

쌀 저장 단백질 글루텔린 유전자 암호 분석

신윤철 · 김주곤* · 남백희

명지대학교 생물학과, *농촌진흥청 농업유전공학 연구소

초록 : 글루텔린은 쌀 종자단백질의 80% 수준에 이르는 산 또는 알칼리에 용해되는 가장 중요한 저장 단백질로, 분자량은 54 kDa 내외이며 22kDa의 산성사슬과 32kd의 염기성사슬로 구성되어 있다. 지금까지 보고된 13개의 글루텔린 유전자의 염기서열로부터 유추된 단백질 서열의 다중분석에 의하여 계통발생적으로 1군에서 5군까지 5개군으로 나눌 수 있다. 각 군 상호유전자간의 염기서열과 아미노산서열의 유사성은 60%에서 90%에 이르어 이들 각 군은 매우 유사한 아미노산 조성을 보여주고 있다. 그러나 lysine 함량에 있어서는 2.5%에서 3.6%로 약간의 변화를 보여주고 있는데, 이는 arginine이 point mutation에 의하여 lysine으로 변화된 결과임을 알 수 있었다. 각 군에서의 산성단백질과 염기성 사슬의 등전점은 각각 9.0, 10.0 내외로 매우 유사한 반면, 산성사슬은 각각 6.6, 6.7, 7.2, 8.4, 7.9의 독특한 값을 가지고 있다. 아울러 유전자 암호에서 세번째 염기가 G와 C로 끝나는 암호의 비율(fraction of G+C at synonymous site in codons: GC3s)과 유효 암호수(effective codon number, Nc), 우선 암호수(Preferred codon number)의 분석과 상관그래프는 글루텔린 유전자들의 전이효율의 차이는 거의 없어 유사한 수준으로 발현될 가능성을 제시하고 있다.(1993년 11월 5일 접수, 1993년 11월 26일 수리).

지금까지 쌀에서의 저장 단백질의 분획별 분포¹⁻²⁾와 글루텔린 단백질 군의 전기등전점에 의한 비교가 보고된 바 있으나,³⁾ 유전자수준에서 유전자산물을 동정하고 쌀에서의 분포를 연관시키지 못하고 있다. 이는 저장단백질이 대부분 유사한 구조를 가지는 단백질군으로 구성되어 있으며, 이들의 발현에 관여하는 유전자도 유전자군으로 여러개 존재하기 때문에 유전자와 유전자산물인 단백질의 동정이 어려운 것으로 알려지고 있다.

산성 또는 염기성 단백질로 알려진 글루텔린은 총 쌀단백질의 80%를 차지하는 쌀에서 매우 중요한 저장 단백질이다. 미성숙 단백질 및 산성과 염기성사슬의 분자량은 각각 56 kDa, 32 kDa, 22 kDa으로,⁴⁾ 지금까지 13개의 글루텔린 유전자의 염기서열이 밝혀진 바 있다.⁵⁻¹⁵⁾

이러한 글루텔린 유전자군의 다양한 유전자의 염기서열이 밝혀짐에 따라 본 연구에서는 유전자와 유전자산물인 단백질간을 연관짓기 위하여 지금까지 발표된 13개의 글루텔린 유전정보를 분석하고, 이로부터 유추되는 단백질의 성질을 비교, 분석하였다.¹⁶⁻¹⁷⁾ 이러한 분석은 쌀에서의 저장단백질의 분포와 이에 따르는 품종개량에 있어 중요한 지표로 이용될 수 있을 것으로 여겨진다.

실험방법

글루텔린 유전자들의 DNA염기서열은 EMBL¹⁸⁾과 GeneBank¹⁹⁾ DNA Sequence database로부터 구하였다. 이러한 염기서열로부터 유추된 아미노산 서열은 DNASIS (Pharmacia ver. 7.06)를 이용하였으며, 다중분석과 계통발생학적 분석은 CLUSTAL program²⁰⁾에 의해서, 단백질의 유사도, 분자량, 등전점은 PROSIS(Pharmacia ver 7.06)를 이용하여 분석하였다. 또한 유효암호수(effective codon number: Nc)²¹⁾와 유전자 암호에서 세번째 염기가 G와 C로 끝나는 암호의 비율(fraction of G+C at synonymous site in codons: GC3s)은 computer program인 CODONS에 의해 계산되었다.²²⁾ 우선암호수(Preferred codon number)는 지금까지 사용되어진 원리들을 이용하여,²³⁾ 본 연구진에 의하여 개발된 CODUSA(Codon Usage Analyzer)를 사용하여 구하였다.²⁴⁾

결과 및 고찰

글루텔린 유전자군의 분류

지금까지 발표된 13개의 글루텔린 유전자로부터 유추

된 아미노산서열을 다중분석으로 계통발생학적 연관관계를 그려보면 Fig. 1에서와 같이 크게 5개군으로 나눌 수 있다. 이러한 글루텔린의 아미노산서열을 다중배열하여 보면, Fig. 2에서 보는 바와 같이 글루텔린의 다양한 구조는 사슬 길이가 조금씩 변하거나 치환에 의한 아미노산의 변이에 의한 것이나, 이들은 거의 유사하였다.

이들의 염기서열은 Table 1에서 보는 바와 같이 60~90%의 유사성을 보여주었으며, 이들의 유사성을 단백질 수준에서 비교하여 보아도 60~90% 정도로 염기서열의 유사성과 동일한 수준의 유사성을 보여주고 있다. 한편, 유사한 성질을 가지는 아미노산을 동일한 아미노산으로 고려하여 유사성을 비교하여 보면, 10% 정도씩 다소 증가하는 경향을 보여주었다.

