온대자포니카 벼 품종의 등숙기 종자 내 카드뮴 함량에 관한 전장유전체연관분석

이상범^{1,a} · 김규원^{2,a} · 윤민영² · 김경진¹ · 류지혁¹ · 김원일¹ · 문병철¹ · 박상원^{1*} · 박용진^{2*}

¹국립농업과학원 농산물안전성부 화학물질안전과, ²공주대학교 산업과학대학 식물자원학과

Genome-wide Association Study on Cadmium Contents of Temperate Japonica Varieties

Sang-Beom Lee^{1,a}, Kyu-Won Kim^{2,a}, Min-Young Yoon², Gyeong-Jin Kim¹, Ji-Hyock Yoo¹, Won-Il Kim¹,
Byeong-Churl Moon¹, Sang-Won Park^{1*}, and Yong-Jin Park^{2*}

¹Chemical Safety Division, National Institute of Agriculture Science (NAIS), Wanju 55365, Korea

²Department of Plant resources, College of Industrial Sciences, Kongju National University, Yesan 32439, Korea

Abstract : Objective of this study was to investigate the difference of cadmium (Cd) levels in rice grains from non-polluted fields and to define the gene associated with Cd uptake for producing safety food. Cd was analyzed by Inductively Coupled Plasma Mass Spectrometer (ICP-MS). The average concentration of Cd in rice grains was 0.943 $\mu\text{g/kg}$ and Cd levels ranged from 0.050 to 5.699 $\mu\text{g/kg}$. Genome-Wide Association study (GWAS) based on phenotype data for Cd levels was performed. However, results of GWAS were affected by subpopulation structure and caused false positive. Therefore, GWAS for rice ecotypes (temperate Japonica, tropical Japonica, Indica, Aus, Aromatic, and Admixture) was performed to minimize false positive. GWAS results showed that Os01g0611300, Os01g0611900, Os01g0611950, Os01g0612000, Os01g0612200, Os11g0444400, Os11g0444700, Os11g0444800, and Os11g0444900 genes have significant correlation with Cd levels in rice grains. The sequences of these genes were compared to sequence positions of each other gene (haplotype analysis). According to the results of haplotype analysis, Cd levels of non-synonymous group were higher than other groups and sequence of non-synonymous group was similar to that of Indica. These results were corresponding to the previous research result that Cd levels of Indica were higher than Japonica. Therefore, candidate genes detected through GWAS need to be examined by knock-out or cross breeding.

Keywords : Genome-wide association study, Cadmium, Haplotype, Ecotype, Candidate genes

서 언

카드뮴은 납, 수은과 더불어 환경에 폭넓게 존재하고 있는 중금속 중 하나로 한번 환경에 노출되면 토양과 수질에 축적되어 식물의 생육을 저하시킨다(Uraguchi et al. 2012). 그리고 카드뮴에 오염된 물과 음식을 사람이 섭취하였을 경우 골연화 증상과 신장 이상을 유발한다(Horiguchi et al. 2010). 따라서 카드뮴 노출을 최소화하기 위해 중국과 EU에서는 곡류에 있어 카드뮴 함량 기준을 0.2 $\mu\text{g/kg}$, 호주는 0.1 $\mu\text{g/kg}$ 로 설정하였으며,

FAO/WHO 합동 식품첨가물전문가위원회(Expert Committee on Food Additives)와 국제식품규격위원회(Codex Alimentarius Commission)는 0.4 $\mu\text{g/kg}$ 수준으로 설정하고 있다(Joint FAO/WHO Expert Committee on Food Additives 2010; Codex committee on contaminants in Foods, CCCF8, www.fao.org/fao-who-codexalimentarius/). 한편 우리나라는 2008년부터 식품의약품안전처 고시에 의거하여 곡류의 카드뮴 함량 기준을 Codex 및 주요 외국과 유사한 수준의 0.1 $\mu\text{g/kg}$ (밀과 쌀은 0.2 $\mu\text{g/kg}$)으로 설정하고 있다(식품의약품안전처, 식품기준 및 규격에 의한 고시. 제2016-101호). Kim(2008)은 토양오염 대책기준인 4.0 $\mu\text{g/kg}$ 을 초과하여 우리나라에서 영농활동이 제한된 농경지를 대상으로 벼 품종 별 백미 카드뮴 함량을 분석한

*Corresponding author (E-mail: swpark@korea.kr, Tel: +82-63-238-3247 / E-mail: yjpark@kongju.ac.kr, Tel: +82-41-330-1201)

(Received on October 12, 2016. Accepted on November 9, 2016.)

결과, 대부분 0.173~1.188 $\mu\text{g/kg}$ 의 범위로 허용기준을 초과한다고 보고하였다.

벼의 카드뮴 함량을 줄이기 위해서는 담수와 토양개량제 처리, 객토와 같은 재배학적 방법을 이용할 수 있다. 그러나 이러한 재배학적 방법은 비용이 많이 들어 대체적으로 실용가능성이 낮은 것으로 알려져 있다(Takijima et al. 1973). 최근에는 품종 간의 유전적 다양성을 바탕으로 한 작물 내 카드뮴 함량을 저감하는 연구들이 이루어지고 있다.

Arao & Ae (2003)는 포트 재배 및 수경 재배를 통해 벼 품종 간 카드뮴 함량에 대한 유전적 다양성이 나타남을 확인하였으며, 물 관리가 벼 종자의 카드뮴 함량에 영향을 줄 수 있다고 보고하였다. 또한 Arao & Ishikawa (2006)는 콩과 벼의 카드뮴 함량과 분포에 대한 유전형의 차이를 통해 카드뮴과 다른 금속이온(Cu, Fe, Mn, Zn)들이 서로 상호작용하여 줄기로 흡수·이동한다고 보고하였다. 일반적으로 벼의 경우 카드뮴은 뿌리에서 흡수되어 물관부에 의해 줄기로 이동하고, 다시 마디에 축적되어 잎과 종자로 이동하게 된다. 이러한 벼의 카드뮴 흡수 메커니즘과 관련하여 OsIRT1, OsIRT2, OsHMA3 및 OsLCT1 유전자들이 순차적으로 벼의 카드뮴 흡수에 관여하는 것으로 알려져 있다(Lee & An, 2009).

본 연구에서는 주요 식량작물 중 하나인 벼 280개 품종을 대상으로 등숙기 종자 내 카드뮴 함량을 유도결합 플라즈마 질량분석기(ICP-MS)를 사용하여 $\mu\text{g/kg}$ 단위로 카드뮴 함량을 분석하였다. 그리고 카드뮴 함량 데이터를 전장유전체연관분석(GWAS)을 통하여 후보유전자를 동정하였고, 동정된 후보유전자들을 haplotype 분석하여 아미노산 서열을 분석하였다. 따라서 본 연구 결과는 향후 중금속 오염에 안전한 벼 품종 선발 및 계발을 위한 육종의 기초 자료로 사용될 것이다.

재료 및 방법

공시재료

공주대학교에서는 국립농업과학원 유전자원센터와 국제미작 연구소(IRRI)가 세계 각국에서 수집한 24,368점의 유전자원 중 특성 정보가 있는 도입종(introduced) 893점, 한국 유전자원 중 육성종(breeding line) 1,116점과 재래종(landrace) 394점, 잡초성 벼(weedy rice) 1,955점 및 국제미작연구소(IRRI)의 core set 48점을 포함한 총 4,046점을 대상으로 Kim et al. (2007)이 고안한 Heuristic algorithm 방법인 Powercore method를 적용하여 선발하였고 한국에서 재배하기 힘든 종 30

점을 제외하여, 137점을 핵심집단으로 선발하였다. 또한 한국 육성계통 158점을 추가하여 최종적으로 유전자원 295점을 선발하였다. 선발된 유전자원은 유전적 특성을 파악하기 위해 전장유전체(whole genome)에 대한 유전자 분석 및 각각의 염색체에 대한 단일염기다형성(single nucleotide polymorphism, SNP)이 분석되었다.

본 연구에서는 공주대학교에서 선발하여 보유한 295점의 벼 유전자원과 유전체 재분석 데이터, 카드뮴 함량 분석이 가능했던 벼 280점을 대상으로 등숙기 종자 내 카드뮴 함량에 관해 전장유전체연관분석을 하였다.

