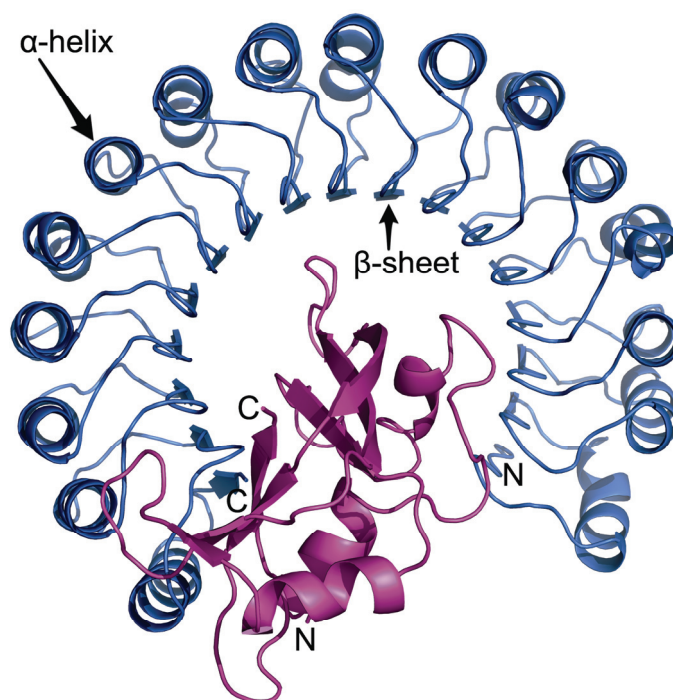
## 引用文獻 (Literature cited)

- 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.
- Belkadir, Y., R. Subramaniam, and J. L. Dangl. 2004. Plant disease resistance protein signaling: NBS-LRR proteins and their partners. *Curr. Opin. Plant Biol.* 7:391–399.
- Chang, S. J. 1995. Physiological and Genetical Studies on Bacterial Blight (*Xanthomonas oryzae* pv. *oryzae*) Resistance in Rice (*Oryza sativa* L.). Doctor Dissertation. Graduate Institute of Agronomy, NCHU, Taichung. 172 pp. (in Chinese with English abstract)
- Dangl, J. L. and J. D. Jones. 2001. Plant pathogens and integrated defense responses to infection. *Nature* 411:826–833.
- Finn, R. D., J. Mistry, B. Schuster-Böckler, S. Griffiths-Jones, V. Hollich, T. Lassmann, S. Moxon, M. Marshall, A. Khanna, R. Durbin, S. R. Eddy, Erik L. L. Sonnhammer, and A. Bateman. 2006. Pfam: clans, web tools and services. *Nucleic Acids Res.* 34:D247–251.
- 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.
- Hall, T. 2001. BioEdit Version 5.0.6. North Carolina State Univ. 192 pp.
- Jones, D. A. and J. D. G. Jones. 1997. The role of leucine-rich repeat proteins in plant defences. *Adv. Bot. Res.* 24:89–167.
- Khush, G. S. and E. R. Angles. 1999. A new gene for resistance to race 6 of bacterial blight in rice, *Oryzae sativa* L. *Rice Genet. Newslet.* 16:92–93.
- Kikuchi, S., K. Satoh, T. Nagata, N. Kawagashira, K. Doi, N. Kishimoto, J. Yazaki, M. Ishikawa, H. Yamada, H. Ooka, I. Hotta, K. Kojima, T. Namiki, E. Ohneda, W. Yahagi, K. Suzuki, C. J. Li, K. Ohtsuki, T. Shishiki, Y. Otomo, K. Murakami, Y. Iida, S. Sugano, T. Fujimura, Y. Suzuki, Y. Tsunoda, T. Kurosaki, T. Kodama, H. Masuda, M. Kobayashi, Q. Xie, M. Lu, R. Narikawa, A. Sugiyama, K. Mizuno, S. Yokomizo, J. Niikura, R. Ikeda, J. Ishibiki, M. Kawamata, A. Yoshimura, J. Miura, T. Kusumegi, M. Oka, R. Ryu, M. Ueda, K. Matsubara, J. Kawai, P. Carninci, J. Adachi, K. Aizawa, T. Arakawa, S. Fukuda, A. Hara, W. Hashidume, N. Hayatsu, K. Imotani, Y. Ishii, M. Itoh, I. Kagawa, S. Kondo, H. Konno, A. Miyazaki, N. Osato, Y. Ota, R. Saito, D. Sasaki, K. Sato, K. Shibata, A. Shinagawa, T. Shiraki, M. Yoshino, and Y. Hayashizaki. 2003. Collection, mapping, and annotation of over 28000 cDNA clones from japonica rice. *Science* 301:376–379.
- Kobe, B. and J. Deisenhofer. 1995. Proteins with leucine-rich repeats. *Curr. Opin. Struct. Biol.* 5:409–416.
- Kurata, N. and Y. Yamazaki. 2006. Oryzabase. An integrated biological and genome information database for rice. *Plant Physiol.* 140:12–17.
- Mathews, C. K., K. E. van Holde, K. G. Ahern. 2000. *Biochemistry*. 3<sup>rd</sup> ed. Addison-Wesley Inc., USA. 1200 pp.
- Ogawa, T., L. Lin, R. E. Tabien, and G. S. Khush. 1987. A new recessive gene for resistance to bacterial blight of rice. *Rice Genet. Newslet.* 4:98–100.
- Petpisit, V., G. S. Khush, and H. E. Kauffman. 1977. Inheritance of resistance to bacterial blight in rice. *Crop Sci.* 17:551–564.
- Sanchez, A. C., L. L. Ilag, D. Yang, D. S. Brar, F. Ausubel, G. S. Khush, M. Yano, T. Sasaki, Z. Li, and N. Huang. 1999. Genetic and physical mapping of *xa13*, a recessive bacterial blight resistance gene in rice. *Theor. Appl. Genet.* 98:1022–1028.
- Shi, X. H., X. R. Liu, L. Luo, W. B. Liu, and J. Xu. 2005. A research of amino acid order based on amino acid classification. *J. Biomath.* 20:491–495.
- 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.
- Song, W. Y., L. Y. Pi, G. L. Wang, J. Gardner, T. Holsten, and P. C. Ronald. 1997. Evolution of the rice *Xa21* disease resistance gene family. *Plant Cell* 9: 1279–1287.
- Sun, X. H., T. G. Lu, S. R. Jia, and D. F. Huang. 2004. Bioinformatics analysis of the LRR-NBS gene family in rice. *Sci. Agric. Sinica* 37:1–7.

