

쌀 저장 단백질 프로라민 유전자 암호 분석

이태호 · 김주곤* · 남백희

명지대학교 이과대학 생물학과, *농촌진흥청 농업 유전공학 연구소

초록: 쌀의 주요 저장 단백질중의 하나인 알콜용해성 프로라민의 성질을 분석하고 이들을 동정하기 위하여 유전자 database로부터 얻은 17개의 프로라민 유전자의 염기서열을 상호비교 분석하였다. 유전자로부터 유추된 단백질 서열의 다중분석결과 프로라민들은 계통발생적으로 type I에서 type IV의 4개의 군으로 크게 분류할 수 있었다. 이러한 분류는 아미노산 서열 중간과 카복실 말단쪽의 짧은 결손에 의해 여러 형태의 아미노산 서열에 의한 것임을 알 수 있었다. 1군에서 4군까지의 군들은 methionine과 cysteine과 같은 황을 포함하는 아미노산이 각각 1, 4, 10, 30%로 구성된 특징을 보여 주었다. 또한 각 군들은 등전점이 9.2, 8.2, 6.7, 7.4인 각군별로 매우 상이한 등전점을 나타내었다. 아울러 GC3s에 대한 유효 암호수(effective codon number, Nc)와 우선 암호수(preferred codon number)의 분석과 상관 그래프를 통해서 프로라민 유전자들의 군별 전이 효율의 차이가 현저하여 프로라민 생산 수준의 군별차이의 가능성을 제시하였다(1993년 11월 5일 접수, 1993년 11월 26일 수리).

지금까지 쌀에서의 저장 단백질의 분획별 분포¹⁻²⁾와 프로라민 단백질 군의 전기등전점에 의한 분포의 비교가 보고된바 있으나,³⁾ 유전자수준에서 이러한 프로라민 유전자군의 유전자산물을 동정하고 쌀에서의 분포를 연관시키지 못하고 있다. 이는 저장단백질이 대부분 유사한 구조를 가지는 다양한 단백질로 구성되어 있으며, 이들의 발현에 관여하는 유전자도 여러개 존재하므로 유전자와 유전자산물인 단백질의 동정이 어렵기 때문으로 알려지고 있다.⁴⁾ 그러나 최근 저장 단백질의 다양한 유전자의 염기서열이 밝혀짐에 따라, 유전자염기서열에서의 유전자 암호 분석과⁵⁻⁶⁾ 이로부터 유추되는 단백질군의 분류와 이들의 성질의 유추와 이러한 성질을 바탕으로 한 비교분석이 가능하게 되었다.

따라서 본 연구에서는 저장 단백질의 질적향상을 위한 저장단백질의 동정과 품종별 분포분석을 통하여 품종개발에 중요한 지표로 이용하기 위한 연구의 일환으로 프로라민 유전자정보를 분석하였다. 프로라민은 쌀의 단백질 총량중 약 10%를 차지하며 분자량은 12~17 kDa 정도인 알콜용해성 저장 단백질중의 하나로, 지금까지 17개의 프로라민 유전자정보가 보고되었다.⁷⁾ 이들 유전자 정보와 이로부터 유추되는 단백질의 성질을 비교, 분석하여 이들성질의 분석지표로 사용하고자 시도하였다.

실험방법

프로라민 유전자들의 DNA염기서열은 EMBL(release 30.0, March, 1992)과 GeneBank(release 74.0, December, 1992) database로부터 구하였다. 이러한 염기서열로부터 아미노산 서열의 유추는 DNASIS(Pharmacia ver. 7.06) Program을 이용하였으며, 단백질의 다중분석과 계통발생학적 분석은 CLUSTAL V Program⁸⁾을 이용하였다. 아울러 단백질의 유사성 및 분자량과 등전점은 PROSIS(Pharmacia ver 7.06)를 이용하여 분석하였다. 또한 유효암호수(effective codon number: Nc)⁹⁾와 유전자 암호에서 세번째 염기가 G와 C로 끝나는 암호의 비율(fraction of G+C at synonymous site in codons: GC3s)은 Computer program "CODONS"에 의해 계산되었다.¹⁰⁾ 우선암호수(preferred codon number)는 지금까지 이용되어온 원리를 이용하여,¹¹⁾ 본 연구진에 의하여 개발된 CODUSA(codon usage analyzer)를 사용하여 구하였다.¹²⁾