DNA 추출 및 유전체 재분석

유전자원 295점의 어린 잎을 각각 채취하여 액체질소를 넣고 막자사발에서 마쇄한 후, DNeasy® Plant Mini Kit(QIAGEN)를 이용하여 genomic DNA를 추출하였고 정량은 최소 농도 30ng/ μl 이상으로 하였다. 그리고 DNA QC를 위해 Quant-iT BR assay kit (Q32850, Invitrogen)를 이용하여 추출된 시료를 희석하여 ds_DNA의 형광 농도를 측정하였다. 광학밀도(optical density) 값의 측정은 Tecan F200 (Tecan, Switzerland)를 사용하였고, 추출된 DNA를 0.7% agarose gel에서 전기영동을 하여 확인 하였다. 유전체 재분석은 HiSeq 2500(Illumina) NGS (Next Generation DNA Sequencer)를 이용하여 short read sequence를 생산하였고, resequencing 방법으로 유전체 분석을 실시하였다.

벼 등숙기 종자 내 카드뮴 함량 분석

카드뮴에 오염되지 않은 논에서 재배된 핵심집단 137점과 한국 육성계통 158점을 포함한 295점 중 시료 양이 부족한 15개 품종을 제외한 280점을 대상으로 등숙기 종자 내 카드뮴 함량을 분석하였다. 등숙기에 채취한 종자를 건조 후 제현기로 현미를 만든 다음 믹서기로 마쇄한 시료를 각각 1g씩 비커에 담아 HNO_3 7ml, H_2O_2 1ml를 첨가하고 전기로 열판에서 180°C 조건으로 분해하였다. 분해된 시료는 초순수로 25ml를 정용한 뒤 유도결합플라즈마질량분석기 (ICP-MS 7700E, Agilent Technologies, U.S.)를 사용하여 카드뮴 함량을 분석하였다. 그리고 후보유전자 그룹 간 카드뮴 함량의 차이를 비교하기 위해 SAS (Statistical Analysis System ver. 9.0)를 이용하여 분산분석 (ANOVA test) 및 던칸다중검정 (Duncan's multiple range test)을 실시하였다.

Table 1. Cadmium range of 280 accessions.(units; $\mu\text{g/kg}$)

	Total (n=280)	Japonica (n=206)	Indica (n=61)
Mean	0.943176	0.787193	1.495608
Median	0.63415	0.5719	0.9453
Maximum	5.699	3.9064	5.699
Minimum	0.0528	0.0528	0.0953

Variant calling

Samtools과 GATK tool을 이용하여 variant calling을 하였다. variant calling에는 SNP position과 reference SNP, allele SNP, insertion, deletion의 정보를 갖고 있으며, 이러한 정보는 mapping 결과를 해석한 것이다. 최종적으로 variant calling File (VCF)을 작성하여 SNP의 영역을 찾을 수 있었다.

전장유전체연관분석(Genome-Wide Association studies, GWAS)

미국 코넬 대학교에서 개발한 Genomic Association and Prediction Integrated Tool (GAPIT, version 2)을 이용하여 전장유전체연관분석(GWAS)을 수행하였다(Lipka et al. 2012). GAPIT은 개화기, 종실중, 출수기 혹은 등숙기 종자 내 이온체 (Cd, As, Mg, Zn, Fe, Ca, Cu, etc.) 함량과 같은 식물의 양적 형질에 적용할 수 있는 프로그램으로 본 연구에서는 GAPIT의 혼합선형모형을 이용하였다(Zhang et al. 2010). 혼합선형모형은 기존 선형모형에 SNP들의 임의효과(random effect)를 추가한 것으로 일반적인 혼합선형모형은 Henderson의 행렬을 사용하면 다음과 같이 표현할 수 있다.

$$y=X\beta+Zv+e$$

여기서 $v \sim N(0, K\sigma_v^2)$ 이고 $e \sim N(0, I\sigma_e^2)$ 이라고 가정한다. 또한 y 는 $n \times 1$ 벡터로 관측된 표현형을 나타내며 n 은 표현형 개수로 표본의 개수와 동일하다. 위 식의 오른쪽 항들을 살펴보면 다음과 같다. X 는 0과 1로 구성되는 $n \times q$ 디자인 행렬로 질편, SNP, 그리고 집단 구조화(population structure) 등의 고정된 효과를 나타내며, β 는 $q \times 1$ 벡터로 q 개의 고정된 효과를 위해 추정하여야 할 회귀계수이다. Z 는 0과 1로 구성되는 $n \times p$ 디자인 행렬로 개체들 간의 친족효과(kinship effect)를 포함하는 임의

효과를 나타내며, v 는 $p \times 1$ 벡터로 임의 효과를 위해 추정하여야 할 회귀계수이다. p 는 친족행렬(kinship matrix)을 위한 그룹의 개수를 나타냄으로, $p \leq n$ 을 만족한다. 표본을 그룹화하지 않는 경우에는 $p = n$ 이 된다. 마지막으로 e 는 $n \times 1$ 벡터로 잔차효과(residual effects)를 나타낸다.

GWAS 결과는 100만개의 SNP를 사용하여 형질과 연관성을 동시에 검정하는 다중검정을 하기 때문에 위양성(false positive)이 나타날 수 있으며, 이것을 검정하는 방법은 다음과 같다. 모든 SNP가 서로 독립적이라 가정한다면 SNP에 대한 유의수준이 α , 적어도 1개의 SNP에 대해 위양성(false positive)을 보일 확률을 ζ 라 할 때, SNP가 잘못되지 않은 결론을 내릴 확률은 $\zeta=1-(1-\alpha)^n$ 이 된다(식물분자유종사업단, 2013, 식물 분자 육종 최신 기술, p235-238). 따라서 본 연구에서는 다중검정문제를 해결하기 위해 FDR 검정 (False Discovery Rate Test)를 사용하였으며 온대자포니카와 열대자포니카, 인디카, Aus, Aromatic, Admixture로 집단 구조화(population structure)를 달리하여 GWAS 분석을 하였다.

결과 및 고찰

벼 등숙기 종자 내 카드뮴 함량분석 결과

280점의 품종별 카드뮴 함량 분석 데이터 결과 전체 카드뮴 함량의 평균은 $0.9431 \mu\text{g/kg}$ 이었고, 최소값은 $0.052 \mu\text{g/kg}$, 최대값은 $5.699 \mu\text{g/kg}$ 이었다. 종자의 카드뮴 함량은 대체로 한국 및 Codex 및 주요 외국의 허용기준인 $0.2 \mu\text{g/kg}$ 을 넘지는 않았으나 오염되지 않은 토양임에도 불구하고 상대적으로 다른 품종들에 비해 상대적으로 높은 카드뮴 함량을 보이는 품종들이 있었다. 또한 인디카 계통의 품종들의 평균 카드뮴 함량은 $1.495 \mu\text{g/kg}$ 으로 자포니카 계통 품종의 카드뮴 함량 $0.943 \mu\text{g/kg}$ 보다 상대적으로 높았다(Table 1).

Table 2. Summary of sequence data and mapping statistics.

Sample Size	Sequence Read	Mapping Rate(%)	Mean Depth	Mean Depth in gene Region	SNP	InDel
295	42,414,594	96.40	9.83X	9.38X	11,632,676	1,285,506

Table 3. SNP and InDel numbers in the re-sequenced rice genome.

Group	Sample Number	Variants			
		SNP	HQ ^c _SNP	InDel	HQ ^c _InDel
Heuristic set (HS)	137	11,032,184	1,998,840	1,198,144	265,426
Korea Bred. (KB)	158	8,565,179	3,957,109	999,782	252,344
Total	295	11,632,676	1,481,705	1,285,506	190,828

Table 4. SNP genes position associated with the phenotypes of 280 accessions.

