Supplementary Information

Title: Genome-wide identification of DNA methylation provides insights into the association of gene expression in rice exposed to pesticide atrazine

Name of authors: Yi Chen Lu^{a,b}, Sheng Jun Feng^c, Jing Jing Zhang^{a,b}, Fang Luo^{a,b}, Shuang Zhang^d, Hong Yanga*

Institute: ^aJiangsu Key Laboratory of Pesticide Science, College of Sciences, Nanjing Agricultural University, Nanjing 210095, China; ^bKey Laboratory of Monitoring and Management of Crop Diseases and Pest Insects, Ministry of Agriculture, Nanjing Agricultural University, Nanjing, China; c Department of Biochemistry and Molecular Biology, College of Life Science, Nanjing Agricultural University, Nanjing 210095, China; ^dState key laboratory of food science and technology, Jiangnan University, Wuxi 214122, China

Authors:

Yi Chen Lu, 2013202040@njau.edu.cn Sheng Jun Feng, 2014216035@njau.edu.cn Jing Jing Zhang, 2014202046@njau.edu.cn Fang Luo, 2013111013@njau.edu.cn Shuang Zhang, shuangzhang@jiangnan.edu.cn Hong Yang, hongyang@njau.edu.cn

TABLE OF CONTENT : Supplementary Figures S1-14, Supplementary Tables S1-7 and Supplementary Note

• Supplementary Figures S1-14

Supplementary Figure S1. Sequencing and mapping summary. (A) The cumulative methylcytosine distribution of effective sequencing depth. The *y* axis indicates the cumulative methylcytosine distribution of CK (control, ATR-free) and ATR-exposed rice under each effective sequencing depth of cytosine (*x* axis). (B) The genomic coverage under different depths of sequencing reads. This figure reflects the overall trend between genomic coverage and sequencing depth. (C) Distribution of the methylation level in each sequence context. The *y* axis indicates the fraction of all methylcytosines that display each methylation level (*x* axis), where methylation level is the mC/C ratio at each reference cytosine (at least 10 reads required). (D) Levels of methylated cytosines and non-methylated cytosines in whole genome. The bar charts on the top of pie graphs indicate levels of methylated cytosines for each pattern. (E) Percentage of identified methylated cytosines for each sequence context. Blue numbers indicate the number of methylcytosines.

Supplementary Figure S2. Bisulphite-PCR validation of non-CG DNA methylation in an essential detoxification gene LOC_Os10g38470.1.

Supplementary Figure S3. The density profiling of methylcytosines in chromosomes and different genomic regions. (A) Levels of CG methylation and non-CG methylation in chromosomes and different genetic regions (3'-UTR, 5'-UTR, CDS, intron and mRNA). (B–E) Blue dots indicate methylcytosine density in CK and ATR-treated tissues in 10-kb windows throughout chromosome 2 and 7. Smoothed lines represent the methylcytosine density in each context in CK and ATR-treated tissues. Red arrows indicate distinct regions of CG and non-CG methylation between the two treatments (B versus C and D versus E).

Supplementary Figure S4. (A) The density profile of methylcytosines in chromosomes of rice (Data of chromosome 5 are in Fig. 2A, B). Blue dots indicate methylcytosine density in CK and

ATR-exposed rice in 10 kb windows throughout chromosomes. Smoothed lines represent the methylcytosine (CG、CHG and CHH) density in each context in CK (ATR-free) and ATR-exposed rice. (B) The methylation (CG, CHG and CHH) distributions in different genomic regions. The heatmap of genomic region distribution characteristics displays the number of methylcytosines (4X). The *x* axis indicates the number of methylcytosines in 200 bp windows displaying each mean of methylation levels (*y* axis). Black thread represents the methylated median value in specific density of CHG and CHH. Red zone reflects the particular methylated levels from shallow to deep and the number of methylcytosines. The above blue bar chart represents the density distribution of CHG and CHH and mapping to the horizontal axis. Green column on the right side display the distribution of methylation level. CK: ATR-free.