결과 및 고찰

프로라민 유전자군의 분류

지금까지 보고된 17개의 프로라민 유전자로부터 유추

된 아미노산서열을 다중분석으로 계통발생적 연관관계를 보면 Fig. 1에서와 같이 크게 4개군으로 분류할 수 있

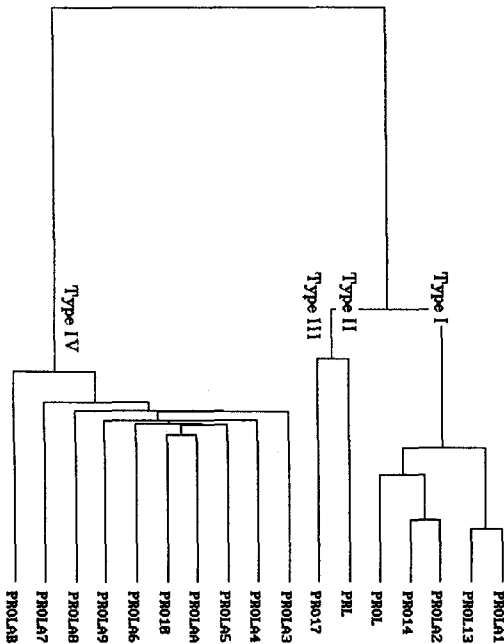


Fig. 1. The phylogenetic relationship of the 17 rice prolamines.

The dendrogram was obtained from the comparison of the amino acid sequence deduced from the published DNA coding sequences.

었다. 이들중 1군에 속하는 다섯개의 유전자는 다시 세 개군으로 나눌 수 있었다. 이러한 프롤라민의 아미노산 서열을 다중배열 하여보면, Fig. 2에서 보는 바와 같이 1군, 2군, 3군의 프롤라민은 각각 군에 따라 몇개의 아미노산이 절단된 형태의 배열을 보여주고 있으며, 부분적으로 아미노산의 치환 되어 있음을 알 수 있었다. 반면 4군의 프롤라민은 군내의 단백질 상호간에 전체적으로 매우 유사한 아미노산 서열을 보여 주고 있다. 단지 C-말단 아미노산 서열의 차이를 보여주고 부분적으로 아미노산의 치환이 나타나고 있다.

이들의 염기서열은 Table 1에서 보는 바와 같이 80%내외의 유사성을 보여주었다. 반면 2군과 3군은 1군과 비교하여 볼 때 각각 단계적으로 65%, 55% 내외의 유사성을 보여주었다. 이들의 유사성을 단백질 수준에서 비교하여 보면 각 군별로, 75~85%, 70%, 60%의 수준으로 염기서열의 유사성과 동일한 수준의 유사성을 보여주고 있으며, 유사한 성질을 가지는 아미노산을 동일한 아미노산으로 고려하여 유사성을 비교하여 보면, 10% 정도씩 증가하는 경향을 보여주고 있다. 그러나 제 4군에 속하는 유전자군은 다른 군의 유전자들과는 유사성이 없으며, 유사아미노산을 동일 아미노산으로 고려하여야 50% 수준의 유사성만을 보여주는 독특한 군임을 알 수 있었다.

프롤라민 단백질군의 성질

프롤라민 아미노산 서열의 다중분석결과 얻어진 4군의

Table 1. The sequence homology of 4 types of rice prolamine genes and proteins

Type	Name	I			II	III
		PROLA1	PROLA2	PROL	PRL	PRO17
Gene	PROLA1					
	I PROLA2	83.7				
	PROL	78.7	89.0			
	II PRL	66.7	67.6	65.6		
	III PRO17	56.4	57.0	51.4	70.4	
	IV PRO10	—*	—*	—*	—*	—*
Protein	PROLA1					
	I PROLA2	84(90)**				
	PROL	75(83)	88(92)			
	II PRL	69(77)	69(79)	64(76)		
	III PRO17	60(71)	58(70)	53(66)	71(82)	
	IV PRO10	27(51)	26(49)	27(48)	28(52)	32(56)

The value are expressed as percentage.