SNP	Chromosome	Position	P value	-log (P)	maf	FDR_Adjusted P-values
Chr_20067501	2	20,067,501	2.10E-06	5.68	0.123656	0.381048916
Chr_24368066	1	24,368,066	2.54E-06	5.60	0.216846	0.381048916
Chr_24237296	1	24,237,296	5.13E-06	5.29	0.222222	0.381048916
Chr_24799373	1	24,799,373	5.84E-06	5.23	0.141577	0.381048916
Chr_23754508	1	23,754,508	6.92E-06	5.16	0.18638	0.381048916

* maf: minor allele frequency

벼 유전체 재분석

벼 핵심집단 137점과 한국 육성계통 158점을 포함한 295점의 유전체 재분석 결과, sequence read는 42,414,594개이며, 평균 depth는 9.83X, mapping rate는 96.40%로 나타났다. 또한 유전자내의 SNP와 InDel은 각각 11,632,676과 1,285,506개였다 (Table 2). 그리고 핵심집단(Heuristic Set) 137점의 Total SNP는 11,032,184개이고, 전체 InDel은 1,198,144개였으며, 국내 육성종의 전체 SNP는 8,565,179개, 전체 InDel은 999,782개로 분석되었다(Table 3).

유전체재분석 자료의 전장유전체분석(GWAS) 및 후보유전자 확인

GAPIT을 이용한 GWAS 분석은 보통 100만개 정도의 SNP를 사용해서 형질과 연관성을 동시에 검정한다(식물분자유종사업단. 2013. 식물분자유종최신기술. p235-238). 유전체 재분석

을 한 280점에 대해 다중검정을 한 결과 염색체 2번과 1번에서 각각 high peak를 보인 p-value 값 2.10×10^{-6} , 2.54×10^{-6} 은 기존에 발표된 Huang(2012)의 벼 개화기와 생산량에 관한 GWAS 연구와 Kim(2016)의 Korean rice core set의 전장유전체 재분석 연구의 p-value값과 비교하였을 때 SNP와 형질 사이에 연관성이 있었다.

염색체 2번의 경우 20,067,501bp 위치에서 $-\log_{10}(p)$ 값이 5.68로 가장 높은 값을 보였다. 그리고 염색체 1번의 24,368,066, 24,237,296, 24,799,373bp 위치에서 각각 5.60, 5.29, 5.23, 5.16의 $-\log_{10}(p)$ 값을 보였다(Table 4). 각 염색체에서 가장 높은 값을 보인 위치를 기준으로 Rap-DB (<http://rapdb.dna.affrc.go.jp/>)를 활용하여 ± 50 Kbp 범위 내에서 확인한 결과 염색체 1번과 2번에서 각각 33개와 9개의 후보유전자를 확인하였다 (Table 5).

등숙기 종자 내 카드뮴 함량과 SNP의 연관성을 Quantile-Qua

Table 5. Candidate gene associated with cadmium contents on chromosome 1 and 2 of 280 accessions.

Chromosome	Name	Description	Position
1	Os01g0612900	Lipase, GDSL domain containing protein (Os01t0612900-01)	chr01:24312652..24317113 (+ strand)
1	Os01g0613100	Similar to predicted protein (Os01t0613100-00)	chr01:24334363..24336159 (+ strand)
1	Os01g0613300	Conserved hypothetical protein CHP02058 domain containing protein (Os01t0613300-01)	chr01:24341214..24343585 (+ strand)
1	Os01g0613400	Non-protein coding transcript (Os01t0613400-01)	chr01:24341881..24342143 (+ strand)
1	Os01g0613500	Peptidase C1A, papain family protein (Os01t0613500-01)	chr01:24343518..24345172 (- strand)
1	Os01g0613800	Peptidase C1A, papain family protein (Os01t0613800-01)	chr01:24351308..24353345 (- strand)
1	Os01g0613900	Hypothetical conserved gene (Os01t0613900-01); Similar to predicted protein (Os01t0613900-02); LPPG:FO 2-phospho-L-lactate transferase CofD/UPF0052 domain containing protein (Os01t0613900-03)	chr01:24354168..24360484 (- strand)
1	Os01g0614300	Protein of unknown function DUF231, plant domain containing protein (Os01t0614300-01)	chr01:24364146..24367678 (+ strand)
1	Os01g0614400	Conserved hypothetical protein (Os01t0614400-01)	chr01:24366702..24367660 (- strand)
1	Os01g0614500	Nucleotide-binding, alpha-beta plait domain containing protein (Os01t0614500-01)	chr01:24369103..24372156 (- strand)
1	Os01g0614750	Hypothetical protein (Os01t0614750-00)	chr01:24375913..24379864 (- strand)
1	Os01g0614700	Similar to Adaptin ear-binding coat-associated protein 1. (Os01t0614700-01); Adaptin ear-binding coat-associated protein 1 NECAP-1 family protein (Os01t0614700-02)	chr01:24375877..24379899 (+ strand)
1	Os01g0614900	NAD(P)-binding domain containing protein (Os01t0614900-01); Similar to autophagy-related 7 (Os01t0614900-02)	chr01:24380637..24386666 (- strand)
1	Os01g0615000	Hypothetical protein (Os01t0615000-01)	chr01:24387024..24387834 (+ strand)
1	Os01g0615100	Similar to Substilin /chymotrypsin-like inhibitor (Proteinase inhibitor) (Os01t0615100-01)	chr01:24392841..24393423 (+ strand)
1	Os01g0615050	Proteinase inhibitor I13, potato inhibitor I family protein (Os01t0615050-00)	chr01:24390676..24391763 (- strand)
1	Os01g0615200	Transferase family protein (Os01t0615200-01)	chr01:24395081..24396820 (- strand)
1	Os01g0615250	Hypothetical protein (Os01t0615250-00)	chr01:24395291..24396703 (+ strand)
1	Os01g0615300	Transferase family protein (Os01t0615300-01); Transferase family protein (Os01t0615300-02)	chr01:24397564..24400652 (- strand)
1	Os01g0615400	Hypothetical protein (Os01t0615400-00)	chr01:24397782..24400564 (+ strand)
1	Os01g0615500	ABC transporter-like domain containing protein (Os01t0615500-00)	chr01:24408462..24410310 (- strand)
1	Os01g0615850	Hypothetical gene (Os01t0615850-01)	chr01:24418549..24419274 (- strand)
1	Os01g0611000	Protein of unknown function DUF642 domain containing protein	chr01:24181111..24183979 (+ strand)
1	Os01g0611050	Hypothetical gene (Os01t0611050-00)	chr01:24187513..24189601 (- strand)
1	Os01g0611300	Similar to Metal tolerance protein 7 (Os01t0611300-01)	chr01:24195504..24196616 (+ strand)
1	Os01g0611900	Pentatricopeptide repeat domain containing protein.	chr01:24240257..24245230 (- strand)

Chromosome	Name	Description	Position
1	Os01g0611950	Hypothetical gene (Os01t0611950-00)	chr01:24244539..24245291 (+ strand)
1	Os01g0612000	Similar to DNA methyl transferase4 (Os01t0612000-01)	chr01:24245869..24249361 (- strand)
1	Os01g0612200	Cytochrome c oxidase, subunit Vb family protein (Os01t0612200-01)	chr01:24255379..24259940 (+ strand)
1	Os01g0612350	Hypothetical gene (Os01t0612350-01)	chr01:24261982..24267594 (- strand)
1	Os01g0612425	Hypothetical protein (Os01t0612425-00)	chr01:24271398..24272349 (+ strand)
1	Os01g0612500	Phospholipase/carboxylesterase domain containing protein (Os01t0612500-01)	chr01:24277095..24280999 (+ strand)
1	Os01g0612600	Protein of unknown function DUF1644 family protein (Os01t0612600-01)	chr01:24282911..24285306 (- strand)
2	Os02g0539600	SANT domain, DNA binding domain containing protein (Os02t0539600-01)	chr02:20011829..20013851 (+ strand)
2	Os02g0539900	WD40 repeat domain containing protein (Os02t0539900-00)	chr02:20026633..20030067 (- strand)
2	Os02g0540000	Similar to Zinc finger, C3HC4 type family protein, expressed (Os02t0540000-01)	chr02:20034215..20036481 (+ strand)
2	Os02g0540400	Similar to Yarrowialipolytica chromosome C of strain CLIB99 of Yarrowialipolytica (Os02t0540400-01)	chr02:20056435..20057359 (- strand)
2	Os02g0540700	Similar to Arm repeat-containing protein (Os02t0540700-00)	chr02:20082159..20083676 (+ strand)
2	Os02g0541000	Conserved hypothetical protein (Os02t0541000-01)	chr02:20102615..20103218 (+ strand)
2	Os02g0541300	Hypothetical conserved gene (Os02t0541300-00)	chr02:20109037..20109343 (+ strand)
2	Os02g0541150	Hypothetical protein (Os02t0541150-00)	chr02:20109198..20110897 (- strand)
2	Os02g0541325	Similar to Serine decarboxylase (Os02t0541325-00)	chr02:20109369..20112225 (+ strand)

Fig. 1. Plotting of genome-wide association scan for cadmium contents in 280 accessions (total). a) Manhattan plot, b) Q-Q plot.