- Sun, X., Y. Cao, and S. 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.
- The Rice Chromosome 10 Sequencing Consortium. 2003. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* 300:1566–1569.
- Wang, J. and W. Wang. 1999. A computational approach to simplifying the protein folding alphabet. *Nat. Struc. Biol.* 6:1033–1038.
- Wang, A. J., C. X. Zhu, and F. J. Wen. 2002. Molecular biology of the resistance to blight disease in rice. *J. Shandong Agric. Univ. (Nat. Sci.)* 33:101–106. (in Chinese with English abstract)
- Yamazaki, Y. and P. Jaiswal. 2005. Biological ontologies in rice databases. An introduction to the activities in Gramene and Oryzabase. *Plant Cell Physiol.* 46:63–68.
- Yoshimura, S., Y. Umehara, N. Kurata, Y. Nagamura, T. Sasaki, Y. Minobe, and N. Iwata. 1996. Identification of a YAC clone carrying the *Xa-1* allele, a bacterial blight-resistance gene in rice. *Theor. Appl. Genet.* 93:117–122.
- 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. USA* 95:1663–1668.
- Zhang, X. 1998. Leucine-rich repeat receptor-like kinases in plants. *Plant Mol. Biol. Rep.* 16:301–311. (in Chinese with English abstract)

附錄 1. LRR 結構之帶狀圖 (Kobe & Deisenhofer 1996)。

**Appendix 1.** A diagram showing ribbon structure of the leucine-rich repeat (LRR) (Kobe & Deisenhofer 1996).



# Leucine-rich Repeat (LRR) Distribution and Amino Acid Composition in Protein Sequences of Genes with Resistance to Bacterial Blight (*Xanthomonas oryzae*) of Rice<sup>1</sup>

Yi-Chen Lin<sup>2</sup>, Chun-Tang Lu<sup>3</sup>, Meng-Li Wei<sup>3</sup>, and Hsiu-Ying Lu<sup>4,5</sup>

## Abstract

Lin, Y. C., C. T. Lu, M. L. Wei, and H. Y. Lu. 2009. Leucine-rich repeat (LRR) distribution and amino acid composition in protein sequences of genes with resistance to bacterial blight (*Xanthomonas oryzae*) of rice. *J. Taiwan Agric. Res.* 58:93–105.

Since leucine-rich repeat (LRR) is known as a domain of protein specific to disease resistance genes, it is used to study the genes of rice with resistance to bacterial blight caused by *Xanthomonas oryzae* pv. *oryzae*. The CDS (coding domain sequence) and full-length cDNA sequences of the completely sequenced genes were obtained through the access from public databases of NCBI (National Center for Biotechnology Information) and KOME (Knowledge-based Oryza Molecular Biological Encyclopedia) and they were used to identify each LRR fragment in translated protein sequences through Pfam protein families database. Results showed that only *Xa1*, *Xa21* and *Xa26* genes were found to possess LRRs in their protein sequences. When bioinformatics and statistical methods were used for further comparative analysis of distribution and amino acid composition for these LRRs among *Xa* protein sequences, LRRs of *Xa1* were loosely located after NB-ARC (nucleotide-binding adaptor shared by APAF-1, R proteins, and CED-4) in the protein sequences, whereas LRRs of *Xa21* and *Xa26* were densely located before NB-ARC or P-kinase (protein-kinase). It appeared that the first one or two LRRs of the gene *Xa1* proteins in proximity to NB-ARC were distant from the other LRRs, but no similar phenomenon was observed within the protein sequences of the *Xa21* and *Xa26* genes. The LRRs of these three *Xa* genes mainly consisted of hydrophilic amino acids, with the highest frequency for leucine, followed by neutral serine. The LRRs of *Xa1* protein sequences had different frequencies for some amino acids compared to LRRs of *Xa21* and *Xa26*. These findings are useful in understanding of the structural variation of LRRs in protein sequences among *Xa* genes with resistance to bacterial blight of rice.

**Key words:** Rice, *Xa* genes, Bacterial blight, *Xanthomonas oryzae* pv. *oryzae*, Leucine-rich repeat, Amino acids.

- 
1. Contribution No.2354 from Taiwan Agricultural Research Institute (TARI), Council of Agriculture. Accepted: June 26, 2009.
  2. Graduate Student, Graduate Institute of Biotechnology, Chaoyang University of Technology, Wufeng, Taichung, Taiwan ROC.
  3. Assistant Researcher, Crop Science Division, TARI, Wufeng, Taichung, Taiwan, ROC.
  4. Senior Researcher and Director, Technical Service Division, TARI, Wufeng, Taichung, Taiwan, ROC.
  5. Corresponding author, e-mail: iying@tari.gov.tw; Fax: (04)23325176.