*indicates no homology.

**The value in parenthesis shows the degree of homology when the equivalent amino acid are considered as the homologous amino acids.

FROL	MKIIIF--VFALLAIAAC RPLPSLMFLGQSYRQYQLQSPVLLQQQVLSFYNEFVRQQYGIA	57
PRO14	MKIIIF--VFALLAIAAC SASAQFDVVGQSYRQYQLQSPVLLQQQVLSFYNEFVRQQYGIA	58
PROLA2	MKIIIF--VFALLAIAAC RP-LQFDVVGQSYRQYQLQSPVLLQQQHVLSFYNEFVRQQYGIA	57
PROLA1	MKIIIF--VFALLAIAAC -ATAQFDVVGQINRQYQVQSPLLLQQQVLSLYNEFVRQQYSIA	57
PRL	MKIIIF--VFALLAIVACN-RS RFDFLSQSYRQYQLQSHLLLQQQVLSFYNEFVRQQYSIV	57
PRO17	MKIIIF--FFALLAEAA CASAQFDVAVTQVYRQYQL-----QQQMLSPCGEFVRQQCSTV	52
PRO10	MAAYTSKIFALFALIALSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	54
PROLAA	----- ALLPLIALSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	45
PROLA5	----- FALIALSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	43
PROLA6	----- ALFALIALSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	45
PROLA4	----- LFALIALSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	44
PROLA9	----- LFALIXLSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	44
PROLA3	----- IALSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	40
PROLA8	----- IALSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	40
PROLA7	----- LSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	39
PROLAB	----- LFALIXLSASATTA ITTMQY-----FPPTLAMGMTMDPCRQYMMQTLGGMG	45
*	
PROL	ASPFLQSAAFQLRNNQVWQH-----QA-GGQQSRYQDINIVQAIAYELQLQQFGDLYFDR	111
PRO14	ASPFLQSAAFQLRNNQVWQ-----LALVAQQSHYQDINIVQAIQQQLQLQQFGDLYFDR	113
PROLA2	ASPFLQSAAFQLRNNQVWQ-----LALVAQQSHYQDINIVQAIQQQLQLQQFGDLYFDR	112
PROLA1	ASPFLQSAVFLRNNQVWLQ-----LRLVAQQSHYQDINIVQAIQQQLHLQQFGDLYIDR	112
PRL	ATPEWQPATFQLINNVQVQQCCQQLRLVAQQSHYQAI SIVQAI VQQLQLQQFSGVYFDQ	117
PRO17	ATPEFQSPVFQLRNCQVMQQCCQQLRMIAQQSHCQAI SSVQAI VQQLQLQQFSGVYFDQ	112
PRO10	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	103
PROLAA	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	94
PROLA5	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	92
PROLA6	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	94
PROLA4	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	93
PROLA9	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	93
PROLA3	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	89
PROLA8	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	89
PROLA7	SS----TAMFMSQPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----ICAGL	88
PROLAB	SS----TTMFMSPMALLQQCCMQLQGMPQCHCGTSCQMMQSMQQV-----VCAGL	94
 * .. * . *	
PROL	NQAQAQALLAFNVPSRYGIYPRYYGAPSTITTLGGVL-	148
PRO14	NLAQAQ--LAFNVPSRYGIYPRYYGAPSTITTLGGVL-	148
PROLA2	NLAQAQALLAFNVPSRYGIYPRYYGAPSTITTLGGVL-	149
PROLA1	NLAQAQRLLAFNLPSTYGIYPRYYRAPGSTITTLGGVLY	150
PRL	TQAQAQTLTFLNLP SICGIYPNYYSAPRSIATVGGVWY	155
PRO17	AQAQAQAMGLNLP SICGIYPSYNTVP-EIPTVGGI WY	149
PRO10	GQQQMM-KMAMQMPYMCNMAFVNFQLSSCGC-----C	134
PROLAA	GQQQMM-KMAMQMPYMCNMAFVN-----	116
PROLA5	GQQQMM-KMAMQMPYMCNMAFV-----	113
PROLA6	GQQQMM-KMAMQMPYMCNMAF-----	114
PROLA4	GQQQMM-KMAMQMPYMCNMAH-----	113
PROLA9	GQQQMM-KMAMQMPYMCNMAF-----	113
PROLA3	GQQQMM-KMAMQMPYMCNMAFV-----	110
PROLA8	GQQQMM-KMAMQMPYMCNMAF-----	109
PROLA7	GQQQMM-KMAMQMPYMCNMAF-----	108
PROLAB	GQQQMMKAMQMPYMCNMAF-----	115
	.. * *	

Fig. 2. Multiple alignment of 16 prolamine protein sequences from rice. The signal sequences are shown in bold letters. The character '*' shows perfectly conserved positions, and character '.' well conserved positions.