Supplementary Figure S5. Quantitative real-time PCR validation of genes which were both |Log2foldchange|≥1 in differentially expressed genes (DGEs) profiling and (Log2µ**treatment/**µ**CK ≥ 1,** *x* **axis) indifferenttially methylated valuses. These genes encoding proteins presented as follows.** (A) LOC Os06g13560.1, SAM dependent carboxyl methyltransferase; (B) LOC Os06g37300.1, cytochrome P450; (C) LOC Os10g38470.1, glutathione S-transferase; (D) LOC Os03g28940.1, ZIM domain containing protein; (E) LOC Os07g48870.1, MYB family transcription factor; (F) LOC_Os05g41080.1, histone H3; (G) LOC_Os02g12380.1, histone deacetylase; (H) LOC_Os02g19924.1, aminotransferase; (I) LOC_Os05g34854.2, gibberellin 20 oxidase 2; (J) LOC_Os03g40540.1, cytochrome P450.

Supplementary Figure S6. Methylation at various genic regions (upstream, genebody and downstream) differentially associated with gene expression. Relative methylation density (*x* axis) as a function of gene expression (*y* axis), with the logarithm value of expression fold change increasing from bottom to top. Colored lines represent data point density.

Supplementary Figure S7. Profiling of the DNA methylation marks in the specific regions of genes LOC_Os03g28940.1 (*OsJAZ6*, A) and LOC_Os07g48870.1 (MYB transcription factor, B).

Supplementary Figure S8. The expression and methylation level of ATR metabolism-based genes in shoot and root under ATR exposure. ML, methylation level.

Supplementary Figure S9. Expression of genes associated with DNA methylation under ATR exposure. The data were retrieved from datasets from the high-throughput sequencing of rice exposed to ATR (see Materials and Methods) and are represented as the ratio of Log₂ (TPM_{ATR}/TPM_{CK}) .

Supplementary Figure S10. The effect of ATR on DNA methylation and expression of transposons and retrotransposons. (A) The numbers of transposons and retrotransposons with

differential DNA methylation levels (≥ 2-fold, *p*<0.05) under ATR stress at various genomic regions. (B) The numbers of transposons and retrotransposons with differential expression $(Log₂|fold)$ change ≥ 1 , $p \leq 0.05$) under ATR stress in rice shoots and roots. (C) Venn diagrams display the overlap between differentially expressed genes (DEGs) and differentially methylated genes (DMGs) in shoots and roots. (D) The expressions of transposons and retrotransposons with differential expression and methylation level. Data were presented as Log₂foldchange in left chart and log_{10} FPKM in heat map. (E) Venn diagrams indicate the overlap between the genes nearby transposons/retrotransposons and those with differential expression in shoots and roots. The eight genes which are the nearest distance to each transposons and retrotransposons, present as the nearby genes. (F) Go enrichment analysis of 178 DEGs nearby ATR-responsive transposons/retrotransposons. (G) Heat map represents the DEGs nearby the eight transposons/retrotransposons with differential expression and methylation level. Data were presented as log_{10} FPKM in heat map.

Supplementary Figure S11. UPLC/Q-TOF-MS² analysis of atrazine (ATR) and its metabolites and conjugates in rice. The extracted ion chromatograms were obtained from full-scan chromatogram. (A) Extracted ion chromatogram of *m/z* 216.1011 (ATR), *m/z* 228.1286 (AME, internal standard) and their negative control (CK). The metabolites as follow: (B) the peak of *m/z* 232.0955 at 4.26 min and *m/z* 214.0849 at 5.74 min; (C) the peak of *m/z* 212.1499 at 4.31 min and *m/z* 202.0849 at 7.53 min; (D) the peak of *m/z* 198.1342 at 2.07 min and *m/z* 188.0701 at 3.71 min; (E) the peak of *m/z* 174.0538 at 2.43 min and *m/z* 146.0229 at 1.66 min. The conjugates as follow: (F) the peak of *m/z* 599.2006 at 8.42 min and *m/z* 487.2061 at 2.82 min; (G) the peak of *m/z* 430.1852 at 2.78 min and *m/z* 388.1751 at 1.76 min; (H) the peak of *m/z* 358.1642 at 2.83 min, *m/z* 335.1056 at 6.48 min and *m/z* 318.1411 at 2.01 min; (I) the peak of *m/z* 315.1596 at 5.33 min and *m/z* 301.1426 at 3.08 min.