Table 6. SNP genes position associated with the phenotypes of 182 accessions (only temperate Japonica).

SNP	Chromosome	Position	P value	-log (P)	maf	FDR_Adjusted P-values
Chr_24486393	1	24,486,393	9.25E-07	6.03	0.051913	0.044195254
Chr_24524351	1	24,524,351	9.25E-07	6.03	0.051913	0.044195254
Chr_24237296	1	24,237,296	9.25E-07	6.03	0.051913	0.044195254
Chr_14703809	11	14,703,809	2.27E-06	5.64	0.112022	0.044195254
Chr_14309642	11	14,309,642	6.10E-06	5.21	0.139344	0.044195254

* maf: minor allele frequency

Table 7. Candidate gene associated with cadmium contents on chromosome 1 and 11 of 182 accessions (only temperate Japonica).

Chromosome	Name	Description	Position
1	Os01g0611300	Similar to Metal tolerance protein 7 (Os01t0611300-01)	chr01:24195504..24196616 (+ strand)
1	Os01g0611900	Pentatricopeptide repeat domain containing protein	chr01:24240257..24245230 (- strand)
1	Os01g0611950	Hypothetical gene (Os01t0611950-00)	chr01:24244539..24245291 (+ strand)
1	Os01g0612000	Similar to DNA methyl transferase4 (Os01t0612000-01)	chr01:24245869..24249361 (- strand)
11	Os11g0444400	Conserved hypothetical protein(Os11t0444400-01)	chr11:14687711..14688398 (+ strand)
11	Os11g0444700	Octicosapeptide/Phox/Bem1p domain containing protein (Os11t0444700-01)	chr11:14705242..14706300 (+ strand)
11	Os11g0444800	Conserved hypothetical protein(Os11t0444800-01)	chr11:14713141..14714215 (- strand)
11	Os11g0444900	Octicosapeptide/Phox/Bem1p domain containing protein(Os11t0444900-01)	chr11:14713272..14714398 (+ strand)

ntile(Q-Q) plot과 Manhattan plot 그래프를 통해 시각적으로 확인한 결과 카드뮴 함량과 SNP간에 연관성이 있는 것으로 나타났다(Fig. 1a). 한편, Manhattan plot에서 가로축은 염색체 번호로 2번에서 $-\log_{10}(p)$ 값이 5.68로 가장 높게 나타났고 그 다음으로 염색체 1번에서 $-\log_{10}(p)$ 값이 5.60으로 높게 나타났다(Fig. 1b).

하지만 카드뮴 함량을 분석한 자원 280점을 온대자포니카와 열대자포니카, 인디카, Aus, Aromatic, Admixture로 생태형을 달리하여 GWAS 분석을 한 결과, 염색체 2번의 SNP position은 인디카에서만 나타났고 자포니카에서는 찾을 수 없었으며, 인디카에서 $-\log_{10}(P)$ 값이 3.63을 보여 SNP와 형질 간의 연관성이 낮게 나타났다. 염색체 2번의 경우 전체 280점의 GWAS 분석 결과와 생태형에 따른 GWAS 결과가 서로 대조적으로 나타나게

된 것은 GWAS 결과에 집단구조화가 영향을 미친 것으로 사료된다. 반면, 염색체 1번에서 전체 자원 280점과 동일한 SNP position을 온대자포니카 182점에서 확인할 수 있었으며, $-\log_{10}(P)$ 값은 6.03으로 높게 나타났다(Table 6).

온대자포니카 품종의 전장유전체분석 (GWAS) 및 후보유전자 확인

온대자포니카 182점의 GWAS 분석을 한 결과 염색체 1번의 24,237,296, 24,486,393, 24,524,351bp 위치에서 $-\log_{10}(p)$ 값이 모두 6.03으로 높게 나타났으며, 염색체 11번은 position 14,703,809bp에서 $-\log_{10}(p)$ 값이 5.64로 높았다(Table 6). 또한 등숙기 종자 내 카드뮴 함량과 관련하여 표현형과 유전형 간에 연관성이 높게 나타남을 알 수 있었다(Fig. 2a, 2b). 특히 $-\log_{10}(p)$

Fig. 2. Plotting of genome-wide association scan for cadmium contents in 182 accessions (only temperate Japonica). a) Manhattan plot, b) Q-Q plot.

값이 6.03인 염색체 1번의 24,237,296bp 위치는 전체 280점의 GWAS 분석 결과에서도 $-\log_{10}(p)$ 값이 5.29로 높게 나타났다. 따라서 염색체 1번의 24,237,296bp 위치가 카드뮴 함량과 유전적인 연관성이 가장 높다고 판단되었으며, 염색체 1번에서 $-\log_{10}(p)$ 값이 6.03인 24,237,296bp 위치의 $\pm 50\text{kbp}$ 범위에서 Rap-DB를 활용하여 후보 유전자를 확인하고 다시 $\pm 25\text{kbp}$ 범위에서 연관성이 가장 높은 후보유전자를 선발하였다. 염색체 11번 또한 $-\log_{10}(p)$ 값이 5.64로 높았던 14,703,809bp 위치의 $\pm 50\text{kbp}$ 범위에서 후보유전자를 확인하였고 $\pm 25\text{kbp}$ 범위에서 연관성이 가장 높은 후보유전자를 선발하였다(Table 7).

염색체 1번의 동정된 후보 유전자 중 카드뮴과 연관성이 높을 것으로 추정되는 유전자는 $\pm 50\text{kbp}$ 내의 Os01g0611300 (similar to metal tolerance protein 7)와 $\pm 25\text{kbp}$ 내의 Os01g0611900 (pentatricopeptide repeat domain containing protein), Os01g0611950 (Hypothetical gene), Os01g0612000 (similar to DNA methyl transferase4), Os01g0612200 (cytochrome c oxidase, subunit Vb family protein) 이었다. 한편, 염색체 11번의 $\pm 25\text{kbp}$ 범위에서 동정된 후보유전자는 Table 7과 같다. 후보유전자 중 Os11g0444700 (octicosapeptide / Phox / Bem1p domain containing protein) 유전자는 14,703,809bp에서 20kbp 범위 안에 있어 카드뮴 함량에 가장 많은 영향을 미칠 것으로 판단되었다. 그 외 Os11g0444400 (conserved hypothetical protein), Os11g0444800 (conserved hypothetical protein), Os11g0444900 (octicosapeptide / Phox / Bem1p domain containing protein) 유전자가 $\pm 25\text{kbp}$ 범위에서 동정되었다.

후보유전자들의 haplotype 분석

염색체 1번 24,237,296bp와 염색체 11번 14,703,809bp의 $\pm 25\text{kbp}$ 위치에서 선발된 후보유전자를 확인하여 haplotype 분석을 하였다. Haplotype 분석 시 Nipponbare (temperate Japonica) 염기서열을 기준으로 삼았으며, 그룹 간의 표현형 데이터를 비교하여 아미노산 서열을 분석하였다.

