Insights into the species-specific metabolic engineering of glucosinolates in radish (*Raphanus sativus* L.) based on comparative genomic analysis

Jinglei Wang¹, Yang Qiu¹, Xiaowu Wang¹, Zhen Yue², Xinhua Yang², Xiaohua Chen¹, Xiaohui Zhang¹, Di Shen¹, Haiping Wang¹, Jiangping Song¹, Hongju He³ & Xixiang Li¹

Glucosinolates (GSLs) and their hydrolysis products present in Brassicales play important roles in plants against herbivores and pathogens as well as in the protection of human health. To elucidate the molecular mechanisms underlying the formation of species-specific GSLs and their hydrolysed products in *Raphanus sativus* L., we performed a comparative genomics analysis between *R. sativus* and *Arabidopsis thaliana*. In total, 144 GSL metabolism genes were identified, and most of these GSL genes have expanded through whole-genome and tandem duplication in *R. sativus*. Crucially, the differential expression of *FMOGS-OX2* in the root and silique correlates with the differential distribution of major aliphatic GSL components in these organs. Moreover, *MYB118* expression specifically in the silique suggests that aliphatic GSL accumulation occurs predominantly in seeds. Furthermore, the absence of the expression of a putative non-functional epitiospecifier (ESP) gene in any tissue and the nitrile-specifier (NSP) gene in roots facilitates the accumulation of distinctive beneficial isothiocyanates in *R. sativus*. Elucidating the evolution of the GSL metabolic pathway in *R. sativus* is important for fully understanding GSL metabolic engineering and the precise genetic improvement of GSL components and their catabolites in *R. sativus* and other Brassicaceae crops.

Glucosinolates (GSLs), a large class of sulfur-rich secondary metabolites whose hydrolysis products display diverse bioactivities, function both in defence and as an attractant in plants, play a role in cancer prevention in humans and act as flavour compounds. GSLs are present in the pungent plants of the order Brassicales, which consists of sixteen families, including Brassicaceae, Capparidaceae, Caricaceae and Euphorbiaceae (specifically the genus *Drypetes*). There is great variation in GSL components and contents from species to species, with more than 200 types of GSLs identified. These diverse GSLs have been classified into three major groups based on the structures of their various amino acid precursors: the aliphatic GSLs, derived from methionine, isoleucine, leucine or valine; the aromatic GSLs, derived from phenylalanine or tyrosine; and the indole GSLs, derived from tryptophan. The basic biosynthesis pathway of aliphatic and aromatic GSLs comprises three phases: side-chain elongation, core structure formation and secondary modifications.

The glucosinolate–myrosinase system involved in Brassicales secondary metabolism has been well studied. GSLs and myrosinase (thioglucoside glucohydrolase, EC 3.2.1.1471) are normally separated into the partitioned spaces of cells. GSLs can be enzymatically hydrolysed into several different types of breakdown products, such as isothiocyanates (ITCs), epithionitriles, and nitriles, which differ among species, depending on the enzyme, substrate, pH, and presence of iron ions, while GSLs and myrosinase are encountered when tissues have been disrupted. ITC, a pungent compound specific to Brassicaceae plants, is well known to exert antimicrobial
activity against bacteria and fungi in plants\textsuperscript{12,13} and to effectively decrease the carcinogenic risk of colon and lung cancer\textsuperscript{14,15}. In contrast, epithionitriles and nitriles have been shown to have little potential for conferring health-benefits\textsuperscript{16,17}.

Radish (\textit{Raphanus sativus} \textit{L}., \textit{2n} = \textit{2x} = 18), a member of the Brassicaceae tribe in the plant family Brassicaceae\textsuperscript{14}, is a historically cultivated crop worldwide. \textit{R. sativus} is a relative of \textit{B. rapa} and \textit{B. oleracea}. The main GSLs in the seeds of \textit{B. rapa} and \textit{B. oleracea} are progoitrin and glucopinan\textsuperscript{18,19}, both of which are aliphatic; however, in the roots of \textit{B. rapa} and \textit{B. oleracea}, the predominant GSL is glucoraphanin, which is aromatic\textsuperscript{19}. Glucoraphanin, which accounts for 70–95\% of the total GSLs in seeds of \textit{R. sativus}, and glucoraphatin, which can account for 50–90\% of the total GSLs in roots, are aliphatic GSLs specific to \textit{R. sativus}\textsuperscript{20–22}. Aliphatic GSLs are predominant in the seeds of \textit{R. sativus}, which has also been demonstrated in \textit{B. rapa} and \textit{B. oleracea}\textsuperscript{18,19}. In \textit{B. rapa} and \textit{B. oleracea}, the major GSL hydrolysis products are nitriles and epithionitriles rather than ITCs\textsuperscript{23,24}. In contrast, ITCs are preferentially produced in \textit{R. sativus}, whereas a small amount of nitriles and epithionitriles are formed\textsuperscript{25}.