단백질 성질을 각 균별로 비교하여 보면 Table 2와 같다. 각 균에 있어서의 특이한 점은 glutamine이 균별로 18~22%로 가장 높은 함량을 유지하고 있으며, 소수성 아미노산인 valnine, leucine, isoleucine, phenylalanine, ty-

rosine 등의 함량이 전체의 37%, 34%, 30%, 15%의 수준으로 점차 감소하는 특징을 보여주고 있다. 또한 arginine의 함량도 5% 수준에서 1% 수준으로 감소하는 경향을 보여주고 있다. 또한 산성 아미노산인 aspartic

acid는 각군별로 2.5%, 1.5%, 1.0% 내외, glutamic acid는 1% 내외의 수준을 유지하고 있었다. 반면 각군에서의 meththionine과 cysteine과 같은 황을 포함하는 아미노산 구성비율은 각각 1, 4, 10, 30% 수준을 보여주고 있어서, 군에 따라 쌀에서의 황함유아미노산의 주요한 공급원으로 프롤라민이 이용되고 있음을 알 수 있었다. 이는 황함유아미노산의 함량이 매우 부족한 쌀에서의 단백질 효용가치를 증가시킬 수 있는 매우 중요한 점이라 하겠다. 이들의 분자량을 군별로 비교하여 보면, 각군별로 15.5, 16.4, 15.2, 12.0 kDa 등으로 4군의 프롤라민을 제외하고는 1, 2, 3군은 큰 차이를 보여 주고 있지 않아, 분자량의 차이에 의한 프롤라민의 비교분석에는 어려움이 있을 것으로 예상된다. 반면 이들 각각의 등전점을 살펴보면 9.2, 8.2, 6.7 그리고 7.4의 독특한 등전점을 나타내고 있어, 등전점 차이에 의한 프롤라민 군의 분류가 가능할 것으로 여겨지고 있다. 이는 품종에 따르는

프롤라민 단백질의 분포분석에 있어 매우 중요한 단서로 전기 등전점에 의한 프롤라민의 비교분석에 이용될 수 있는 매우 중요한 지표라 하겠다.

프롤라민의 아미노산 서열에 의한 각군의 소수성 양상을 분석하여 보면 Fig.3과 같이 각군이 매우 유사한 형태를 보여주고 있다. 1, 2, 3군에서의 소수성 양상은 단백질의 유사성정도의 차이에 관계없이 매우 유사하게 나타나고 있으며, 4군에서의 소수성 양상 또한 카복시 말단 부분을 제외하고는 전체적으로 매우 유사한 양상을 보여 주고 있어 아미노산 서열에서 유사성이 나타나지 않으나, 단백질의 3차원적 입체구조가 매우 유사한 것이 아닌가 유추된다. 이는 저장 단백질의 축적에 있어 프롤라민이 가지는 고유한 구조가 군별로 매우 유사한 구조를 가지는 것으로 유추할 수 있겠다.