글루텔린 단백질군의 성질

글루텔린 아미노산 서열의 다중분석결과 얻어진 5군의 단백질 성질을 각 군별로 비교하여 보면 Table 2와 같다. 각 군에 있어서 glutamine이 11% 내외로 가장 높은 함량을 보여주었으며, 소수성 아미노산인 valine, leucine, isoleucine, phenylalanine, tyrosine 등의 함량은 전체의 30% 수준으로 유사하였다. 산성 아미노산인 aspartic acid는 각 군별로 3.15%, 4.6%, 2.1%, 2.5%, 2.5% 내외의 수준이고, glutamic acid는 5%의 수준을 유지하고 있었다. 반면, lysine함량은 군별로 2.5에서 3.6% 정도까지 증가하는 경향을 보여주는 반면, arginine은 8% 수준에서 6% 수준으로 감소하는 경향을 보여주고 있다. 이는 다중분석(Fig. 2)에서 보는 바와 같이 각 군에서의 arginine이 lysine으로 point mutation에 의한 것임을 알 수 있다. 이러한 lysine의 변이는 이와 유사한 성질을 가지는 arginine의 암호중 AGA가 AAA로의 변화에 의해 lysine함량이 증가된 것임을 알 수 있겠다(Table 3). 따라서 이러한 변이에 의한 arginine으로부터의 lysine으로의 변환은 앞으로 lysine함량의 인위적인 증가에 이용할 수 있는 한 방안으로 제시될 수 있겠다. 분자량을 군별로 비교하여 보면, 각 군별로 53.7, 53.4, 54.2, 54.0, 53.5 kDa 등으로 각 군간에 큰 차이를 보여 주고 있지 않아, 분자량의 차이에 의한 글루텔린의 비교분석에는 어려움이 있을 것으로 예상된다. 반면 이들 각각의 등전점을 살펴보면(Table 2), 성숙단백질은 8.5~9.5 정도이고, 염기성 사슬의 등전점은 10.0 정도로 군간의 차이는 볼 수 없으나, 산성 사슬은 6.6, 6.7, 7.2, 8.4, 7.9로 차이를 보여주고 있어, 산성 사슬에서의 등전점 차이에 의한 글루텔린 군의 분류가 가능할 것으로 여겨지고 있다. 이는 품종에 따르는 글루텔린 단백질의 분포분석에 있어 매우 중요한 단서로 전기 등전점에 의한 글루텔린의 비교분

석에 이용될 수 있는 매우 중요한 지표라 하겠다.

이들의 아미노산 서열에 의한 각 군의 소수성양상을 분석하여 보면 Fig. 3과 같이 매우 유사한 형태를 보여 주고 있다. 각 군에서의 소수성 양상은 단백질의 유사성 정도와 마찬가지로 매우 유사하게 나타나고 있어 단백질의 3차원적 입체구조 또한 매우 유사한 것이 아닌가 유추된다. 이는 저장 단백질의 축적에 있어 글루텔린이 가지는 고유한 구조가 군별로 매우 유사한 것으로 유추할 수 있겠다.

글루텔린 유전자 암호의 분석

5군으로 분류된 13개의 유전자들 중에서 각 군을 대표하는 5개의 유전자 암호를 분석하여 각 유전자 암호의 사용빈도를 보면(Table 3), 글루텔린을 구성하는 유전자 암호중 4개의 암호를 가지고 있는 valine과 glycine, 3개의 암호를 가지고 있는 isoleucine, 2개의 암호를 가지고 있는 tyrosine, histidine, asparagine, glutamine, phenylalanine 등의 유전자암호사이의 사용빈도는 다소 차이는 있지만 전체적으로 편중되어 있지 않다. 반면, 6개의 암호를 사용하는 leucine, arginine, serine, 4개의

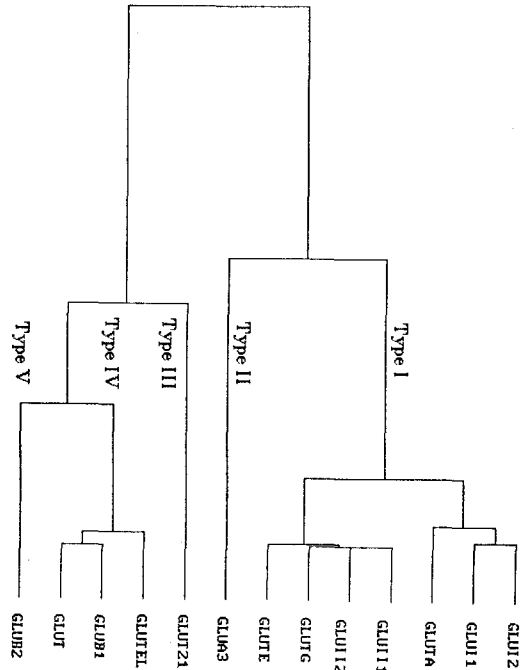


Fig. 1. The phylogenetic relationship of the rice glutelins.