Supplementary Figure S12. Analysis of ATR metabolites. The ion MS² spectra were obtained from (A) *m/z* 228.1286 precursor ion, (B) *m/z* 216.1011 precursor ion, (C) *m/z* 232.0955 precursor ion, (D) *m/z* 214.0849 precursor ion, (E) *m/z* 212.1499 precursor ion, (F) *m/z* 202.0849 precursor ion, (G) *m/z* 198.1342 precursor ion, (H) *m/z* 188.0701 precursor ion, (I) *m/z* 174.0538 precursor ion and (J) *m/z* 146.0229 precursor ion of degradation products.

Supplementary Figure S13. Analysis of ATR conjugates. The ion MS² spectra were obtained from (A) *m/z* 599.2006 precursor ion, (B) *m/z* 487.2061 precursor ion, (C) *m/z* 430.1852 precursor ion, (D) *m/z* 388.1751 precursor ion, (E) *m/z* 358.1642 precursor ion, (F) *m/z* 335.1056 precursor ion, (G) *m/z* 318.1411 precursor ion, (H) *m/z* 315.1596 precursor ion and (I) *m/z* 301.1426 precursor ion of conjugates.

Supplementary Figure S14. Gene expression of the T-DNA insertion mutants. Relative to WT, OsRDR1 (accession # H0643 form the rice retrotransposon Tos17 insertion lines http://tos.nias.affrc.go.jp/), JMJ706 (accession # PFG_K-00085.R form http://signal.salk.edu/cgi-bin/RiceGE), Met (PFG_K-02237.L), SDJ724 (PFG_3A-02454.L), SDJ714 (PFG_2A-30024.L), OsDRM2 (PFG_3A-05515.R) and OsRos1 (PFG_1B-00939.R)

expression was silenced in seedlings of each T-DNA insertion mutants, respectively, as evidenced by both qRT-PCR (top) and semi-quantitative RT-PCR (bottom) amplifications with gene-specific primers. Rice actin and ubiquitin were used as internal reference, respectively.

• Supplementary Tables S1-7

Supplementary Table S1. Output data of BS-Seq

Sample	Raw reads	Raw data	Mapped reads	Mapped data	Average	Whole genome average
	(M)	(Gb)	(M)	(Gb)	map rate $(\%)$	coverage depth (X)
CK (control)	220.53	19.85	115.25	10.37	52.26	27.70
ATR	196.67	17.70	71.39	6.42	36.30	17.16

Supplementary Table S2. Data description of BS-Seq reads for the two rice samples

Supplementary Table S3. The effective coverage of chromosomes and intergenic regions

Supplementary Table S4. Distribution of DMRs in chromosome

Supplementary Table S5. 674 genes from the interconnection of DNA methylome (differential DNA methylation levels \geq 2-fold, $p < 0.05$) and transcriptome (differential genes expression \geq 2-fold, $p < 0.05$).

Supplementary Table S6. Summary of all MS2 data for metabolites of atrazine in rice

#-1, internal standard; #-2, chemical technical; t_R^a , retention time; MS² main fragments^b, base peak of MS² fragment ions are shown in bold. ^c Compounds that have been reported for the first time.

Supplementary Table S7. Primer sequences used for this study

• Supplementary Note

Description of effect of DNA methylation inhibitors on accumulation of ATR residues and its metabolites (conjugates) in rice