프롤라민 유전자 암호의 분석

Table 2. The Characteristics of 4 Types of Rice Prolamines

Type	I		II		III		IV	
	PROL	PROLA2	PROLA1	PRL	PRO17	PRO10	PROLA3	PROLAB
Gly	6.67	5.19	4.41	2.84	4.44	6.76	6.48	6.48
Ala	8.89	9.63	8.82	6.38	8.15	7.63	8.33	7.41
Val	5.19	6.67	7.35	7.09	8.15	1.69	1.85	1.85
Leu	11.85	13.33	14.71	9.93	6.67	6.78	6.48	6.48
Ile	4.44	4.44	5.15	5.67	5.19	1.69	1.85	0.93
Ser	6.67	5.93	5.88	9.22	8.15	7.63	6.48	6.48
Thr	2.22	2.22	2.94	4.26	3.70	7.63	8.33	9.26
Cys	0.74	0.74	0.74	3.55	5.93	9.32	7.41	7.41
Met	0.74	0.00	0.00	0.71	2.96	18.64	20.37	21.30
Asp	2.22	2.96	2.94	1.42	1.48	0.85	0.93	0.93
Asn	4.44	4.44	5.15	3.55	2.22	1.69	0.93	0.93
Glu	1.48	0.74	0.74	0.71	1.48	0.00	0.00	0.00
Gln	17.78	18.52	18.38	21.28	22.22	16.10	16.67	16.67
Arg	5.93	5.19	5.88	4.26	2.96	0.85	0.93	0.93
Lys	0.00	0.00	0.00	0.00	0.00	0.85	0.93	0.93
His	0.74	1.48	1.47	1.42	0.74	0.85	0.93	0.93
Phe	5.19	5.19	4.41	4.96	5.19	2.54	1.85	1.85
Tyr	8.15	7.41	7.35	6.38	4.44	2.54	2.78	2.78
Trp	0.74	0.74	0.00	1.42	0.74	0.00	0.00	0.00
Pro	5.93	5.19	3.68	4.96	5.19	5.93	6.48	6.48
M.W.**	15.4	15.4	15.6	16.3	15.2	13.0	11.9	12.0
PI	9.45	9.19	9.50	8.41	6.67	7.32	7.47	7.47
Source***	sat	sat	sat	sat	sat	sat	ruf	lon
	Tainung 67			Nipponbare				

*The amino acids are expressed in 3 letter code and their compositions are shown in Mol percentage.

**The molecular weight of each proteins are expressed in kilodation.

***Abbreviations used: sat; Oryzae sativa, ruf; rufipogon, lon; longistaminata

4군으로 분류된 17개의 유전자들 중에서 각 군을 대표하는 4개의 유전자 암호를 분석하여 각 유전자 암호의 사용빈도를 보면 Table 3에서 보아 알 수 있듯이, 각 군의 유전암호 사용의 빈도는 다르게 나타나고 있다. 각 군별로 나타난 유전자암호 사용빈도를 종합하여보면 프롤라민을 구성하는 유전자 암호중 6개의 암호를 가지고 있는 leucine과, 4개의 암호를 가지고 있는 valine, proline, alanine, 2개의 암호를 가지고 있는 tyrosine, histidine, glutamine, phenylalanine 등의 유전자암호의 사용은 암호사이의 다소 차이는 있지만 전체적으로 유사한 수준의 사용빈도를 보여주고 있다. 반면, 6개의 암호를 사용하는 arginine, serine, 4개의 암호를 사용하는 threonine, 3개의 암호를 사용하는 isoleucine, 2개의 암호를 사용하는 asparagine, aspartic acid, glutamic acid, cysteine 등은 유전자 암호사용의 편중현상이 두드러지게 나타나고 있는데, 그 특징을 살펴보면 Table 4와 같다. Table에서 보아 알 수 있듯이 사용된 총암호의 수(total

codons used)와 우선암호수(preferred codon numbers)는 3군과 4군의 프롤라민 유전자에서 다소 낮게 나타나고 있다. 이는 분자량이 작은 4군에서 두드러지게 나타나는 것으로 보아, 각군에서의 총아미노산의 수에 의한 것으로 여겨진다. 그러나, 유효암호수(effective codon number, Nc)에 있어서는 2군과 3군이 가장 낮은 값인 47개 정도를 보여준 반면, 1군은 50~55개 수준, 4군은 유효암호수의 최대값인 61(정지암호 3개 제외)을 보여주고 있다. 반면 사용된 암호에서의 3번째 염기가 G 또는 C인 암호의 비율(fraction of G/C ending codons at synonymous site, GC3s)은 1, 2군이 0.45 수준을 나타내주며, 3, 4군은 0.55 내외의 수준을 보여주고 있다. 이는 각각 사용암호의 빈도에 있어, 1, 2군이 A/T로 끝나는 암호의 사용빈도가 다소 높은 반면, 3, 4군은 G/C로 끝나는 암호의 사용빈도가 다소 높게 나타나고 있다. 이러한 각군에서의 유효암호수와 GC3s의 값을 도식화 하여보면 Table 4와 같이 각군별로 모두 암호사용에서의 GC3s 값으로 추정되는 암호의 편중현상은 매우 유사한 수준으로 나타나나, 유효암호수는 각군별로 그룹을 이루고 있음을 알 수 있다. 이러한 현상은 유전자의 발현에 있어 사용암호의 편중도가 유사한 수준에서 전이(translation) 효율은 유효암호수에 의하여 비교될 수 있다는 이론⁹⁾에 입각하여