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

Multiple Sequence Alignment of Rice Glutelin

RICGLUI2	MASI--NRPIVFTVCLFLLCNGSLAQQLLGGQTSQWQSSRRGSPRECRFDRLQAFEP	58
RICGLUI1	MASI--NRPIVFTVCLFLLCNGSLAQQLLGGQTSQWQSSRRGSPRECRFDRLQAFEP	58
RICGLUTA	MASI--NRPIVFTVCLFLLCNGSLAQQLLGGQTSQWQSSRRGSPRECRFDRLQAFEP	58
ICGLUI1	MASI--NRPIVFTVCLFLLCDGSLAQQLLGGQTSQWQSSRRGSPRCRFDRLQAFEP	58
ICGLUI2	MASI--NRPIVFTVCLFLLCDGSLAQQLLGGQTSQWQSSRRGSPRCRFDRLQAFEP	58
RICGLUT6	MASI--NRPIVFTVCLFLLCDGSLAQQLLGGQTSQWQSSRRGSPRCRFDRLQAFEP	58
RICGLUT5	MASI--NRPIVFTVCLFLLCDGSLAQQLLGGQTSQWQSSRRGSPRCRFDRLQAFEP	58
RICGLUA3	MATI--KFPVFTVCLFLLCNGSLAQ--LLSQTSQWQSSRRGSPRECRFDRLQAFEP	57
ICGLUT21	MATIAFSRLSIYF--CVLLCHGSMQAQ--LFGPNVNFVHNPRQGGFRECREFDRLQAFEP	57
ICGLUTEL	MASSVFSRFSIYF--CVLLCHGSMQAQ--LFPNSTNPWHSPRQGSFRECREFDRLQAFEP	57
RICGLUB1	MASSVFSRFSIYF--CVLLCHGSMQAQ--LFPNSTNPWHSPRQGSFRECREFDRLQAFEP	57
RICGLUT	MASSVFSRFSIYF--CVLLCHGSMQAQ--LFPNSTNPWHSPRQGSFRECREFDRLQAFEP	57
RICGLUB2	MATIFSRSIYF--CAMLCCGSMQAQ--LFPNSTNPWHSPRQGSFRECREFDRLQAFEP	57
	** * * * * * *	
RICGLUI2	SVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	118
RICGLUI1	SVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	118
RICGLUTA	SVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	118
ICGLUI1	SVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	118
ICGLUI2	SVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	118
RICGLUT6	SVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	118
RICGLUT5	SVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	118
RICGLUA3	TVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	117
ICGLUT21	KVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	117
ICGLUTEL	KVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	117
RICGLUB1	KVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	117
RICGLUT	KVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	117
RICGLUB2	KVRSQAGTTEFFDVSNEQFQCTGVSVRRVIEPRGLLPHYTNGLASLVYIIQGRGITGPT	117
	***** * * * * *	
RICGLUI2	FPGCPESYQQQFQQSGQAQLTESQSQSKFKDEHQKIHFRQGDVIALPAGVAHWYNDG	178
RICGLUI1	FPGCPESYQQQFQQSGQAQLTESQSQSKFKDEHQKIHFRQGDVIALPAGVAHWYNDG	178
RICGLUTA	FPGCPESYQQQFQQSGQAQLTESQSQSKFKDEHQKIHFRQGDVIALPAGVAHWYNDG	178
ICGLUI1	FPGCPETYQQQFQQSGQAQLTESQSQSKFKDEHQKIHFRQGDVIALPAGVAHWYNDG	178
ICGLUI2	FPGCPETYQQQFQQSGQAQLTESQSQSKFKDEHQKIHFRQGDVIALPAGVAHWYNDG	178
RICGLUT6	FPGCPETYQQQFQQSGQAQLTESQSQSKFKDEHQKIHFRQGDVIALPAGVAHWYNDG	178
RICGLUTE	FPGCPETYQQQFQQSGQAQLTESQSQSKFKDEHQKIHFRQGDVIALPAGVAHWYNDG	178