염색체 1번에서 선발된 후보유전자는 Os01g0611300, Os01g0611900 및 Os01g0611950, Os01g0612000, Os01g0612200이었다. Os01g0611300 유전자는 haplotype 분석 결과 2개의 그룹(Hap1, Hap2)으로 구분하였으며 각 그룹간의 평균은 0.734, 2.014 $\mu\text{g/kg}$ 이었다. 낮은 카드뮴 함량을 보인 Os01g0611300-Hap1 그룹은 Exon 부분에서 Nipponbare와 유사한 염기를 갖는 반면, 높은 카드뮴 함량을 보인 Os01g0611300-Hap2 그룹은 Nipponbare의 염기와 비교하였을 때, Exon의 1057번 위치에서 G에서 A로 변환되었으며, Serine(AGT)에서 Asparagine(AAT)으로 아미노산 서열이 바뀌었다. Os01g0611300-Hap1 그룹에 속하는 품종은 173개로 카드뮴 함량이 0.052~3.050 $\mu\text{g/kg}$ 의 범위였으며, Os01g0611300-Hap2 그룹에 속하는 품종은 9개로 카드뮴 함량이 0.369~3.906 $\mu\text{g/kg}$ 의 범위를 보였다(Fig. 3a).

Os01g0611900 유전자는 전사방법에 따라 Os01t0611900-01과 Os01t0611900-02로 세분되며 각각의 전사체는 5' UTR과 Exon, 3' UTR에서 차이가 있다. Haplotype 분석결과 3개의 그룹(Hap1, Hap2, Hap3)으로 구분하였으며, 각 그룹간의 평균 카드뮴 함량은 2.014, 0.702, 1.167 $\mu\text{g/kg}$ 이었다. 가장 높은 카드뮴 함량을 보인 Os01g0611900-Hap1 그룹은 Nipponbare의 염기와 비교하였을 때, Exon 부분의 각각 2917, 3065, 4358번

Fig. 3. a) Os01t0611300-01

Chromosome 1	Exon 1	Means of cadmium contents (Units: $\mu\text{g/kg}$)	F-value	Acc No.
SNP Position	1057			
Ref. Sequence	G			
Allele	G/A			
Amino acid substitution	Ser->Asn			
Codon Variant	AGT -> AAT			
Hap1	G	0.734 ^b	30.99 ^{**}	173
Hap2	A	2.014 ^a		9

Fig. 3. b) Os01t0611900-01 **Os01t0611900-02**

Chromosome 1	Exon 1											Means of cadmium contents (Units: $\mu\text{g/kg}$)	F-value	Acc No.
SNPPosition	2583	2917	2956	3065	3086	3088	3613	4050	4283	4358	4648			
Ref. Sequence	T	C	C	A	G	G	G	T	C	C	G			
Allele	T/C	C/G	C/G	A/T	G/A	G/A	G/A	T/C	C/T	C/T	G/A			
Amino acid substitution	Arg ↓ Gly	Leu ↓ Phe	Gly ↓ Gly	Phe ↓ Tyr	Ala ↓ Val	Ile ↓ Ile	Ile ↓ Ile	Thr ↓ Ala	Arg ↓ Lys	Arg ↓ His	Leu ↓ Leu			
Codon Variant	AGG ↓ GGG	TTG ↓ TTC	GGG ↓ GGC	TTT ↓ TAT	GCT ↓ GTT	ATC ↓ ATT	ATC ↓ ATT	ACG ↓ GCG	AGG ↓ AAG	CGC ↓ CAC	CTC ↓ CTT			
Hap1	T	G	C	T	G	G	G	T	C	T	G	2.014 ^a	18.61 ^{**}	9
Hap2	T	C	C	A	G	G	G	T	C	C	G	0.702 ^b		161
Hap3	T	C	C	A	G	G	A	T	C	C	G	1.167 ^b		12

Fig. 3. c) Os01t0611950-00

Chromosome 1	Exon 1	Means of cadmium contents (Units: $\mu\text{g/kg}$)	F-value	Acc No.
SNPPosition	4358			
Ref. Sequence	C			
Allele	C/T			
Amino acid substitution	Ala -> Val			
Codon Variant	GCG -> GTG			
Hap1	T	2.014 ^a	30.99 ^{**}	9
Hap2	C	0.734 ^b		173

Fig. 3. d) Os01t0612000-01

Chromosome 1	Exon1		Exon 3	Exon 5	Means of cadmium contents (Units: $\mu\text{g}/\text{kg}$)	F-value	Acc No.
SNPPosition	1279	1280	1641	2293			
Ref. Sequence	G	-	T	T			
Allele	G/A	G/GA	T/A	T/C			
Amino acid substitution	Ser ↓ Phe	-	Lys ↓ Met	Leu ↓ Leu			
Codon Variant	TCT ↓ TTT	-	AAG ↓ ATG	TTA ↓ TTG			
Hap1	G	-	T	C	0.779 ^b	15.54 ^{**}	39
Hap2	G	-	T	T	0.721 ^b		134
Hap3	G	-	T	C	2.014 ^a		9

Fig. 3. e) Os01t0612200-01

Chromosome 1	Exon 1		Exon 6	Means of cadmium contents (Units: $\mu\text{g}/\text{kg}$)	F-value	Acc No.
SNPPosition	174	205	4148			
Ref. Sequence	C	C	G			
Allele	C/T	G/GGCC	G/T			
Amino acid substitution	Ser ↓ Ser	-	Val ↓ Phe			
Codon Variant	TCC ↓ TCT	-	GTC ↓ TTC			
Hap1	C	-	G	1.014 ^b	17.22 ^{**}	16
Hap2	C	-	G	0.705 ^a		157
Hap3	C	-	G	2.014 ^a		9

Fig. 3. Haplotype analysis of the sequence covering the staple candidate genes on chromosome 1 coding regions in 182 accessions and phenotypic variation among different haplotypes. The statistical analysis was performed in SAS 9.0 package. a) Os01t0611300-01, b) Os01t0611900-01, Os01t0611900-02, c) Os01t0611950-00, d) Os01t0612000-01, e) Os01t0612200-01.

위치에서 C에서 G로, A에서 T, C에서 T로 변환되었다. 이러한 염기변화는 Leucine(TTG)에서 Phenylalanine(TTC), Phenylalanine(TTT)에서 Tyrosine(TAT), Leucine(CGC)에서 Histidine(CAC)으로 아미노산 서열을 바꾸었다. 한편, 가장 카드뮴 함량이 낮은 Os01g0611900-Hap2 그룹은 Nipponbare의 염기와 유사하였으며 Os01g0611900-Hap3 그룹은 Exon의 3613번

위치에서 G에서 A로 염기변화가 있었다. 그리고 아미노산서열이 Isoleucine(ATC)에서 Isoleucine(ATT)으로 바뀌었다. Os01g0611900-Hap1 그룹에 속하는 품종은 9개로 카드뮴 함량이 0.369~3.906 $\mu\text{g}/\text{kg}$ 의 범위였으며, Os01g0611900-Hap2 그룹은 161개로 카드뮴 함량이 0.052~3.050 $\mu\text{g}/\text{kg}$ 의 범위였다. 그리고 Os01g0611900-Hap3 그룹은 12개의 품종이 속해 있으며

카드뮴 함량이 0.177~2.661 $\mu\text{g/kg}$ 범위로 나타났다(Fig. 3b).

Os01g0611950 유전자는 Os01g0611900 유전자와 염기가 중복되어 2개의 그룹(Hap1, Hap2)으로 구분하였으며, 각 그룹 간 평균 카드뮴 함량은 2.014, 0.734 $\mu\text{g/kg}$ 이었다. 높은 카드뮴 함량을 보인 Os01g0611950-Hap1 그룹은 Nipponbare의 염기와 비교하였을 때, Exon의 4358번 위치에서 C에서 T로 변환되었고, Alanine(GCG)에서 Valine(GTG)으로 아미노산 서열이 바뀌었다. Os01g0611950-Hap1 그룹에 속하는 품종은 9개로 카드뮴 함량이 0.369~3.906 $\mu\text{g/kg}$ 범위이며, Os01g0611950-Hap2 그룹에 속하는 품종은 173개로 카드뮴 함량이 0.052~3.050 $\mu\text{g/kg}$ 범위로 나타났다(Fig. 3c).