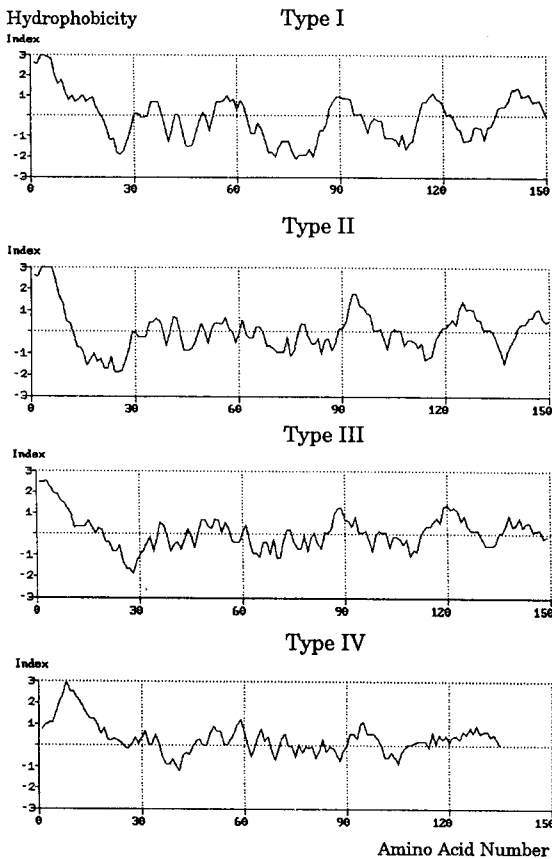


Fig. 3. Hydrophobicity of 4 types of rice prolamines. The degree of hydrophobicity is expressed in (+) values, and that of hydrophilicity in (-) values.

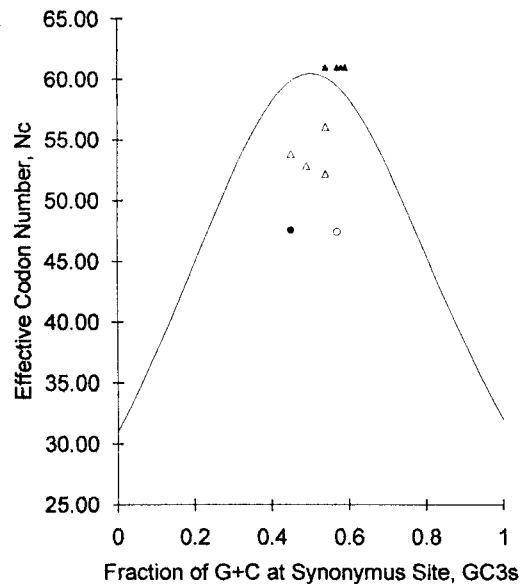


Fig. 4. The plot of fraction of G/C synonymous (GC 3s) versus effective codon numbers (Nc) of 4 types of rice prolamine genes.

The effective codon numbers and GC3s of 4 types of prolamine genes are denoted as follows: Δ , type I; \circ , type II; \bullet , type III; \blacktriangle , type IV.

불 때, 1, 2군의 전이 효율이 3, 4군의 전이 효율보다 높을 것으로 추정될 수 있다. 따라서 프롤라민 단백질의 생산 수준이 군별로 현격한 차이를 보여 줄것으로 예상된다. 따라서 3, 4군에 속하는 황함유 아미노산의 수준이 매우 높은 유용한 프롤라민 유전자의 발현은 우선

암호수가 작으나 유효암호수가 높아 단백질 전이 효율에 있어 불리하기 때문에 그 생합성 수준이 낮을 것으로 추정된다.