RICGLUA3	FPGCPETYQQQFQQSQEQQQLLEGQSQSKFKDEHQKIHFRQGDVIALPAGVAHWYNDG	177
ICGLUT21	FPGCPATYQQQFQQF----PEGQSQSKFRDEHQKIHFRQGDVIALPAGVAHWYNDG	173
ICGLUTEL	FPGCPATYQQQFQQF----SQGQSQSKFRDEHQKIHFRQGDVIALPAGVAHWYNDG	173
RICGLUB1	FPGCPATYQQQFQQF----SQGQSQSKFRDEHQKIHFRQGDVIALPAGVAHWYNDG	173
RICGLUT	FPGCPATYQQQFQQF----SQGQSQSKFRDEHQKIHFRQGDVIALPAGVAHWYNDG	173
RICGLUB2	FPGCPATYQQQFQQF----SQGQSQSKFRDEHQKIHFRQGDVIALPAGVAHWYNDG	173
	***** * * * * *	
RICGLUI2	EVVVAIYVTDLNNGANQLDPRQDFLLAGNKR--NPQ--AYRREVEERS--QNI FSGFSTE	234
RICGLUI1	EVVVAIYVTDLNNGANQLDPRQDFLLAGNKR--NPQ--AYRREVEERS--QNI FSGFSTE	234
RICGLUTA	EVVVAIYVTDLNNGANQLDPRQDFLLAGNKR--NPQ--AYRREVEERS--QNI FSGFSTE	234
ICGLUI1	EVVVAIYVTDLNNGANQLDPRQDFLLAGNKR--NPQ--AYRREVEERS--QNI FSGFSTE	234
ICGLUI2	EVVVAIYVTDLNNGANQLDPRQDFLLAGNKR--NPQ--AYRREVEERS--QNI FSGFSTE	234
RICGLUT6	EVVVAIYVTDLNNGANQLDPRQDFLLAGNKR--NPQ--AYRREVEERS--QNI FSGFSTE	234
RICGLUTE	EVVVAIYVTDLNNGANQLDPRQDFLLAGNKR--NPQ--AYRREVEERS--QNI FSGFSTE	234
RICGLUA3	DAPIVAIVYTDIYNSANQLDPRHRDFLLAGNKR--NPQ--LYRREVEERS--QNI FSGFSTE	234
ICGLUT21	DAPVVAIVYFDLNNANQLEPRQKFLLAGNKR--EQQMYGRSIEQHSQNI FSGFNNE	231
ICGLUTEL	DAPIVAIVYVYDNNANQLEPRQKFLLAGNKR--EQQMYGRSIEQHSQNI FSGFVNE	233
RICGLUB1	DRHIVAVYVYDNNANQLEPRQKFLLAGNKR--EQQMYGRSIEQHSQNI FSGFVNE	233
RICGLUT	DRHIVAVYVYDNNANQLEPRQKFLLAGNKR--EQQMYGRSIEQHSQNI FSGFVNE	233
RICGLUB2	DASVVAIVYVYDNNANQLEPRQKFLLAGNKR--QVYSSIEQHSQNI FSNFGTE	231
 * * * * *	
RICGLUI2	LLSEALGVSGVARQLQCNQDQGEIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	293
RICGLUI1	LLSEALGVSGVARQLQCNQDQGEIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	293
RICGLUTA	LLSEALGVSGVARQLQCNQDQGEIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	293
ICGLUI1	LLSEALGVSNQVARQLQCNQDQGEIVRVERGLSLLQP--YASLQEQEQGQVSRERYQEG	293
ICGLUI2	LLSEALGVSNQVARQLQCNQDQGEIVRVERGLSLLQP--YASLQEQEQGQVSRERYQEG	293
RICGLUT6	LLSEALGVSNQVARQLQCNQDQGEIVRVERGLSLLQP--YASLQEQEQGQVSRERYQEG	293
RICGLUTE	LLSEALGVSNQVARQLQCNQDQGEIVRVERGLSLLQP--YASLQEQEQGQVSRERYQEG	293
RICGLUA3	LLSEALGVSSGVARQLQCNQDQGEIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	292
ICGLUT21	LLSEALGVNALVAKRLQCNQDQGEIIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	290
ICGLUTEL	MLSEALGINAVAARLQSNQDQGEIIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	289
RICGLUB1	MLSEALGINAVAARLQSNQDQGEIIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	289
RICGLUT	MLSEALGINAVAARLQSNQDQGEIIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	289
RICGLUB2	LLSEALGINTVAARLQSNQDQGEIIVRVEHGLSLLQP--YASLQEQEQGQVSRERYQEG	285
 * * * * *	