In the model plant \textit{A. thaliana}, many genes have been investigated in the context of controlling GSL biosynthesis\textsuperscript{26–28}. Furthermore, in comparative studies with \textit{A. thaliana}, the genes that control GSL biosynthesis in Brassicaceae vegetables have been identified and characterized. One hundred two and 105 GSL biosynthesis genes were identified in \textit{B. rapa}\textsuperscript{27} and \textit{B. oleracea}\textsuperscript{28,29,30}, respectively. Moreover, 87 and 104 GSL biosynthesis genes were reported in two \textit{R. sativus} genomes\textsuperscript{29,30}, respectively. The greatest variation between the GSL profiles of \textit{R. sativus} and \textit{B. rapa}, \textit{B. oleracea} and \textit{R. sativus} is mainly attributed to the duplication or loss of two genes, \textit{GRS1} and \textit{AOP}\textsubscript{2}\textsuperscript{31}. \textit{GRS1} catalyses the dehydrogenation reaction to generate the unsaturated 4-carbon GSL from glucoerucin or glucoraphanin to obtain glucoraphasin or glucoraphenin, respectively; these reactions are specific to \textit{R. sativus}\textsuperscript{31}. \textit{AOP2} and \textit{AOP3} catalyse the formation of alkenyl and hydroxyalkyl GSLs, respectively\textsuperscript{28}. Only one \textit{AOP2} is functional in \textit{B. oleracea}, and no \textit{AOP2} homologue has been identified in \textit{R. sativus}; however, three copies of \textit{AOP2} have been identified in \textit{B. rapa}, and no \textit{AOP3} homologue has been identified in these three vegetables. Therefore, low amounts of sinigrin and glucopinan were found in \textit{R. sativus}\textsuperscript{27–29}. Epithiospecifier protein (ESP) and nitrile-specifier protein (NSP) can catalyse GSL to produce nitriles and epithionitriles, respectively\textsuperscript{32,33}, but the enzyme cofactor ESP is inactivated in \textit{R. sativus}\textsuperscript{34}. Although several studies on \textit{R. sativus} genes have involved GSL biosynthetic and degradation pathways\textsuperscript{29–31}, the mechanism responsible for the existence and distribution of both species-specific GSLs and the hydrolysis products of these GSLs remains unclear, which leads to the questions of why \textit{R. sativus} seeds show substantial accumulation of aliphatic GSLs rather than aromatic and indole GSLs and why low amounts of nitriles exist in \textit{R. sativus}. Furthermore, the tissue-specific distribution of the two dominating GSLs (glucoraphenin and glucoraphasin) remains to be elucidated in \textit{R. sativus}.

In this study, we systematically identified GSL metabolism genes in \textit{R. sativus}. Furthermore, we present several interesting observations that may explain the formation and distribution of species-specific GSLs and their hydrolysed products in \textit{R. sativus}. The results of this study contribute to a more thorough understanding of how to precisely genetically modify and improvement of the the GSL metabolism pathway in \textit{R. sativus} and its relatives.