이상의 여러가지 프롤라민 유전자군의 유전자 암호번호는 이로부터 발현되는 단백질의 여러가지 성질 비교는

Table 3. The frequency of codons used in rice glutelin genes

Type	I	II	III	IV	Frequency	Type	I	II	III	IV	Frequency		
Codon	PROL	PRL	PRO17	P R O - LAB	Total	Codon	PROL	PRL	PRO17	P R O - LAB	Total		
					%						%		
Leu-TTA	1	0	0	2	3	5%	Gly-GGT	6	4	4	1	15	58%
TTG	4	3	2	1	10	18%	GGC	3	0	2	4	9	35%
CTT	3	2	2	2	9	16%	GGA	0	0	0	1	1	4%
CTC	3	4	2	2	11	20%	GGG	0	0	0	1	1	4%
CTA	3	4	3	1	11	20%	Ile-ATT	5	6	4	1	16	52%
CTG	4	3	2	2	11	20%	ATC	2	2	4	1	9	29%
Arg-CGT	0	0	0	0	0	0%	ATA	2	3	1	0	6	19%
CGC	1	1	0	0	2	11%	Tyr-TAT	7	4	1	1	13	45%
CGA	0	0	0	0	0	0%	TAC	4	5	5	2	16	55%
CGG	1	1	0	0	2	11%	His-CAT	1	1	0	0	2	40%
AGA	2	0	1	0	3	16%	CAC	0	1	1	1	3	60%
AGC	2	0	1	0	3	16%	Ser-TCT	2	2	3	1	8	20%
Ser-TCT	2	2	3	1	8	20%	TCC	0	1	0	1	2	5%
TCC	0	1	0	1	2	5%	TCA	1	0	2	0	3	8%
TCA	1	0	2	0	3	8%	TCG	1	1	0	1	3	8%
TCG	1	1	0	1	3	8%	TGT	3	6	1	2	12	30%
TGT	3	6	1	2	12	30%	AGC	2	3	5	2	12	30%
AGC	2	3	5	2	12	30%	Val-GTT	1	3	3	2	9	27%
Val-GTT	1	3	3	2	9	27%	GTC	4	3	4	0	11	33%
GTC	4	3	4	0	11	33%	GTA	1	3	1	0	5	15%
GTA	1	3	1	0	5	15%	GTG	2	3	3	0	8	24%
GTG	2	3	3	0	8	24%	Pro-CCT	3	2	2	2	9	31%
Pro-CCT	3	2	2	2	9	31%	CCC	2	2	2	0	6	21%
CCC	2	2	2	0	6	21%	CCA	2	3	2	4	11	38%
CCA	2	3	2	4	11	38%	CCG	1	0	1	1	3	10%
CCG	1	0	1	1	3	10%	Thr-ACT	0	3	2	4	9	38%
Thr-ACT	0	3	2	4	9	38%	ACC	3	2	2	3	10	42%
ACC	3	2	2	3	10	42%	ACA	0	0	1	2	3	13%
ACA	0	0	1	2	3	13%	ACG	0	1	0	1	2	8%
ACG	0	1	0	1	2	8%	Ala-GCT	9	5	7	1	22	42%
Ala-GCT	9	5	7	1	22	42%	GCC	1	3	3	4	11	21%
GCC	1	3	3	4	11	21%	GCA	4	3	2	3	12	23%
GCA	4	3	2	3	12	23%	GCG	2	1	3	1	7	13%
GCG	2	1	3	1	7	13%	Cys-TGT	0	2	1	4	7	32%
Cys-TGT	0	2	1	4	7	32%	TGC	1	3	7	4	15	68%
TGC	1	3	7	4	15	68%	Phe-TTT	6	5	4	1	16	52%
Phe-TTT	6	5	4	1	16	52%	TTC	3	4	6	2	15	48%
TTC	3	4	6	2	15	48%	Trp-TGG	1	2	1	0	4	100%
Trp-TGG	1	2	1	0	4	100%	Met-ATG	2	2	5	23	32	100%
Met-ATG	2	2	5	23	32	100%							

Table 4. The characteristics of codon usage pattern of 4 types of prolamine genes