Os01g0612000 유전자는 3개의 그룹(Hap1, Hap2, Hap3)으로 구분하였으며, 각 그룹의 평균 카드뮴 함량은 0.779, 0.721, 2.014 $\mu\text{g/kg}$ 이었다. Os01g06112000-Hap1 그룹과 Os01g06112000-Hap3 그룹은 Exon5의 2293번 위치에서 T에서 C로 염기변화가 있었으며, 아미노산 서열은 Leucine(TTA)에서 Leucine(TTG)로 바뀌었다. 반면, Os01g06112000-Hap1 그룹과 Os01g06112000-Hap3 그룹 모두 Exon1의 1279번 위치와 Exon3의 1641번 위치에서 아미노산 서열이 Serine(TCT)에서 Phenylalanine(TTT), Lysine(TTA)에서 Methionine(TTG)으로 바뀌었으나, Exon1과 Exon6에서 염기변화는 없었다. Os01g06112000-Hap2 그룹은 Exon1, Exon3, Exon5 모두에서 Nipponbare와 같은 염기를 가졌다. Os01g06112000-Hap1 그룹에 속하는 품종은 39개로 카드뮴 함량이 0.052~3.034 $\mu\text{g/kg}$ 범위이며, Os01g0611200-Hap2 그룹에 속하는 품종은 134개로 카드뮴 함량이 0.118~3.050 $\mu\text{g/kg}$ 범위로 나타났다. 그리고 Os01g0611200-Hap3 그룹에 속하는 품종은 9개로 카드뮴 함량이 0.369~3.906 $\mu\text{g/kg}$ 의 범위를 보였다(Fig. 3d).

Os01g0612200 유전자는 3개(Hap1, Hap2, Hap3)의 그룹으로 구분하였다. 각 그룹의 평균 카드뮴 함량은 1.014, 0.705, 2.014 $\mu\text{g/kg}$ 이었다. 3개의 그룹 모두 Exon1과 Exon6에서 염기변화는 없었으며, Exon1의 174번 위치에서 아미노산 서열이 Serine(TCC)에서 Serine(TCT)으로 Exon 4148번 위치는 Valine(GTC)에서 Phenylalanine(TTC)으로 바뀌었다.

Os01g06112200-Hap1 그룹에 속하는 품종은 16개로 카드뮴 함량이 0.0528~2.661 $\mu\text{g/kg}$ 범위이며, Os01g0611200-Hap2 그룹에 속하는 품종은 157개로 카드뮴 함량이 0.117~3.050 $\mu\text{g/kg}$ 범위로 나타났다. 그리고 Os01g0611200-Hap3 그룹에 속하는 품종은 9개로 카드뮴 함량이 0.369~3.906 $\mu\text{g/kg}$ 의 범위를 보였다(Fig. 3e).

염색체 11번에서 선발된 후보 유전자는 Os11g0444400, Os11g0444700, Os11g0444800, Os11g0444900이었다. Os11g0444400 유전자는 Haplotype 분석 결과 3개의 그룹(Hap1, Hap2, Hap3)으로 구분하였으며 그룹의 평균 카드뮴 함량은 각각 0.189, 1.382, 0.684 $\mu\text{g/kg}$ 이었다. Os11g0444400-Hap1 그룹은 Nipponbare 염기에 대해 첫 번째 Exon의 403번과 459번, 472번 위치에서 A에서 C로, C에서 A, C에서 T로 염기변화가 있었으며, Glutamic acid(GAA)에서 Aspartic acid(GAC), Alanine(GCG)에서 Glutamic acid(GAG), Aspartic acid(GAC)에서 Aspartic acid(GAT)으로 아미노산 서열이 바뀌었다. Os11g0444400-Hap2 그룹은 Exon 403번 위치에서만 염기변화가 있었으며 아미노산 서열이 Glutamic acid(GAA)에서 Aspartic acid(GAC)로 바뀌었다. 반면 Os11g0444400-Hap3 그룹은 모든 Exon 부분에서 아미노산 서열만 바뀌었을 뿐 염기변화는 없었다. 카드뮴 함량이 가장 낮은 Os11g0444400-Hap1 그룹에 속하는 품종은 2개로 카드뮴 함량의 범위는 0.119~0.260 $\mu\text{g/kg}$ 이었다. 그리고 Os11g0444400-Hap2 그룹에 속하는 품종은 31개로 카드뮴 함량의 범위는 0.1721~3.9064 $\mu\text{g/kg}$ 이었으며, Os11g0444400-Hap3 그룹에 속하는 품종은 149개로 카드뮴 함량의 범위는 0.052~3.034 $\mu\text{g/kg}$ 이었다(Fig. 4a).

Os11g0444700 유전자는 3개의 그룹(Hap1, Hap2, Hap3)으로 구분하였으며 그룹의 평균 카드뮴 함량은 각각 0.189, 1.382, 0.684 $\mu\text{g/kg}$ 이었다. Os11g0444700-Hap1 그룹은 Exon1의 181번과 302번, Exon2의 553번과 828번 위치에서 C에서 A로, C에서 T, A에서 C로 염기변화가 있었으며, Leucine(CTC)에서 Isoleucine(ATC), Alanine(GCG)에서 Valine(GTG), Asparagine(AAC)에서 Threonine(ACC)으로 아미노산 서열이 바뀌었다. Os11g0444700-Hap2 그룹은 Exon1의 302번과 Exon2의 553번 위치에서 C에서 T, A에서 C로 염기변화가 있었으며, Alanine(GCG)에서 Valine(GTG), Asparagine(AAC)에서 Threonine(ACC)으로 아미노산 서열이 바뀌었다. 반면, Os11g0444700-Hap3 그룹은 Exon에서 아미노산 서열만 바뀌었을 뿐, Exon에서 염기변화는 없었다. 카드뮴 함량이 가장 낮은 Os11g0444700-Hap1 그룹에 속하는 품종은 2개로, 카드뮴 함량의 범위는 0.119~0.260 $\mu\text{g/kg}$ 이었다. 그리고 Os11g0444700-Hap2 그룹에 속하는 품종은 31개로 카드뮴 함량의 범위는 0.172~3.906 $\mu\text{g/kg}$ 이었으며, Os11g0444700-Hap3 그룹에 속하는 품종은 149개로 카드뮴 함량의 범위는 0.052~3.034 $\mu\text{g/kg}$ 이었다(Fig. 4b).

Fig. 4. a) Os11t0444400-01

Chromosome 11	Exon 1			Means of cadmium contents (Units: $\mu\text{g/kg}$)	F-value	Acc No.
SNPPosition	403	459	472			
Ref. Sequence	A	C	C			
Allele	A/C	C/A	C/T			
Amino acid substitution	Glu ↓ Asp	Ala ↓ Glu	Asp ↓ Asp			
Codon Variant	GAA ↓ GAC	GCG ↓ GAG	GAC ↓ GAT			
Hap1	C	A	T	0.189 ^b	14.41 ^{**}	2
Hap2	C	C	C	1.382 ^a		31
Hap3	A	C	C	0.684 ^{ab}		149

Fig. 4. b) Os11t0444700-01

Chromosome 11	Exon 1		Exon 2		Means of cadmium contents (Units: $\mu\text{g/kg}$)	F-value	Acc No.
SNP Position	181	302	553	828			
Ref. Sequence	C	C	A	T			
Allele	C/A	C/T	A/C	T/G			
Amino acid substitution	Leu ↓ Ile	Ala ↓ Val	Asn ↓ Thr	Tyr ↓ Asp			
Codon Variant	CTC ↓ ATC	GCG ↓ GTG	AAC ↓ ACC	TAC ↓ GAC			
Hap1	A	T	C	T	0.189 ^b	14.41 ^{**}	2
Hap2	C	T	C	T	1.382 ^a		31
Hap3	C	C	A	T	0.684 ^{ab}		149

Os11g0444800 유전자는 3개의 그룹(Hap1, Hap2, Hap3)으로 구분하였으며 그룹의 평균 카드뮴 함량은 각각 0.684, 0.189, 1.382 $\mu\text{g/kg}$ 이었다. Os11g0444800-Hap1 그룹은 Exon에서 염기변화가 없었으며, Os11g0444800-Hap2 그룹은 Exon 495번과 564번 위치를 제외한 모든 Exon부분에서 allele를 가지는 동시에 아미노산 서열이 바뀌었다. Exon 495번과 564번 위치에서 SNP deletion(-1, -3)이 나타났다. 그리고 Os11g0444800-Hap3 그룹은 Exon 745번 위치에서만 염기변화가 없었고, 나머지

Exon 부분에서 allele를 가지는 동시에 아미노산 서열이 바뀌었다. Os11g0444800-Hap1 그룹에 속하는 품종은 149개로 카드뮴 함량의 범위는 0.052~3.034 $\mu\text{g/kg}$ 이었다. 그리고 카드뮴 함량이 가장 낮은 Os11g0444800-Hap2 그룹에 속하는 품종은 2개로 카드뮴 함량의 범위는 0.119~0.260 $\mu\text{g/kg}$ 이었으며, Os11g0444800-Hap3 그룹에 속하는 품종은 31개, 카드뮴 함량의 범위는 0.172~3.906 $\mu\text{g/kg}$ 이었다(Fig. 4c).