**Results**

**Identification and analysis of GSL genes in \textit{R. sativus}**. By comparing the draft \textit{R. sativus} genome\textsuperscript{35} with the \textit{A. thaliana} genome, we discovered 144 GSL genes, which were slightly fewer than the 161 GSL genes present in \textit{B. rapa}, greater than the 113 GSL genes of \textit{Arabidopsis lyrata} and the 117 GSL genes of \textit{B. oleracea}, and two-fold higher than the 68 found in \textit{A. thaliana} (Fig. 1 and Supplementary Table S1). Notably, \textit{R. sativus} homologues corresponding to 17 \textit{A. thaliana} GSL genes were not identified in this study. These \textit{A. thaliana} genes include MAM3, IPMI-SSU3, IPMDH3, CYP79F2, three EMOGS-OX genes and two transcription factors (Supplementary Table S1). In addition, TGG2 was not found in the ‘XYB36-2’ genome, but two TGG2 genes were reported by Leong \textit{et al.}\textsuperscript{30}. Regarding gene numbers, there was less variation, with 40\% more side-chain elongation genes observed in \textit{R. sativus} compared to \textit{A. thaliana}, whereas in the co-substrate pathway, the number of genes found in \textit{R. sativus} was 2.33 times the number present in \textit{A. thaliana} (Fig. 1). As aliphatic GSLs are the major
GSLs present in R. sativus, the loss, retention and expansion of the GSL genes may have led to the accumulation of species-specific GSLs in R. sativus (Fig. 2).

Expansion of the GSL metabolic genes in R. sativus. Gene copy numbers can be expanded in four major ways: genome duplication, segmental duplication, tandem duplication and via transposable elements. Most of the retained genes (36 out of 51) were present as multi-copies in R. sativus (Supplementary Table S1), which may have resulted from tandem duplication or whole-genome triplication (WGT) occurring after the divergence of R. sativus and A. thaliana. Of the 36 expanded genes, 25 AtGSL genes have syntenic relationships to 51 RsGSL genes (Supplementary Table S2). Furthermore, 16 of the 36 over-retained AtGSL genes were tandemly duplicated and distributed in 17 tandem arrays in R. sativus (Supplementary Table S3). These data showed that tandem duplication and WGT substantially contributed to the expansion of GSL metabolic genes in R. sativus.

We analysed the number and the ratio of single-copy to multi-copy homologous genes to reveal the retention status of the three types of GSL core structure formation genes in R. sativus after the WGT. The gene copy number ratios between R. sativus and A. thaliana were 2 (6:3), 0.8 (5:4) and 1 (4:4) for aliphatic, indole and aromatic GSL core structure formation genes, respectively (Table 1). Furthermore, the downstream genes (GGP1, SUR1, UGT74C1 and ST5b), which retained more than one copy, were more frequently duplicated than upstream genes in the pathway for aliphatic GSL core structure biosynthesis in R. sativus (Fig. 2).

Notably, the ST genes were found to be highly expanded in R. sativus, and most of the ST copies (8 out of 14) in R. sativus were tandemly duplicated. R. sativus contained 11 copies of the ST5b gene, which encodes desulfo glucosinolate sulfotransferase, which is involved in the final step of GSL core structure biosynthesis. The phylogenetic tree of the ST genes in R. sativus and its closely related species revealed that ST5a and ST5b, especially the latter, have undergone the greatest amount of expansion in R. sativus, B. rapa and B. oleracea. The copies of ST5b were clustered into three groups, two of which included only ST5b copies from R. sativus, B. rapa and B. oleracea, implying that they were inherited from a common ancestor and diverged via tandem duplication (Fig. 3).

Expression patterns of FMOGS-OX and MYB118 genes in R. sativus. FMOGS-OX1-4 S-oxygenates both short- and long-chain methylthioalkyl GSLs to produce the corresponding methylsulfanylalkyl GSL.

Figure 2. Putative aliphatic GSL biosynthetic and degradation pathways in R. sativus. Glucoraphasatin is principally found in the root of R. sativus, whereas glucoraphenin is detected mainly in the flowers. The genes shown in purple represent the genes involved in aliphatic GSL biosynthesis in R. sativus, and the numbers in the red brackets are the numbers of corresponding genes in R. sativus. The yellow polygon stands for the translocator on the chloroplast membrane. The symbol ‘?’ represents unknown genes.
enzyme that has a similar function to IPMDH1 \(^44\), which promotes glucosinolate accumulation\(^45,46\). Therefore, the functional

\( R. \text{ sativus} \)

non-nitrile products exist in

\( \text{B. rapa} \) and \( \text{B. oleracea} \) respectively. Interestingly, MYB118 was expressed only in the siliques in \( \text{B. rapa} \) and \( \text{R. sativus} \) and was expressed highly in the siliques but at low levels in other tissues in \( \text{B. oleracea} \) (Fig. 4b).