Fig. 2. Multiple Alignment of amino acid sequences rice glutelins. The signal sequences in N-terminal region and point mutation of lysine are shown in bold letters. The character '*' shows perfectly conserved positions, and character '.' well conserved positions.

RICGLUI2	QYQOSYGSGCSNGLDETFCTLRVRQNI DNPNRADTYNPRAGRVTNLNTOQMFPILSLVQM	353
RICGLUI1	QYQOSYGSGCSNGLDETFCTLRVRQNI DNPNRADTYNPRAGRVTNLNTOQMFPILSLVQM	353
RICGLUTA	QYQOSYGSGCSNGLDETFCTLRVRQNI DNPNRADTYNPRAGRVTNLNTOQMFPILSLVQM	353
ICGLUII1	GYQOSYGSGC PNGLDETFCTMRVRQNI DNPNRADTYNPRAGRVTNLNSQNFPIILNLVQM	353
ICGLUII2	GYQOSYGSGC PNGLDETFCTMRVRQNI DNPNRADTYNPRAGRVTNLNSQNFPIILNLVQM	353
RICGLUTG	GYQOSYGSGC PNGLDETFCTMRVRQNI DNPNRADTYNPRAGRVTNLNSQNFPIILNLVQM	353
RICGLUTE	GYQOSYGSGC PNGLDETFCTMRVRQNI DNPNRADTYNPRAGRVTNLNTOQMFPIILNLVQM	353
ICGLUA3	QYQKQLQGSGCSNGLDETFCTMRVRQNI DNPNLADTYNPRAGRITYLNGOQKFFIILNLVQM	352
ICGLUT21	QYSEEQQPSTRCNGLDENFCTIKARVNIENPSHADTYNPRAGRITRILNSQKFFIILNLVQL	350
ICGLUTEL	QYSERQQTSSRWNGLEENFCTIKVRVNIENPSRADSYNPRAGRITSVNSQKFFIILNLIQM	349
RICGLUB1	QYSERQQTSSRWNGLEENFCTIKVRVNIENPSRADSYNPRAGRITSVNSQKFFIILNLIQM	349
RICGLUT	QYSERQQTSSRWNGLEENFCTIKVRVNIENPSRADSYNPRAGRITSVNSQKFFIILNLIQM	349
RICGLUB2	QYSEEQQTSSRWNGLEENFCTIKARVNIENPSRADSYNPRAGRISSVNSQKFFIILNLIQM * . * . . * .	345
RICGLUI2	SAVKVNLYQNALLS PFWNINAHSVVI TOGRARVQV VNNNGKT VFNGLRRGQLLI I PQH	413
RICGLUI1	SAVKVNLYQNALLS PFWNINAHSVVI TOGRARVQV VNNNGKT VFNGLRRGQLLI I PQH	413
RICGLUTA	SAVKVNLYQNALLS PFWNINAHSVVI TOGRARVQV VNNNGKT VFNGLRRGQLLI I PQH	413
ICGLUII1	SAVKVNLYQNALLS PFWNINAHSVVI TOGRAQVQV VNNNGKT VFNGLRRGQLLI I PQH	413
ICGLUII2	SAVKVNLYQNALLS PFWNINAHSVVI TOGRAQVQV VNNNGKT VFNGLRRGQLLI I PQH	413
RICGLUTG	SAVKVNLYQNALLS PFWNINAHSVVI TOGRAQVQV VNNNGKT VFNGLRRGQLLI I PQH	413
RICGLUTE	SAVKVNLYQNALLS PFWNINAHSVVI TOGRAQVQV VNNNGKT VFNGLRRGQLLI I PQH	413
RICGLUA3	SAVKVNLYQNALLS PFWNINAHSVVI TOGRARVQV VNNNGKT VFDGELRRGQLLI I PQH	412
ICGLUT21	SATRVNLYQNALLS PFWNVAHSVVI VQGHARVQV VSNLKGKT VFDGVL RPQQLLI I PQH	410
ICGLUTEL	SATRVNLYQNALLS PFWNVAHSVVI LVYMI QGSRVQV VSNFQKT VFDGVL RPQQLLI I PQH	409
RICGLUB1	SATRVNLYQNALLS PFWNVAHSVVI LVYMI QGSRVQV VSNFQKT VFDGVL RPQQLLI I PQH	409
RICGLUT	SATRVNLYQNALLS PFWNVAHSVVI LVYMI QGSRVQV VSNFQKT VFDGVL RPQQLLI I PQH	409
RICGLUB2	DATRVNLYQNALLS PFWNVAHSVVI LVYMI QGSRVQV VSNFQKT VFDGVL RPQQLLI I PQH * .	405
RICGLUI2	YAVVKKAEQREGCAY IAFKTNPNMSVSHIAGKSSI FRALPNDVL ANAYRISREEAQR LKHN	473
RICGLUI1	YAVVKKAEQREGCAY IAFKTNPNMSVSHIAGKSSI FRALPNDVL ANAYRISREEAQR LKHN	473
RICGLUTA	YAVVKKAEQREGCAY IAFKTNPNMSVSHIAGKSSI FRALPNDVL ANAYRISREEAQR LKHN	473
ICGLUII1	YVVVKKAEQREGCAY IAFKTNPNMSVSHIAGKSSI FRALPTDVL ANAYRISREEAQR LKHN	473
ICGLUII2	YVVVKKAEQREGCAY IAFKTNPNMSVSHIAGKSSI FRALPTDVL ANAYRISREEAQR LKHN	473
RICGLUTG	YVVVKKAEQREGCAY IAFKTNPNMSVSHIAGKSSI FRALPTDVL ANAYRISREEAQR LKHN	473
RICGLUTE	YVVVKKAEQREGCAY IAFKTNPNMSVSHIAGKSSI FRALPTDVL ANAYRISREEAQR LKHN	473
RICGLUA3	HVVVKKAEQREGCAY IALKTNPDSMVSHMAGKNSI FRALPDDV ANAYRISREEARR LKHN	472
ICGLUT21	YVVLKKAHEGECQY IAFKTNANAFVSHLAGKNSI FRAMPV DVL ANAYRISREQAARSLKNN	470
ICGLUTEL	YAVLKKAEREGCQY IAIKTNANAFVSHLAGKNSV FRALPVDV ANAYRISREQAARSLKNN	469
RICGLUB1	YAVLKKAEREGCQY IAIKTNANTFVSHLAGKNSV FRALPVDV ANAYRISREQAARSLKNN	469
RICGLUT	YAVLKKAEREGCQY IAIKTNANTFVSHLAGKNSV FRALPVDV ANAYRISREQAARSLKNN	469
RICGLUB2	YAVLKKAHEGECQY IAIKTNANAFVSHLAGKNSV FRALPVDV ANAYRISREQAARSLKNN . . * .	465
RICGLUI2	RGDEFGAFTPIQ - YKSYQDVYNAAE -----	497
RICGLUI1	RGDEFGAFTPIQ - YKSYQDVYNAAESS ----	499
RICGLUTA	RGDEFGAFTPIQ - YKSYQDVYNAAESS ----	499
ICGLUII1	RGDEFGAFTPLQ - YKSYQDVYNVAESS ----	499
ICGLUII2	RGDEFGAFTPLQ - YKSYQDVYNVAESS ----	499
RICGLUTG	RGDEFGAFTPLQ - YKSYQDVYNVAESS ----	499
RICGLUTE	RGDEFGAFTPLQ - YKSYQDVYNVAESS ----	499
ICGLUA3	RGDELGFVTPSHAYKSYQDI --- SVSA ---	496
ICGLUT21	RGEELGAFTPRYQQQYYPGFSNESESE - ALE	500
ICGLUTEL	RGEEHGAFTPRFQQQYYPGFSNESESE - TSE	499
RICGLUB1	RGEEHGAFTPRFQQQYYPGFSNESESE - TSE	499
RICGLUT	RGEEHGAFTPRFQQQYYPGFSNESESE - TSE	499
RICGLUB2	RGEEHGAFTPRFQQQYYPGFSNEAKVSRVM * .	496

Fig. 2. continued.