Os11g0444900 유전자는 3개의 그룹(Hap1, Hap2, Hap3)으로

Fig. 4. c) Os11t0444800-01

Chromosome 11	Exon1																Means of cadmium contents (Units: $\mu\text{g}/\text{kg}$)	F-value	Acc No.
SNPPosition	492	495	497	503	512	518	520	524	525	527	528	541	564	578	643	745			
Ref. Sequence	G	-	C	C	A	A	T	A	T	T	C	A	-	T	G	G			
Allele	G/C	AC/A	C/T	C/G	A/T	A/T	T/C	A/T	T/A	T/G	C/T	A/C	TTG/T	T/C	G/A	G/A			
Amino acid substitution	Ala ↓ Gly	-	Gln ↓ Gln	Gln ↓ Asn	Ser ↓ Arg	Thr ↓ Thr	Thr ↓ Ala	Asp ↓ Glu	Asp ↓ Val	Arg ↓ Ser	Arg ↓ Lys	Ser ↓ Ala	-	Gln ↓ Gln	Arg ↓ Cys	Leu ↓ Phe			
Codon Variant	GCA ↓ GGA	-	CAG ↓ CAA	CAG ↓ CAC	AGT ↓ AGA	ACT ↓ ACA	ACT ↓ GCT	GAT ↓ GAA	GAT ↓ GTT	AGA ↓ AGC	AGA ↓ AAA	TCC ↓ GCC	-	CAA ↓ CAG	CGT ↓ TGT	CTT ↓ TTT			
Hap1	G	-	C	C	A	A	T	A	T	T	C	A	-	T	G	G	0.684 ^{ab}	14.41 ^{**}	149
Hap2	C	-1	T	G	T	T	C	T	A	G	T	C	-2	C	A	A	0.189 ^b		2
Hap3	C	-1	T	G	T	T	C	T	A	G	T	C	-2	C	A	G	1.382 ^a		31

Fig. 4. d) Os11t0444900-01

Chromosome 11	Exon 1				Exon 2				Means of cadmium contents (Units: $\mu\text{g}/\text{kg}$)	F-value	Acc No.
SNPPosition	145	221	233	447	512	614	801	831			
Ref. Sequence	A	C	T	T	G	G	A	G			
Allele	A/T	C/T	T/G	T/C	G/A	G/A	A/G	G/A			
Amino acid substitution	Thr ↓ Ser	Ala ↓ Val	Phe ↓ Cys	Leu ↓ Leu	Thr ↓ Thr	Lys ↓ Lys	Ser ↓ Gly	Val ↓ Ile			
Codon Variant	ACG ↓ TCG	GCC ↓ GTC	TTC ↓ TGC	TTG ↓ CTG	ACG ↓ ACA	AAG ↓ AAA	AGC ↓ GGC	GTC ↓ ATC			
Hap1	A	C	T	T	G	G	A	G	0.684 ^{ab}	14.41 ^{**}	149
Hap2	T	T	G	C	A	A	A	G	0.189 ^b		2
Hap3	T	T	G	C	A	G	A	G	1.382 ^a		31

Fig. 4. Haplotype analysis of the sequence covering the candidate genes on chromosome 11 coding regions in 182 accessions and phenotypic variation among different haplotypes. The statistical analysis was performed in SAS 9.0 package. a) Os11t0444400-01, b) Os11t0444700-01, c) Os11t0444800-01, d) Os11t0444900-01.

로 구분하였으며 그룹의 평균 카드뮴 함량은 각각 0.684, 0.189, 1.382 $\mu\text{g kg}^{-1}$ 이었다. Os11g0444900-Hap1 그룹은 Exon1과 2에서 염기변화가 없었으며, Os11g0444900-Hap2 그룹은 Exo

n1의 145번과 221번, 233번 위치에서 A에서 T, C에서 T, T에서 G로 변환되었고, 아미노산 서열이 Threonine(ACG)에서 Serine (TCG), Alanine(GCC)에서 Valine(GTC), Phenylalanine(TT

C)에서 cysteine(TGC)으로 바뀌었다. 그리고 Exon2 447번과 512번, 614번 위치에서 T에서 C, G에서 A, G에서 A로 변하였으며, Leucine(TTG)에서 Leucine(CTG), Threonine(ACG)에서 Threonine(ACA), Lysine(AAG)에서 Lysine(AAA)로 아미노산 서열이 바뀌었다. Os11g0444900-Hap3 그룹은 Exon2의 614번, 801번, 831번 위치를 제외한 모든 부분에서 allele를 가지는 동시에 아미노산 서열이 바뀌었다. Os11g0444900-Hap1 그룹에 속하는 품종은 149개로 카드뮴 함량의 범위는 0.052~3.034 $\mu\text{g/kg}$ 이었다. 그리고 Os11g0444900-Hap2 그룹에 속하는 품종은 2개로 카드뮴 함량의 범위는 0.119~0.260 $\mu\text{g/kg}$ 이었으며, Os11g0444900-Hap3 그룹에 속하는 품종은 31개로 카드뮴 함량의 범위는 0.172~3.906 $\mu\text{g/kg}$ 이었다(Fig. 4d).

기존 연구 결과와의 비교

온대자포니카 182점을 GWAS 분석한 결과 확인된 후보유전자(Os01g0611300, Os01g0611900, Os01g0611950, Os01g0612000, Os01g0612200, Os11g0444400, Os11g0444700, Os11g0444800, Os11g0444900)들은 기존에 밝혀진 벼의 카드뮴 흡수, 이행, 축적 및 내성에 관여하는 유전자(OsIRT1, OsIRT2, OsHMA3 및 OsLCT1 등)들과 달랐다. 이와 같이 GWAS 분석을 통하여 확인된 후보유전자와 기존에 밝혀진 유전자들이 서로 일치하지 않는 원인은 연구방법상의 차이에서 찾을 수 있다. 기존 연구는 주로 양적형질유전자좌위(QTLs) 분석에 의해 RFLP나 SSR 등과 같은 분자표지를 이용하기 때문에 정확한 유전자에 대한 정보를 제공하는데 한계가 있다. 한편 GWAS 분석은 분석 집단의 크기(subpopulation structure)에 영향을 받아 위양성(false positive) 결과를 가져올 수 있지만 (Setakis et al. 2006), FDR 검정(false discovery rate test)을 사용하여 위양성을 최소화 할 수 있다. 그리고 GWAS 분석은 100만개의 SNP와 형질 간의 연관성을 찾는 분석방법이므로 QTL 분석 방법에서 나타나지 않은 유전자가 확인될 수 있다(Yoon 2016).

유전체 재분석과 관련하여 기존에 발표된 연구에 따르면 Yuko et al. (2011)는 Omachi에 대해 229만여 개의 read를 사용하여 89.7%의 mapping rate와 SNP 132,462개, insertion 16,448개, deletion 19,318개를 찾아 유전체재분석을 하였다. 그러나 본 연구에서 사용한 유전체 재분석 결과는 기존 연구에 비해 mapping rate가 96.4%로 높았으며 SNP와 Indel의 개수가 많아 보다 신뢰도가 높다 할 수 있다. 따라서 보다 신뢰도가 높은 유전체 재분석 데이터를 활용하여 GWAS 분석 및 FDR검정을 수행하여 위양성(false positive)을 최소화 하였고, 280점의 분석 결과와

온대자포니카 품종군의 분석 결과에서 일치하는 염색체의 위치에서 후보유전자를 확인하였다. 또한 Nipponbare 벼 (temperate Japonica)의 염기서열을 기준으로 하여 확인된 후보유전자의 영역을 haplotype하고 유사한 염기를 갖는 품종들의 그룹 간 카드뮴 함량을 비교하였다. 또한 haplotype 분석결과를 가지고 염기 변화가 아미노산 서열을 바꾸는 non-synonymous SNP인지를 확인하였다.