Putative non-functional ESP and tissue-specific expression of NSP genes in \( \text{R. sativus} \). When tissues of Brassicales plants are damaged, myrosinase is released, and glucosinolates are degraded to isothiocyanate, thiocyanate, or nitrile derivatives. Although two \( \text{R. sativus} \) myrosinase genes have been cloned by Hara \textit{et al}.\(^40\) and a total of 11 \( \text{R. sativus} \) myrosinase genes were identified by Mitsui \textit{et al}.,\(^29\), it remains unclear why non-nitrile products exist in \( \text{R. sativus} \). We found one ESP gene in \( \text{R. sativus} \); we also identified two copies of ESP in \( \text{B. rapa} \) and five copies in \( \text{B. oleracea} \). The RNA-seq data indicated that the ESP genes were not expressed in \( \text{R. sativus} \) but were expressed in both \( \text{B. rapa} \) and \( \text{B. oleracea} \) (Fig. 5). NSP, whose product promotes the breakdown of glucosinolates into nitrile derivatives, was found to be retained as one copy in \( \text{R. sativus} \) and as two copies in \( \text{B. rapa} \). Interestingly, in \( \text{R. sativus} \), the NSPs gene was expressed at a minimal level in the flowers, seedpods and callus and was not expressed in the roots, but one of the copies in \( \text{B. rapa} \) was highly expressed (Fig. 5).

Discussion

We identified the counterparts of most of the \( \text{A. thaliana} \) GSL metabolic genes, which are present in various copy numbers in \( \text{R. sativus} \). Compared with the GSL genes identified in two previously reported genomes, the copy numbers were the same for most of the identified genes\(^29,30\), with the exception of TGG2. The two \( \text{R. sativus} \) TGG2 genes reported by Jeong \textit{et al}.\(^30\) are not considered to be homologues, since these two genes were identified by using homologue searches based on the sequence of AT4G11150, which does not encode the \( \text{A. thaliana} \) TGG2 homologue. The corresponding homologue of TGG2 (Rs303830) identified by Jeong \textit{et al}.,\(^30\) were identified in the ‘XYB36-2’ and ‘Aokubi’ genomes, but they represented the homologues of TGG1 in these two genomes\(^29\).

The qualitative and quantitative changes in the GSL profiles of \( \text{R. sativus} \) may be due to the loss, expansion, non/neo-functionality and tissue/stage-specific expression of GSL genes. According to our comprehensive and comparative analysis based on genome and transcriptome data from this study and previous research, these changes corresponded to the evolution of the \( \text{R. sativus} \) genome\(^29–31,35\). In terms of side-chain elongation, the absence of MAM3 led to a lack of long-chain aliphatic GSLs in \( \text{R. sativus} \), and this result is consistent with Mitsui \textit{et al}.,\(^29\). We also found that the loss of CYP79F2, of which the knockout decreases the abundance of long-chain aliphatic GSLs in \( \text{A. thaliana} \), may also result in short-side-chain aliphatic GSLs in \( \text{R. sativus} \)\(^31,42\). Considering that functional MAM3 genes were found in \( \text{B. rapa} \) and \( \text{B. oleracea} \), the loss of CYP79F2 might be the main reason for the predominance of short-side-chain aliphatic GSLs observed in these two species. In addition, no homologues of IPMI-SSU3 or IPMDH3 from \( \text{A. thaliana} \) were found in \( \text{R. sativus} \). The IPMIs, which are composed of a large subunit and a small subunit, catalyse the isomerization reaction. IPMI-SSU3 encodes the small subunit. However, knockout mutants of IPMI-SSU3 result in small changes in GSL levels in \( \text{A. thaliana} \)\(^43\). IPMDH3 is a predicted enzyme that has a similar function to IPMDH1\(^44\), which promotes glucosinolate accumulation\(^35,46\). Therefore, the loss of IPMI-SSU3 and IPMDH3 is likely to be offset by IPMI-LSU1, IPMI-LSU2 and IPMDH1.