암호를 사용하는 proline, threonine, alanine, 2개의 암호를 사용하는 glutamine, lysine, aspartic acid 등은 유전자 암호사용의 편중현상이 두드러지게 나타나고 있다. 이러한 유전자 암호의 사용에 있어서의 그 특징을 살펴보면 Table 4에서 보아 알 수 있듯이 사용된 총암호의 수(total codons used)와 우선암호수(preferred codon numbers)가 군에 따라 서로 유사하며, 유효암호수(effective codon number, Nc)에 있어서도 52-56으로 그 값이 비슷한 수준을 보여주고 있다. 또한 사용된 암호에서의 3번째 염기가 G 또는 C인 암호의 비율(fraction of G/C

ending codons at synonymous site, GC3s)도 0.36~0.44 정도로 유사한 수준을 보여주고 있다. 이러한 각 군에서의 유효암호수와 GC3s의 값을 도식화 하여보면 Table 4와 Fig. 4에서와 같이 모든군이 GC3s 값으로 추정되는 암호의 편중현상 및 유효암호수는 매우 유사한 수준을 유지하고 있음을 알 수 있다. 이것을 유전자의 발현에 있어 사용암호의 편중도가 유사한 수준에서 전이(translation)효율은 유효암호수에 의하여 비교될 수 있다는 이론⁹⁾에 입각하여 볼 때, 이 글루텔린 유전자의 전이효율은 높으면서 각 군간의 전이효율은 별 차이가 없을

Table 1. The sequence homology of 5 types of rice glutelin genes and proteins

Type	Name	I	II	III	IV
		GLUTE	GLUA3	GLUT21	GLUB1
Gene	I	GLUTE			
	II	GLUA3	81.2		
	III	GLUT21	68.4	68.1	
	IV	GLUB1	67.5	67.3	83.5
	V	GLUB2	61.7	63.2	83.6
Protein	I	GLUTE			
	II	GLUA3	81(89)		
	III	GLUT21	65(78)	65(78)	
	IV	GLUB1	64(78)	64(78)	83(92)
	V	GLUB2	64(78)	65(78)	81(91)

The value are expressed as percentage.

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

Table 2. The characteristics of 5 types of rice glutelins

Type	AA*	I			II	III	IV	V	
		GLUI2	GLUTA	GLUII2	GLUA3	GLUT21	GLUTEL	GLUB1	GLUB2
Gly	G	7.40	7.16	7.58	7.20	7.14	6.53	6.74	6.57
Ala	A	6.77	6.32	6.32	6.14	7.14	7.16	6.53	6.99
Val	V	7.40	8.00	7.37	7.63	6.51	7.37	7.37	7.63
Leu	L	7.40	7.37	7.16	7.84	8.19	6.74	6.74	6.78
Ile	I	4.44	4.00	4.63	4.87	4.62	5.26	5.26	5.08
Ser	S	7.61	8.42	7.58	7.42	5.25	7.37	7.37	7.63
Thr	T	4.02	4.21	4.21	4.03	3.15	3.58	3.79	3.60
Cys	C	1.69	1.68	1.68	1.69	1.26	1.05	1.05	1.06
Met	M	0.42	0.42	0.84	0.85	1.05	0.84	0.84	0.85
Asp	D	3.17	3.16	3.16	4.66	2.10	2.53	2.53	2.54
Asn	N	6.77	6.53	6.95	5.93	8.40	7.37	7.37	7.42
Glu	E	5.71	5.68	5.37	4.66	6.72	5.68	5.68	5.30
Gln	Q	11.42	11.37	10.95	10.81	11.13	11.58	11.37	11.23
Arg	R	8.25	8.42	7.79	7.63	7.14	6.95	7.16	6.57
Lys	K	2.54	2.53	2.53	2.97	3.15	3.37	3.37	3.60
His	H	1.90	1.89	2.11	2.75	2.31	2.11	2.32	2.33
Phe	F	4.23	4.21	4.42	4.24	5.46	5.26	5.26	5.72
Tyr	Y	4.02	4.00	4.00	4.03	3.57	3.79	3.79	3.81
Trp	W	0.63	0.63	0.84	0.64	0.63	0.84	0.84	0.85
Pro	P	4.23	4.00	4.42	4.03	5.04	4.63	4.63	4.45
M.W.(kDa)**		53.4	53.7	53.7	53.4	54.2	53.9	54.0	53.5
PI m***		9.02	9.15	8.87	8.55	8.97	9.45	9.54	9.53
PI a		6.60	6.60	6.67	6.73	7.20	8.37	8.72	7.94
PI b		10.15	10.25	10.05	10.18	9.87	9.99	9.99	10.18

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

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

***Abbreviations used: pI m, isoelectric point of mature glutelin; pI a, isoelectric point of acidic subunit; pI b, isoelectric point of basic subunit.