염색체 1번에서 선발된 후보유전자들 모두에서 그룹의 평균 카드뮴 함량 (2.014 $\mu\text{g/kg}$)이 높으며 Exon 부분에서 Nipponbare 염기에 대해 대립염기(allele)를 갖는 9개 품종들을 확인할 수 있었다. 또한 9개의 품종은 아미노산 서열을 바꾸는 non-synonymous SNP를 갖고 있었다. 한편 염색체 11번에서 선발된 후보유전자 모두에서 33개의 품종이 Exon 부분에서 대립염기(allele)를 가지며, non-synonymous SNP를 갖는 것으로 확인되었다. 33개의 품종 중 31개의 품종이 속한 그룹의 평균 카드뮴 함량은 1.382 $\mu\text{g/kg}$ 으로 높은 반면, 2개의 품종이 속한 그룹의 평균 카드뮴 함량은 0.189 $\mu\text{g/kg}$ 으로 낮았다.

이상의 결과를 종합해 볼 때 온대자포니카 임에도 reference sequence로 사용된 Nipponbare의 염기에 대해 대립염기(allele)를 갖는 품종들은 인디카에서 유래되어 온 품종들일 가능성이 높다고 판단되었고, 인디카가 대체적으로 자포니카에 비해 등숙기 종자 내 카드뮴 함량이 높다는 기존의 보고들과 유사하였다 (Morishita et al. 1986, Arao & Ae 2003).

우리나라에서 2000년대 초반에 내염성이 강한 간척지 적응형 벼 품종을 개발하기 위해서 인디카 계통의 유전자를 자포니카 계통에 도입하여 품종을 육종하였다는 사실에 비추어 볼 때, 염색체 1번과 11번에서 Nipponbare 염기에 대해 대립염기(allele)를 가지는 품종들이 인디카 계통의 내염성 유전자를 이어 받았을 것이라 추정 할 수 있다. 다만, 인디카 계통의 유전자가 도입된 모든 품종의 카드뮴 함량이 높다고 단정 지을 수 없기 때문에 품종들의 육종 계보도를 통해 인디카 계통과 교배되었는지 여부를 검정할 필요가 있고, 본 연구에서 확인된 후보유전자 내에서 Nipponbare의 염기에 대해 allele를 갖는 온대자포니카 품종들과 인디카 품종들의 염기가 상호 유사한지를 비교·분석할 필요가 있다.

적 요

본 연구는 비오염지 논에서 재배된 벼 품종 280점의 카드뮴 함량의 차이를 조사하고 카드뮴 함량과 관련된 유전자를 확인하

고자 수행하였다. 벼 종자를 산분해법으로 전처리 한 뒤 유도결합 플라즈마질량분석기(ICP-MS)로 카드뮴 함량을 분석한 결과 평균 0.943 (0.050~5.699) $\mu\text{g/kg}$ 수준이었다. 그리고 국내외에서 수집한 유전자원 280점을 선발하여 생태형 (온대자포니카, 열대자포니카, 인디카, Aus, Aromatic 및 Admixture) 별로 구분한 뒤, 전장유전체(whole genome)에 대한 유전자 분석 및 각각의 염색체에 대한 단일염기다형성(SNP)을 분석한 후 온대자포니카 계통 182점의 등숙기 종자 내 카드뮴 함량을 전장유전체 분석하여 카드뮴과 연관된 후보유전자를 확인하였다. 그리고 Os01g0611300, Os01g0611900, Os01g0611950, Os01g0612000, Os01g0612200, Os11g0444400, Os11g0444700, Os11g0444800 및 Os11g0444900 등의 후보유전자가 쌀알 중의 카드뮴 함량과 강한 상관관계가 있음을 확인하였다. 후보유전자들에 대한 haplotype 분석 결과 온대자포니카 임에도 allele들을 갖는 염색체 1번의 9개 품종과 염색체 11번의 33개 품종은 인디카에서 유래되어 온 품종들일 가능성이 높을 것으로 판단되었고, 인디카가 대체적으로 자포니카에 비해 등숙기 종자 내 카드뮴 함량이 높다는 기존의 보고들과 유사하였다.

사 사

본 연구는 농촌진흥청 연구개발사업 (과제번호: PJ01116, PJ010923)의 지원에 의해 이루어진 것임.

REFERENCES

1. Arao T, Ae N. 2003. Genotypic variations in cadmium levels of rice grain. *Soil Sci. Plant Nutr.* 49: 473-479.
2. Arao T, Ishikawa S. 2006. Genotypic differences in cadmium concentration and distribution of soybean and rice. *JARQ.* 40: 21-30.
3. Horiguchi H, Aoshima K, Oguma E, Sasaki S, Miyamoto K, Hosoi Y, Katoh T, Kayama F. 2010. Latest status of cadmium accumulation and its effects on kidneys, bone, and erythropoiesis in inhabitants of the formerly cadmium-polluted Jinzu River Basin in Toyama, Japan, after restoration of rice paddies. *Int.Arch.Occup. Environ. Health.* 83: 953-970.
4. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, Li W, Guo Y, Deng L, Zhu C, Fan D, Lu Y, Weng Q, Liu K, Zhou T, Jing Y, Si L, Dong G, Huang T, Lu T, Feng Q, Qian Q, Li J, Han B. 2012. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature genetics.* 44: 32-39.
5. Joint FAO/WHO Expert Committee on Food Additives. 2010. no. 73.
6. Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwang JG, Kim TS, Cho EG, Park YJ. 2007. PowerCore: a program applying the advanced M strategy with a heuristic search for establishing core sets. *Bioinformatics.* 23: 2371-2162.
7. Kim TS, He Q, Kim KW, Yoon MY, Ra WH, Li FP, Tong W, Yu J, Win TO, Choi BU, Heo EB, Yun BK, Kwon SJ, Kwon SW, Cho YH, Lee CY, Park BS, Park YJ. 2016. Genome-wide resequencing of KRICE_CORE reveals their potential for future breeding, as well as functional and evolutionary studies in the post-genomic era. *BMC Genomics.* 17: 408.
8. Kim WL, Park BJ, Park SW, Kim JK, Kwon OK, Jung GB, Lee JK, KIM JK. 2008. Relationship between fraction of Cd in paddy soils near closed mine and Cd in polished rice cultivated on the same fields. *Korean J. Soil sci. Fert.* 41: 184-189.
9. Lee S, An G. 2009. Over-expression of OsIRT1 leads to increased iron and zinc accumulations in rice. *Plant Cell Environ.* 32: 408-416.
10. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z. 2012. GAPIT: genome association and prediction integrated tool. *Bioinformatics.* 28: 2397-2399.
11. Morishita T, Fumoto N, Yoshizawa T, Kagawa K. 1986. Varietal differences in cadmium levels of rice grains of Japonica, Indica, Javanica, and hybrid varieties produced in the same plot of a field. *Soil Sci. Plant Nutr.* 33: 629-637.
12. Setakis E, Stirnadel H, Balding DJ. 2006. Logistic regression protects against population structure in genetic association studies. *Genome Res.* 16: 290-296.
13. Takijima Y, Katsumi F, Koizumi S. 1973. Cadmium contamination of soils and rice plants caused by zinc mining. III. Effects of water management and applied organic manures on the control of Cd uptake by plants. *Soil Sci. Plant Nutr.* 19: 183-193.
14. Uruguchi S, Fujiwara F. 2012. Cadmium transport and tolerance in rice: perspectives for reducing grain cadmium

- accumulation. *Rice*. 5: 5.
15. Yoon MY. 2016. Genome-wide Association studies (GWAS) of related genes for flowering time in rice core set. Kongju National University.
16. Yuko AK, Yuh S, Hideki N, Kaworu E, Yoshikawa H, Yano M, Wakasa K. 2011. Discovery of genome-wide DNA polymorphisms in a landrace cultivar of Japonica rice by whole-genome sequencing. *Plant Cell Physiol*. 52: 274-282.
17. Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature genetics*. 42: 355-360.