Moreover, duplicated genes may enhance the potential for the quantitative variation of a particular trait\(^47\). Interestingly, most of the GSL genes were present in multiple copies. Syntenic analysis demonstrated that GSL genes have increased their copy numbers through WGT. In addition to WGT, some genes also expanded in number through tandem duplication, such as ST5b, which is present in 11 copies in \( \text{R. sativus} \). Notably, expansion was observed for more aliphatic GSL core structure formation genes than for indole and aromatic GSL genes. Further, the biosynthesis genes for the core structures of aliphatic GSLs that were present in downstream locations were more over-retained than those in upstream locations. The redundancy of the downstream genes may guarantee products for successful synthesis of aliphatic GSLs.

Furthermore, with respect to the secondary modification step, three FMOGS-OX genes (FMOGS-OX1, FMOGS-OX3 and FMOGS-OX4) and two AOP genes (AOP2 and AOP3) were lost. FMOGS-OX1–4 S-oxygenates both short- and long-chained methylthioalkyl GSLs to produce the corresponding methysulfynalkyl GSLs\(^38\). In \( \text{R. sativus} \), we identified one copy of FMOGS-OX2, which catalyses the S-oxygenation of glucoraphasatin and converts glucoraphasatin into glucoraphenin. The results of the GSL content analyses performed in previous

| Types of GSLs | Number of homologues with different copies | Ratio of single to multiple copies |
|---------------|-------------------------------------------|----------------------------------|
|               | 0  | 1  | 2  | 3  | Total |                    |
| aliphatic GSLs| 1  | 3  | 1  | 5  | 10   | 3:6                |
| indole GSLs   | 0  | 5  | 2  | 2  | 9    | 5:4                |
| aromatic GSLs | 0  | 4  | 2  | 2  | 8    | 4:4                |

Table 1. Number and ratio of single-copy to multi-copy homologues of GSL core structure formation genes in \( \text{R. sativus} \).
studies demonstrated that glucoraphasatin is predominant in the roots and leaves of *R. sativus*, whereas glucoraphenin accumulates heavily in the seeds. Correspondingly, *FMOGS-OX2* was highly expressed in the silique and minimally expressed in roots. The tissue-specific differential expression of *FMOGS-OX2* clearly results in the different predominant GSLs in various tissues of *R. sativus*. The *FMOGS-OX2* gene of *R. sativus* shows biotechnological potential, as the cancer-preventive properties of the plant involve sulforaphene, which is the

---

**Figure 3.** Phylogenetic tree of the ST genes in *R. sativus* and four related species. The full-length amino acid sequences were aligned with ClustalW, and the NJ tree was constructed with MEGA using 1000 bootstrap replicates. Each ST gene is indicated along the lines on the right. *R. sativus* proteins are marked with solid red dots. The solid black, yellow, green and blue dots represent *A. thaliana*, *A. lyrata*, *B. rapa* and *B. oleracea*, respectively.
Additionally, we found no AOP2 or AOP3 genes in *R. sativus*, as reported by Mitsui *et al.* The loss of these two genes prevents the conversion of S-oxygenated and hydroxyalkyl GSLs into down-stream GSLs, which is the reason for the accumulation of glucoraphasatin and glucoraphenin in *R. sativus*.

In addition, two transcription factors (MYB76, MYB115) were lost in *R. sativus*. Although aliphatic GSL levels will increase with increased MYB76 expression levels, a knockout mutant of MYB76 was reported to exhibit no significant change in GSLs in *A. thaliana*. MYB115 and MYB118 are functionally redundant and interact to control the expression of GLS genes. Therefore, the loss of MYB76 and MYB115 does not completely abolish GSL biosynthesis. While MYB115 and MYB118 negatively regulate aliphatic GSLs in *A. thaliana*, the double

---

**Figure 4.** Expression levels of *FMOGS-OX2* and MYB118. (a) Expression levels of *FMOGS-OX* genes in various tissues of *R. sativus*, *B. rapa* and *B. oleracea*. (b) Expression levels of MYB118 in various tissues of *R. sativus*, *B. rapa* and *B. oleracea*. The coloured scale bar represents the FPKM values.