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

Type	I	II	III	IV	V	Freq.	Type	I	II	III	IV	V	Freq.
Codon	GLUTE	GLUA3	GLUT21	GLUB1	GLUB2	(%)	Codon	GLUTE	GLUA3	GLUT21	GLUB1	GLUB2	(%)
Leu-TTA	2	5	7	7	6	14	Gly-GGT	7	12	11	11	11	30
TTG	10	12	8	9	8	25	GGC	9	6	11	6	6	22
CTT	7	13	10	8	7	24	GGA	16	14	8	11	12	35
CTC	8	7	4	4	3	14	GGG	5	3	5	5	3	12
CTA	9	3	9	7	7	18	Ile-ATT	9	7	5	10	10	33
CTG	2	1	5	0	3	6	ATC	9	9	14	12	12	45
Arg-CGT	8	8	6	5	4	18	ATA	6	9	5	4	4	22
CGC	5	3	2	3	3	9	Tyr-TAT	10	13	4	8	9	47
CGA	3	2	4	7	6	13	TAC	9	6	14	11	10	53
CGG	1	1	4	4	2	7	His-CAT	6	7	6	8	8	60
AGA	12	12	9	8	7	27	CAC	4	6	6	4	3	40
AGC	9	10	10	8	10	27	Ser-TCT	7	4	7	9	8	19
TCC	4	6	1	2	1	8	TCA	6	3	5	7	7	16
TCA	6	3	5	7	7	16	TCG	0	4	2	2	3	6
TCG	0	4	2	2	3	6	TGT	10	12	5	11	10	27
TGT	10	12	5	11	10	27	AGC	10	7	8	9	10	24
AGC	10	7	8	9	10	24	Gln-CAA	34	43	40	43	48	79
Val-GTT	17	14	13	11	11	36	CAG	18	8	13	11	6	21
GTC	5	10	7	10	11	24	Asn-AAT	16	16	18	14	15	46
GTA	10	10	7	9	9	25	AAC	18	13	22	21	20	54
GTG	5	5	5	7	5	15	Lys-AAA	1	6	3	4	6	27
Pro-CCT	8	10	8	8	7	38	AAG	11	9	12	12	11	73
CCC	4	2	2	3	2	12	Asp-GAT	13	14	8	10	11	78
CCA	6	7	12	8	11	40	GAC	3	8	2	2	1	22
CCG	4	1	2	3	1	10	Glu-GAA	13	9	13	10	9	41
Thr-ACT	9	9	5	6	9	40	GAG	13	13	19	17	16	59
ACC	6	5	4	4	3	23	Cys-TGT	5	4	2	2	2	36
ACA	6	6	5	6	7	32	TGC	5	6	6	5	5	64
ACG	1	0	2	2	0	5	Phe-TTT	7	14	12	15	12	45
Ala-GCT	12	8	13	12	15	36	TTC	17	10	16	13	18	55
GCC	5	5	7	8	6	18	Trp-TGG	4	3	3	4	4	100
GCA	13	18	14	10	13	40	Met-ATG	5	5	7	6	7	100
GCG	2	0	3	3	2	6							

것으로 보인다. 따라서 글루텔린 단백질의 생산 수준은 군별로 거의 동일한 수준에 이를 것으로 예상된다.

이상의 여러가지 글루텔린 유전자군의 암호분석은 이로부터 발견되는 단백질의 성질 비교는 물론 단백질 전이효율의 유추를 가능케 하여 유사한 단백질로 구성되는 글루텔린 단백질군의 동정과 그 분포분석에 매우

중요한 지표로 이용될 수 있다고 본다. 특히, 5개군의 글루텔린 유전자에서 사용된 유전자 암호중 그 사용빈도에 있어 산술평균치보다 높은(즉 6개의 암호의 경우 17%, 4개의 암호의 경우 25%, 3개의 암호의 경우 33%, 2개의 암호의 경우 50% 이상) 암호를 보면, (Fig. 3의 굵은 수치의 암호) 32개의 암호가 편중되어 이용되고

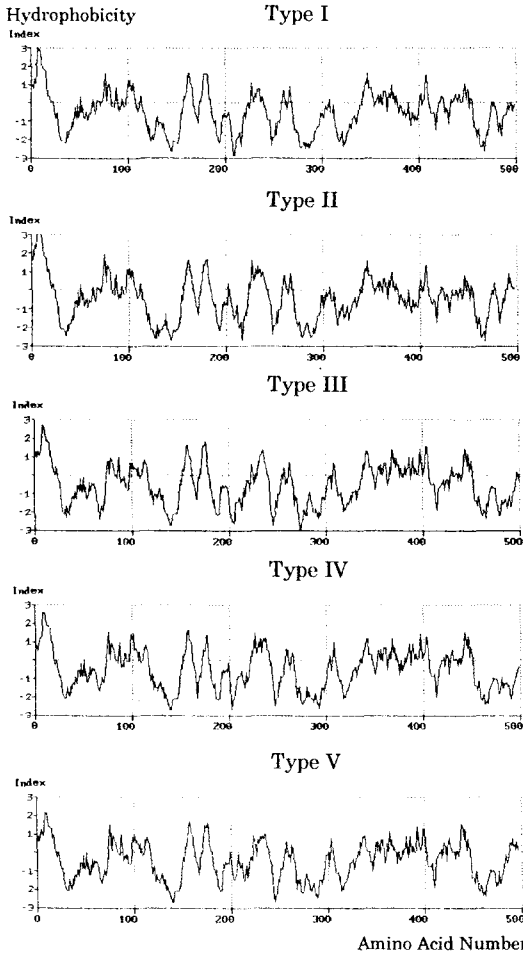


Fig. 3. Hydrophobicity of 4 types of rice glutelins. The degree of hydrophobicity is expressed in (+) values, and that of hydrophilicity in (-) values.

있음을 알 수 있었다. 따라서 이러한 분석을 토대로 하여 글루텔린에서 편중유전자암호를 이용한다면 쌀에서의 외래 유전자발현을 극대화할 수 있는 유전자의 조작이 가능할 것으로 유추된다. 이러한 유전자정보 분석에 의한 동정과 이들의 발현 수준 비교는 쌀 저장 단백질의 질적 향상을 위한 연구에 새로운 접근방향을 제시할 수 있다고 본다.