Under positive ionization, the standard ATR chromatogram showed a retention time of 11.46 min (Fig. S11A), which was confirmed by the presence of $[M+H]$ ⁺ ions with m/z 216 (Fig. S12B). Decomposition of *m/z* 216 precursor ions upon MS/MS analysis produced three fragment ions at m/z 174, m/z 145 and m/z 132 (Fig. S12B). Hydroxyisopropyl atrazine (HIA, metabolite #1), which peaked at 3.89 min was confirmed by loss of H2O and isopropenyl to form *m/z* 214 and *m/z* 174, and the loss of chlorine to form *m/z* 136 from precursor ion of *m/z* 174 (Fig. S11B and S12C). For metabolite #2 (*m/z* 214), which peaked at 5.74 min, the fragment ion of *m/z* 172 from the loss of isopropenylacetylene indicated that the double bond occurred on the ethyl group (Fig. S11B and S12D). The fragment ion of m/z 130 was generated by the loss of -N-CH=CH₂ from m/z 172. Metabolite #2 was identified as dehydrogenated atrazine (DHA). The mass spectrum of metabolite #3 peaked at 4.31 min, and the fragment ion of m/z 212 was produced by loss of chlorine and addition of methoxy group from ATR (Fig. S11C and S12E). Its main product ions were m/z 170 and m/z 158, which was the loss of isopropyl and $-NH$ −CH₂−CH₃ groups, respectively. Thus, metabolite #3 was identified as a triazine herbicide (Atraton), which was more toxic than atrazine¹. Demethylated atrazine (DMA, metabolite $#4$), with the elimination of $-CH_3$ (14 Da) and further elimination of –NH2−CH2−CH3 (42 Da), was identified according to the fragment ion of *m/z* 174 and m/z 132 (Fig. S11C and S12F). Hydroxyatrazine (HA, metabolite #5), which peaked at 2.07 min, was derived from atrazine by substitution of a hydroxyl group for the chlorine, while the loss of isopropenylacetylene and −NH−CH2−CH3 groups resulted in formation of fragment ions of *m/z* 156 and *m/z* 114, respectively (Fig. S11D and S12G). Metabolite #6 peaked at 3.71 min with two main fragment ions at *m/z* 174/146 which were generated by the cleavage of −CH₃ and isopropyl groups (Fig. S11D and S12H). Base on the mass spectra, metabolite #6 was desethylated atrazine (DEA). From the accurate mass value of $[M+H]$ ⁺ ion $(m/z \ 174)$ and its fragments at m/z 146/132/118/104, it can be determined that the metabolite #7 was deisopropylated atrazine (DIA), *N*-demethylated from DMA (Fig. S11E and S12I). Beside the products, a major metabolite was detected with a retention time of 1.66 min (Fig. S11E). The information of MS/MS spectra showed that the elimination of amino group led to occurrence of fragment ions at *m/z* 128, from which metabolite #8 as didealkyl atrazine (DACT) (Fig. S12J).

Compared to the controls (ATR-free) and treatments (ATR- and AZA+ATR-), seven atrazine-glutathione metabolites and two glucosylated-atrazine conjugates were identified. The peak at 8.42 min was detected and revealed a glucosylated ATR-derivative (Conjugate #1) with main $[M+H]^+$ ions at m/z 333/299/267/249 (Fig. S11F and S13A). Ion of m/z 333 was attributed to the elimination of Cys-*β*-Ser (175 Da) and Y^{1,4} ion (89 Da) groups from its precursor. Subsequently, the loss of chlorine, thiol and hydroxyl groups from *m/z* 333 led to the formation of *m/z* 299, *m/z* 267 and *m/z* 249, respectively, suggesting that the conjugate was identified as cysteinyl β-serine and glucosamine conjugate by dehydrogenated atrazine (DHA+Cys & Ser+GlcN-H₂O, conjugate #1). A protonated molecule (m/z 487, t_R =2.82 min) was created by conjugation (Fig. S11F and S13B). The loss of glutamic acid (Glu) for *m/z* 487 formed the fragment ion of *m/z* 358, and the cleavage of thiol group generated the fragment ion of *m/z* 214. These data were consistent with the conjugation of ATR and glutathione (GSH) (ATR-HCl+GSH, conjugate #2). Conjugate #3 (m/z 430, t_R =2.78 min) was determined by its fragment ions at m/z 301/214/172, suggesting that the compound was created from conjugate #2 by the hydrolysis of peptide linkage between glycine and cysteine (Fig. S11G and S13C). Thus, conjugate #3 was identified as cysteinyl γ -glutamic acid conjugate by