**Figure 5.** Expression of the ESP and NSP genes in *R. sativus*, *B. oleracea*, *B. rapa* and *A. thaliana*. Expression analysis of the ESP and NSP genes in: root, stem, flower, silique, callus and leaf tissues. The coloured scale bar in the upper right corner of the figure represents the log2 of FPKM values.
myb115-myb118 mutant exhibits increased levels of most of the short-chain aliphatic GSLs with the exception of 4-methylthiobutyl GSL (glucoerucin, GER). Moreover, GER content increased in the single myb115 mutant but significantly decreased in the myb118 mutant. That is, MYB118 promotes the accumulation of GER but decreases the levels of other aliphatic GSLs in A. thaliana. GER is a precursor of the predominant aliphatic GSLs in R. sativus, B. rapa and B. oleracea. Several studies have shown that aliphatic GSLs, especially GER and its downstream products, are highly predominant in seeds in the three Brassica vegetables. In addition, in other organs, the percentages of aliphatic GSLs are decreased. Considering that MYB118 was expressed only in the siliques, we inferred that its siliqua-specific expression results in the high accumulation of aliphatic GSLs in seeds of these three vegetables. Nonetheless, further work is required to elucidate the molecular basis of the regulation of GSL biosynthesis through MYB118.

The ESP and NSP proteins divert the myrosinase-catalysed hydrolysis of an aliphatic GSL from the formation of ITC to the production of epithionitrile and nitrile. O’Hare et al. found that there were no ESP proteins in R. sativus. We identified one ESP gene showing no detectable expression in any organ of R. sativus, indicating that its function might have been lost in R. sativus. We also observed one NSP gene with no expression profile in the roots and minimal expression in other organs, and we found no NSP1–4 genes within the R. sativus genome. The non-expression of the ESP gene and the lack of expression of the NSP gene in the roots and the leaves were inferred to be the major reasons for the accumulation of distinctive beneficial isothiocyanates, such as sulforaphane, rather than epithionitriles or nitriles.

**Materials and Methods**

**Data resources.** The A. thaliana genome and annotation data were downloaded from TAIR10 (http://www.arabidopsis.org) and B. rapa and B. oleracea assembly and annotation data were downloaded from the BRAD database (http://brassicadb.org/brad). The genome data for A. lyrata were obtained from the JGI database. The three whole-genome sequences of R. sativus sequenced by Zhang et al., Mitsui et al. and Jeong et al. were downloaded from the BRAD database and NODAI Radish genome database (http://www.radish-genome.org/) respectively, and B. rapa and B. oleracea RNA-seq data were obtained from the Gene Expression Omnibus (GEO) database with accession numbers GSE43245 and GSE42891, respectively. R. sativus RNA-seq data are available at EMBL/NCBI/SRA (PRJNA413464).

**GSL gene identification in R. sativus.** The GSL genes of A. thaliana have been reported previously and were aligned with corresponding protein sets from R. sativus, B. rapa, B. oleracea and A. lyrata using BLASTP (E-value ≤ 1 × 10−10; identity ≥ 55).

**Analysis of GSL genes in R. sativus.** To detect the generation mechanism of the expanded genes, we identified syntenic orthologues using SynOrths based on both sequence similarity and the collinearity of flanking genes. We identified tandem duplicate genes using the same standard with GSL gene identification, and only one unrelated gene was allowed to exist between the two genes in a tandem array. CLUSTALW was employed for sequence alignment. The phylogenetic tree of the ST gene family members of R. sativus and other species was constructed using the neighbour-joining method with MEGA (version 7.0.21) software.