참 고 문 헌

1. Kim, S. I and Jo, D. H.: J. Kor. Agr. Chem. Soc., 26 : 65(1983)
2. Yoon, H. H. and Kim, S. H.: J. Kor. Agr. Chem. Soc., 32 : 85(1989)

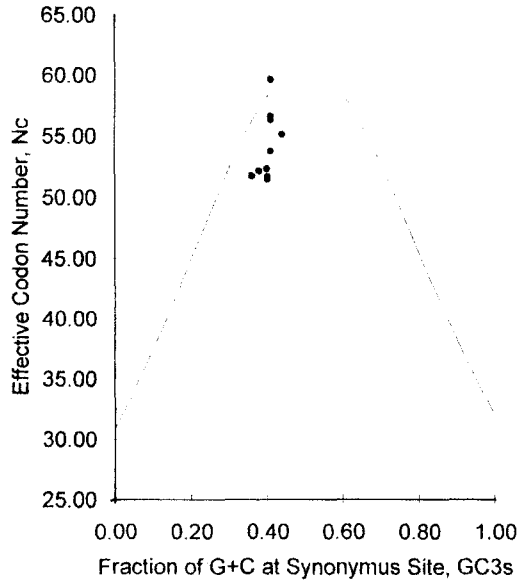


Fig. 4. The plot of Fraction of G/C Synonymous(GC3s) versus effective codon numbers(Nc) of rice glutelin genes.

Table 4. The characteristics of codon usage pattern of rice glutelin genes

Type	I	II	III	IV	V
	GLUTE	GLUA3	GLUT21	GLUB1	GLUB2
Total AA	500	496	500	499	496
Codons					
Used	60	59	61	60	60
Preferred	50	47	46	48	45
Nc	54.37	51.71	55.90	56.56	52.09
GC3s	0.41	0.36	0.44	0.41	0.38

3. Guo, Y. J., R. Bishop, H. Fehrstrom, G. Z. Yu, Y. N. Lian, and S. D. Huang.: Cereal chem., 63 : 1 (1986)
4. Hari B. Krishnan, Thomas W. Okita: Plant Physiol., 81 : 748(1986)
5. Abe, K., Emori, Y., Kawasaki, H., Kondo, H., Suzuki, K. and Arai S.: Agric. Biol. Chem., 53 : 2969(1989)
6. Abe, K. Unpublished, (1990)
7. Higuchi, W. and Fukazawa, C.: Gene, 55 : 245(1987)
8. Kim, S. Y. and Wu, R.: Nucleic Acids Res., 18 : 6845 (1990)
9. Masumura, T., Kidzu, K., Sugiyama, Y., Mitsukawa, N., Hibino, T., Tanaka, K. and Fujii, S.: Plant Mol. Biol., 12 : 723(1989)

10. Takaiwa, F., Kikuchi, S. and Oono, K.: *Mol. Gen. Genet.*, 208 :15(1987)
11. Takaiwa, F., Ebinuma, H., Kikuchi, S. and Oono, K.: *FEBS Lett.*, 221 : 43(1987)
12. Takaiwa, F., Kikuchi, S. and Oono, K.: *Nucleic Acids Res.*, 17 : 3289(1989)
13. Takaiwa, F., Oono, K., Wing, D. and Kato, A.: *Plant Mol. Biol.*, 17 : 875(1991)
14. Takaiwa, F. and Oono, K.: *Jpn. J. Genet.*, 66 : 161 (1991)
15. Wang, C.S., Shastri, K., Wen, L., Huang, J.K., Sonthayanon, B., Muthukrishnan, S. and Reeck, G.R.: *FEBS Lett.*, 222 : 135(1987)
16. Lloyd, A. T. and P. M. Sharp: *Nucleic Acid Res.*, 20 : 5289(1992)
17. Campbell, W. H. and Gowri, G.: *Plant Physiol.*, 92 : 1(1990)
18. EMBL. DNA Sequence Database, R30.0 March (1992)
19. National Institute of Health, Genebank R. 74.0, December (1992)
20. Higgins, D. G. and Sharp, P. M.: *Gene*, 73 : 237 (1988)
21. Wright, F.: *Gene*, 23(1990)
22. Lloyd, A. T. and P. M. Sharp: *J. Hered.*, 83 : 239 (1992)
23. Brown, A. J. P, Bertram, G., Feldmann, P. J. F., Peggie, M. W. and Swoboda, R. K.: *Nucleic Acid Res.*, 19 : 4298(1991)
24. Lee, T. H and B. H. Nahm, Unpublished result (1993)

Codon usage analysis of rice glutelin genes.

Shin, Yun Cheol, Ju-Kon Kim* and Baek Hie Nahm (Department of Biology, MyongJi University, Yongin and *Agricultural Biotechnology Institute, RDA, Suwon, Korea)

Abstract : To characterize glutelins, the most abundant storage protein in rice, 13 complete coding sequences of glutelin genes from the database were analyzed. According to the phylogenic analysis, these genes could be classified into 5 groups, Group I to V. The degrees of homology were calculated to be in the range of 90 to 60%, but the patterns of hydrophobicity were similar in all the groups. Also, each group was found to have similar amino acid composition with variations in lysine content from 2.5 to 3.6% due to the point mutation of arginine to lysine. The isoelectric points of mature proteins and their basic chains of all the groups showed the value of about 9.0 and 10.0, respectively, while the isoelectric points of acidic chains in these groups showed the distinct value of 6.6, 6.7, 7.2, 8.4 and 7.9. The plot of the fraction of G+C at synonymous site in codons (GC3s) against effective codon numbers suggest no major difference in translational efficiency in the expression of glutelin multigenes.