chlorine substitution for atrazine (ATR-HCl+Cys & Glu). A major ATR conjugate (*m/z* 388) eluting at 1.76 min was detected according to $MS²$ (Fig. S11G). The fragment ion of m/z 299 was the result of the loss of serine. Subsequently the loss of imine and hydroxy groups from precursor ion (*m/z* 299) generated the fragment ion of *m/z* 267, confirming that of from conjugate #3, conjugate #5 (ATR) was cysteinyl *β*-serine conjugate by chlorine substitution for atrazine (ATR-HCl+Cys & Ser) (Fig. S13D). The conjugate #5 (*m/z* 358) peaking at 2.83 min were detected on the basis of the fragment ions of *m/z* 301/214, suggesting that cysteinyl *β*-glycine linked to ATR by thiol bond (Fig. S11H and S13E). So the parent ion of *m/z* 358 should point to ATR-HCl+(Cys & Gly). As shown in Fig. S13F, the MS² data with fragment ion of m/z 214 indicated that the compound was formed by conjugation of DHA and cysteine. The loss of $H₂O$ from the parent ion of m/z 335 formed the fragment ion of m/z 317. Thus, the conjugates #6 at 6.48 min was confirmed as a cysteine conjugate for DHA (DHA+Cys) (Fig. S11H). The peak at 2.01 min was detected by MS and revealed a glucosylated ATR-derivative (conjugate #7) with [M+H]+ ions at *m/z* 318 (Fig. S11H and S13F). The precursor ion yielded characteristic fragment ion at *m/z* 156, which could be attributed to the loss of a glucose moiety (162 Da). Thus, the parent ion of m/z 318 was assigned to DIHA+Glc-H₂O, a sugar ring conjugate by substation for dimethyl hydroxyl atrazine via the *O*-glycosidic bond. A peak (t_R =5.33 min) at m/z 315 was detected (Fig. S11I and S13G). The MS² spectrum generated from *m/z* 315 ion showed fragment ions at *m/z* 267, *m/z* 197 and *m/z* 151. Elimination of hydroxyl, amino and methyl groups from *m/z* 315 resulted in the formation of *m/z* 267 ion. Cleavage of −CH2−S− bond and loss of isopropyl group led to the formation of *m/z* 197 ion, suggesting that a cysteine linked to methylated atrazine. The *m/z* 151 was created by the loss of thiol group from the precursor ion (*m/z* 197) (Fig. S13G). A major ATR conjugate (*m/z* 301) eluting at 4.98 min with stronger intensity than other conjugates was detected according to $MS²$ (Fig. S11I, S13H). For m/z 301, the loss of NH₃ produced the fragment ion of m/z 284. The MS² data with the fragment ions of m/z 214 and m/z 172 showed that a protonated thiol atrazine generated (Fig. S13H). The conjugate was clearly identified as a cysteine conjugate by chlorine substitution for atrazine (ATR-HCl+Cys, conjugate #9).

Our recent study characterized several atrazine metabolites in alfalfa (*Medicago sativa* L.)², including that hydroxyisopropyl atrazine (HIA, metabolite #1), dehydrogenated atrazine (DHA, metabolite #2), demethylated atrazine (DMA, metabolite #4), hydroxyatrazine (HA, metabolite #5), desethylated atrazine (DEA, metabolite #6), DHA+Cys (conjugate #6), ATR-HCl+CH₃+Cys (conjugate #8) and ATR-HCl+Cys (conjugate #9). Beside these metabolites, we found metabolite #3 (Atraton), metabolite #8 (DACT), conjugate #2 (ATR-HCl+GSH), conjugate #3 (ATR-HCl+Cys & Glu), conjugate #5 (ATR-HCl+Cys & Gly) and conjugate #7 (DIHA+Glc-H₂O) was reported in rice for the first time. Moreover, conjugate #1 and conjugate #4, were reported in plant for the first time.

References

- 1. Tchounwou, P. B. *et al*. Toxicity assessment of atrazine and related triazine compounds in the microtox assay, and computational modeling for their structure-activity relationship. *Int. J. Mol. Sci.* **1**: 63−74 (2000).
- 2. Zhang J. J, Lu Y. C. & Yang H. Chemical modification and degradation of Atrazine in Medicago sativa through multiple pathways. *J. Agr. Food Chem.* **62**: 9657−9668 (2014).