**References**

1. Brader, G., Mikkelsen, M. D., Halkier, B. A. & Tapio Palva, E. Altering glucosinolate profiles modulates disease resistance in plants. _Plant J._ 46, 758–767 (2006).
2. Halkier, B. A. & Gershenzon, J. Biology and biochemistry of glucosinolates. _Annu. Rev. Plant Biol._ 57, 303–333 (2006).
3. Kim, J. H., Lee, B. W., Schroeder, F. C. & Jander, G. Identification of indole glucosinolate breakdown products with antifeedant effects on _Myzus persicae_ (green peach aphid). _Plant J._ 54, 1015–1026 (2008).
4. Zhang, Y., Kensler, T. W., Schroeder, F. C. & Jander, G. Identification of indole glucosinolate breakdown products with antifeedant effects on _Myzus persicae_ (green peach aphid). _Plant J._ 54, 1015–1026 (2008).
5. Sønderby, I. E., Geu-Flores, F. & Halkier, B. A. Biosynthesis of glucosinolates—gene discovery and beyond. _Trends Plant Sci._ 2, 1322 (2007).
6. Sønderby, I. E., Geu-Flores, F. & Halkier, B. A. Biosynthesis of glucosinolates—gene discovery and beyond. _Trends Plant Sci._ 15, 283–290 (2010).
7. Clarke, D. B. Glucosinolates, structures and analysis in food. _Anal. Methods-UK_ 2, 310–325 (2010).
8. Sønderby, I. E. et al. A systems biology approach identifies a _R2R3 MYB_ gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. _PLoS One_ 2, e1322 (2007).
9. Sønderby, I. E., Geu-Flores, F. & Halkier, B. A. Biosynthesis of glucosinolates—gene discovery and beyond. _Trends Plant Sci._ 15, 283–290 (2010).
10. Fenwick, G. R., Heaney, R. K., Mullin, W. J. & VanEtten, C. H. Glucosinolates and their breakdown products in food and food plants. _Crit. Rev. Food Sci. & Nutr._ 18, 123–201 (1983).
11. Bones, A. M. & Rossiter, J. T. The myrosinase–glucosinolate system, its organisation and biochemistry. _Physiol. Plantarum_ 97, 194–208 (1996).
12. Chumala, P. & Suchy, M. Phytoalexins from _Thlaspi arvense_, a wild crucifer resistant to virulent _Leptosphaeria maculans_; structures, syntheses and antifungal activity. _Phytochemistry_ 64, 949–956 (2003).
13. Zasadz, I. & Ferris, H. Nematode suppression with brassicaceous amendments: application based upon glucosinolate profiles. _Soil Biol. Biochem._ 36, 1017–1024 (2004).
14. Banga, O. Radish: _Raphanus sativus_ ( Cruciferae). _Evolution of Crop Plants. NW Simmonds, ed._ (1976).
15. Baenas, N. et al. Metabolic activity of radish sprouts derived isothiocyanates in _drosophila melanogaster_. _Int. J. Mol. Sci._ 17, 251 (2016).
16. Matusheski, N. V. & Jeffery, E. H. Comparison of the bioactivity of two glucoraphanin hydrolysis products found in broccoli, sulforaphane and sulforaphane nitrile. _J. Agr. Food Chem._ 49, 5743–5749 (2001).
17. Hanschen, F. S. et al. The Brassica epithionitrile 1-cyano–2, 3-epithiopropane triggers cell death in human liver cancer cells in vitro. _Mol. Nutr. Food Res._ 59, 2178–2189 (2015).
18. Hong, E., Kim, S. J. & Kim, G. H. Identification and quantitative determination of glucosinolates in seeds and edible parts of Korean Chinese cabbage. Food chem. 128, 1115–1120 (2011).
19. Bhandari, S. R., Jo, J. S. & Lee, J. G. Comparison of glucosinolate profiles in different tissues of nine Brassica crops. Molecules 20, 15827–15841 (2015).
20. Maldini, M. et al. Identification and quantification of glucosinolates in different tissues of Raphanus raphanistrum by liquid chromatography tandem-mass spectrometry. J. Food Compos. Anal. (2016).
21. Carlson, D. G., Daxenbichler, M., VanIten, C., Hill, C. & Williams, P. Glucosinolates in radish cultivars. J. Am. Soc. Hortic. Sci 110, 634–638 (1985).
22. Ciska, E., Honke, J. & Kozłowska, H. Effect of light conditions on the contents of glucosinolates in germinating seeds of white mustard, red radish, white radish, and rapeseed. J. Agr. Food Chem. 56, 9087–9093 (2008).
23. Hashimoto, S., Miyazawa, M. & Kameoka, H. Volatile flavor sulfur and nitrogen constituents of Brassica rapa L. J. Food Sci. 47, 2084–2085 (1982).
24. Blažević, I. & Mastelić, J. Glucosinolate degradation products and other bound and free volatiles in the leaves and roots of radish (Raphanus sativus L.). Food Chem. 113, 96–102 (2009).