Type Codon	I			II	III	IV	
	PROLA1	PROLA2	PROL	PRL	PRO17	PROLAB	PRO10
Total AA	150	149	149	155	149	115	134
Codons							
Used	49	48	49	48	47	43	44
Preferred	43	40	38	41	37	35	38
Nc	55.69	52.24	49.88	47.56	47.87	61.00	61.00
GC3s	0.54	0.49	0.45	0.45	0.57	0.54	0.57

물론 단백질 전이효율의 유추를 가능케하여 유사한 단백질로 구성되는 프롤라민 단백질군의 동정과 그 분포 분석에 매우 중요한 지표로 이용될 수 있다고 본다. 특히 황 함량이 높은 프롤라민 유전자 정보 분석에 의한 동정과 이들의 발현 수준의 비교는 쌀 저장 단백질의 질적향상을 위한 연구에 새로운 접근방향을 제시할 수 있다고 본다.

참 고 문 헌

- Kim, S. I. and Jo, D. H.: J. Kor. Agr. Chem. Soc., 26 : 65(1983)
- Yoon, H. H. and Kim, S. H.: J. Kor. Agr. Chem. Soc., 32 : 85(1989)
- Guo, Y. J., R. Bishop, H. Fehrnstrom, G. Z. Yu, Y. N. Lian and S. D. Huang: Cereal chem., 63 : 1 (1986)
- Kim, W. T. and Okita, T. W.: Plant Physiol., 88 : 649 (1988)
- Shyur, L. F. and Chen, C. S.: Nucleic Acids Res., 18 : 6683(1990)
- Barbier, P. and Ishihama, A.: Plant Mol. Biol., 15 : 191(1990)
- Yamagata, H., Nomura, T., Arai, S., Tanaka, K. and Iwasaki, T.: Biosci. Biotechnol. Biochem., 56 : 537 (1992)
- Masumura, T., Hibino, T., Kidzu, K., Mitsukawa, N., Tanaka, K. and Fujii, S.: Mol. Gen. Genet., 221 : 1(1990)
- Feng, G., Wen, L., Huang, J. K., Shorrosh, B. S., Muthukrishnan, S. and Reeck, G. R.: Nucleic Acids Res., 18 : 683(1990)
- Kim, W. T. and Okita, T. W.: FEBS Lett., 231 : 308 (1988)
- Lloyd, A. T. and P. M. Sharp: Nucleic Acid Res., 20 : 5289(1992)
- Campbell, W. H. and Gowri, G.: Plant Physiol., 92 : 1(1990)
- Wright, F: Gene, 23(1990)
- Lloyd, A. T. and P. M. Sharp: J. Hered., 83 : 239 (1992)
- Brown, A. J. P, Bertram, G., Feldmann, P. J. F., Peggie, M. W. and Swoboda, R. K.: Nucleic Acid Res., 19 : 4298(1991)
- Lee, T. H. and B. H. Nahm: Unpublished

Codon usage analysis of rice prolamine genes

Lee Tae Ho, Ju-Kon Kim* and Baek Hie Nahm (Department of Biology, MyongJi University, YongIn, Korea 449-728 and *Agricultural Biotechnology Institute, RDA, Suwon, Korea 441-707)

Abstract : To characterize the prolamines in rice cultivars, the complete coding sequences of 17 prolamine genes from the database were analyzed. According to the phylogenic analysis of the sequences, these genes could be classified into 4 groups, Group I to IV. The multiple alignment of the deduced amino acid sequences revealed that the four groups differ from one another in chain length caused by deletion of short internal amino acids or carboxyl terminal fragments. Each group was also found to have different amino acid composition with 1, 4, 10 and 30% of sulfur containing amino acids (methionine and cysteine) in Group I to IV prolamines, respectively. Also the isoelectric points of these groups showed the different values of 9.2, 8.2, 6.7 and 7.4. Finally, from the analysis of codon usage pattern of prolamine genes, the codon usage for arginine, serine, threonine, isoleucine, asparagine, aspartic acid, glutamic acid and cysteine were highly biased. In the analysis of the codon usage pattern, the relation of the fraction of G/C ending codons to effective codon numbers suggests the different translational efficiency in the expression of the prolamine multigenes.