25. Sønderby, I. E., Geu-Flores, F. & Halkier, B. A. Biosynthesis of glucosinolates—gene discovery and beyond. Trends Plant Sci. 15, 283–290 (2010).
26. Wang, H. et al. Glucosinolate biosynthetic genes in Brassica rapa. Gene 487, 135–142 (2011).
27. Liu, S. et al. The Brassica oleracea genome reveals the asymmetrical evolution of polyploid genomes. Nat. Commun. 5 (2014).
28. Mitsui, Y. et al. The radish genome and comprehensive gene expression profile of tuberous root formation and development. Sci Rep. 5 (2015).
29. Juge, N., Mithen, R. F. & Traka, M. Molecular basis for chemoprevention by sulforaphane: a comprehensive review. J. Nat. Prod. 84, 1035–1039 (2011).
30. Thiagarajan, S. & Maffei, R. F. Glucosinolate metabolism and detoxification in radish, Brassica rapa L. J. Food Sci. 47, 2084–2085 (1982).
31. Wink, M. et al. Small variation of glucosinolate composition in Japanese cultivars of radish (Raphanus sativus L.) requires simple quantitative analysis for breeding of glucosinolate component. Breeding Sci. 62, 63–70 (2012).
32. Juge, N., Mithen, R. F. & Traka, M. Molecular basis for chemoprevention by sulforaphane: a comprehensive review. Cell. Mol. Life Sci. 64, 1105–1127 (2007).
33. Gilgalashvili, T., Engqvist, M., Yatushevich, R., Müller, C. & Flügge, U. I. HAG2/MYB26 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in Arabidopsis thaliana. New Phytol. 177, 627–642 (2008).
34. Yi, G. et al. Root glucosinolate profiles for screening of radish (Raphanus sativus L.) genetic resources. J. Agr. Food Chem. 64, 61–70 (2015).
35. Kong, X. Y., Kissen, R. & Bones, A. M. Characterization of recombinant nitrile-specifier proteins (NSPs) of Arabidopsis thaliana expressed in E. coli: Dependency on Fe(II) ions and the effect of glucosinolate substrates and reaction conditions. Phytochemistry 84, 7–17 (2012).
36. Huala, E. et al. The Arabidopsis Information Resource (TAIR): a comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant. Nucleic Acids Res. 29, 102–105 (2001).
37. Cheng, F. et al. BRAD, the genetics and genomics database for Brassica plants. BMC Plant Biol. 11, 1 (2011).
38. Bednarck, P. et al. A glucosinolate metabolism pathway in living plant cells mediates broad-spectrum antifungal defense. Science 332, 101–106 (2009).
39. Grubb, C. D. & Abel, S. Glucosinolate metabolism and its control. Trends Plant Sci. 11, 89–100 (2006).
40. Wang, X. et al. The genome of the mesopolyploid crop species Brassica rapa. Nat. Genet. 43, 1035–1039 (2011).
41. Hu, T. T. et al. The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. Nat. Genet. 43, 476–481 (2011).
61. Feng, C., Jian, W., Lu, F. & Wang, X. Syntenic gene analysis between Brassica rapa and other Brassicaceae species. Front. Plant Sci. 3, 198 (2012).
62. Yu, J. et al. Genome-wide comparative analysis of NBS-encoding genes between Brassica species and Arabidopsis thaliana. BMC Genomics 15, 3–3 (2014).
63. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947–2948 (2007).
64. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874 (2016).

Acknowledgements
This work was supported by grants from the National Key Research and Development Plan of the Ministry of Science and Technology of China (2016YFD0100204-02, 2013BAD01B04), the 973 Program (2013CB127000), the 863 program (2012AA021801-04, 2012AA100101) and the Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS-XTCX2016017, 2016016-4-4, 2016001-5-3).

Author Contributions
X.L. conceived the research and designed the experiments. J.W. and Y.Q. conducted the experiments. J.W. and X.Y. analysed the data. J.W. and Y.Q. wrote the paper. X.C., X.Z., D.S., H.W., J.S. and H.H. participated in the experiment. X.L., Z.Y. and X.W. were involved in manuscript preparation. All of the authors read and approved the final manuscript.

Additional Information
Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-017-16306-4.

Competing Interests: The authors declare that they have no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017